



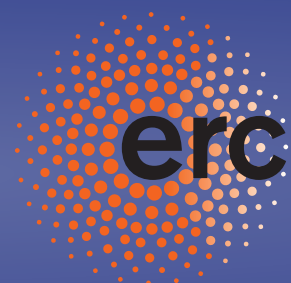
# Reasoning and Ontologies in Data Extraction

**Sergio Flesca, Ermelinda Oro** (University of Calabria), **Tim Furché** (Oxford University)

September 7<sup>th</sup>, 2012 @ Reasoning Web 2012



European Research Council





# What?



# The Age of Big Data

The New York Times

## The Age of Big Data

By STEVE LOHR

Published: February 11, 2012 | 82 Comments

GOOD with numbers? Fascinated by data? The sound you hear is opportunity knocking.

 [Enlarge This Image](#)



Mo Zhou was snapped up by I.B.M. last summer, as a freshly minted Yale M.B.A., to join the technology company's fast-growing ranks of data consultants. They help businesses make sense of an explosion of data —

 RECOMMEND

 TWITTER

 LINKEDIN

 COMMENTS (82)

 SIGN IN TO E-MAIL

 PRINT

 REPRINTS



# The Age of Big Data

The New York Times

## The Age of Big Data

By STEVE LOHR

Published: February 11, 2012 | 82 Comments

GOOD with numbers? Fascinated by opportunity knocking.

[Enlarge This Image](#)

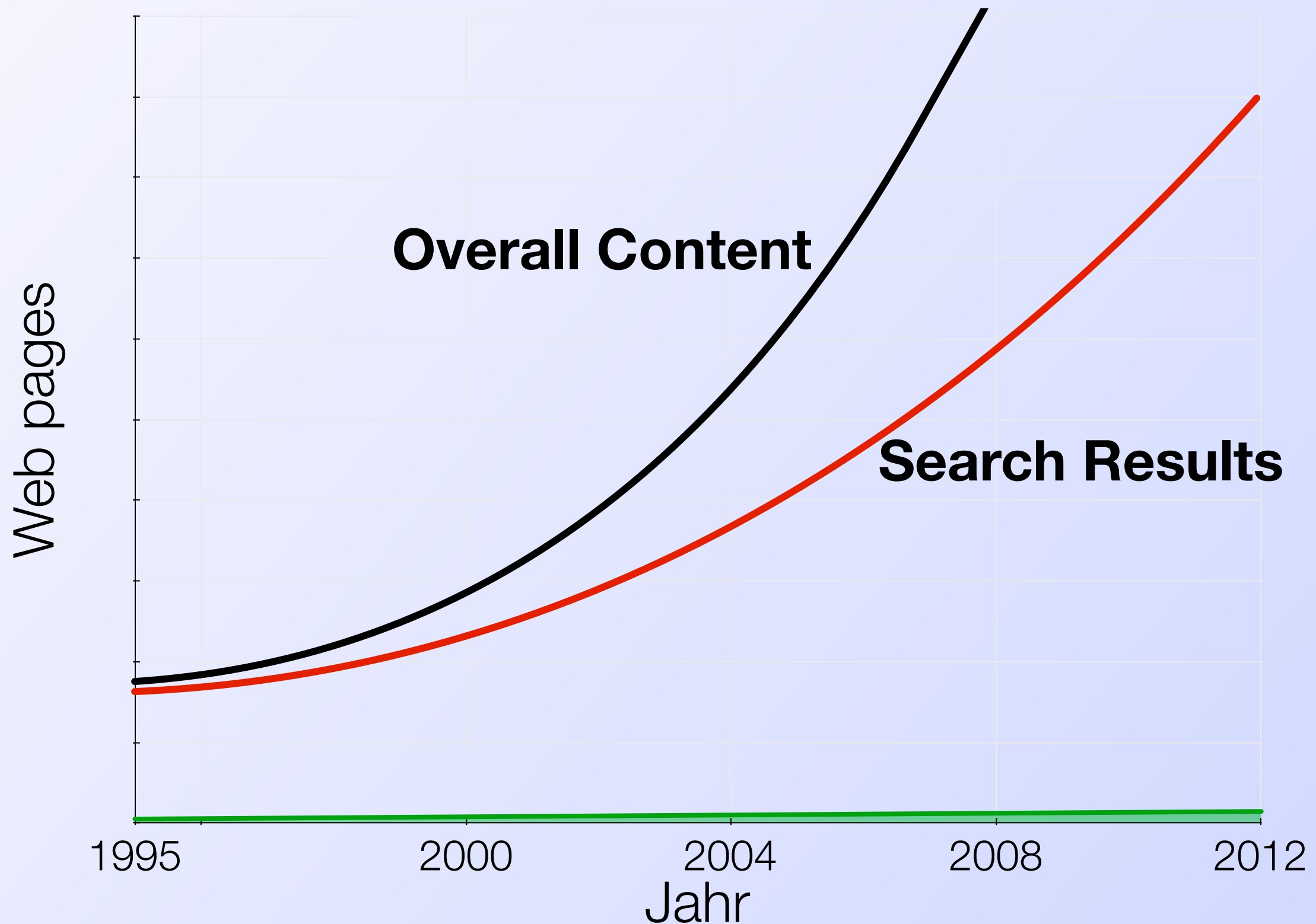


Mo 2  
last s  
M.B.  
com  
cons  
mak



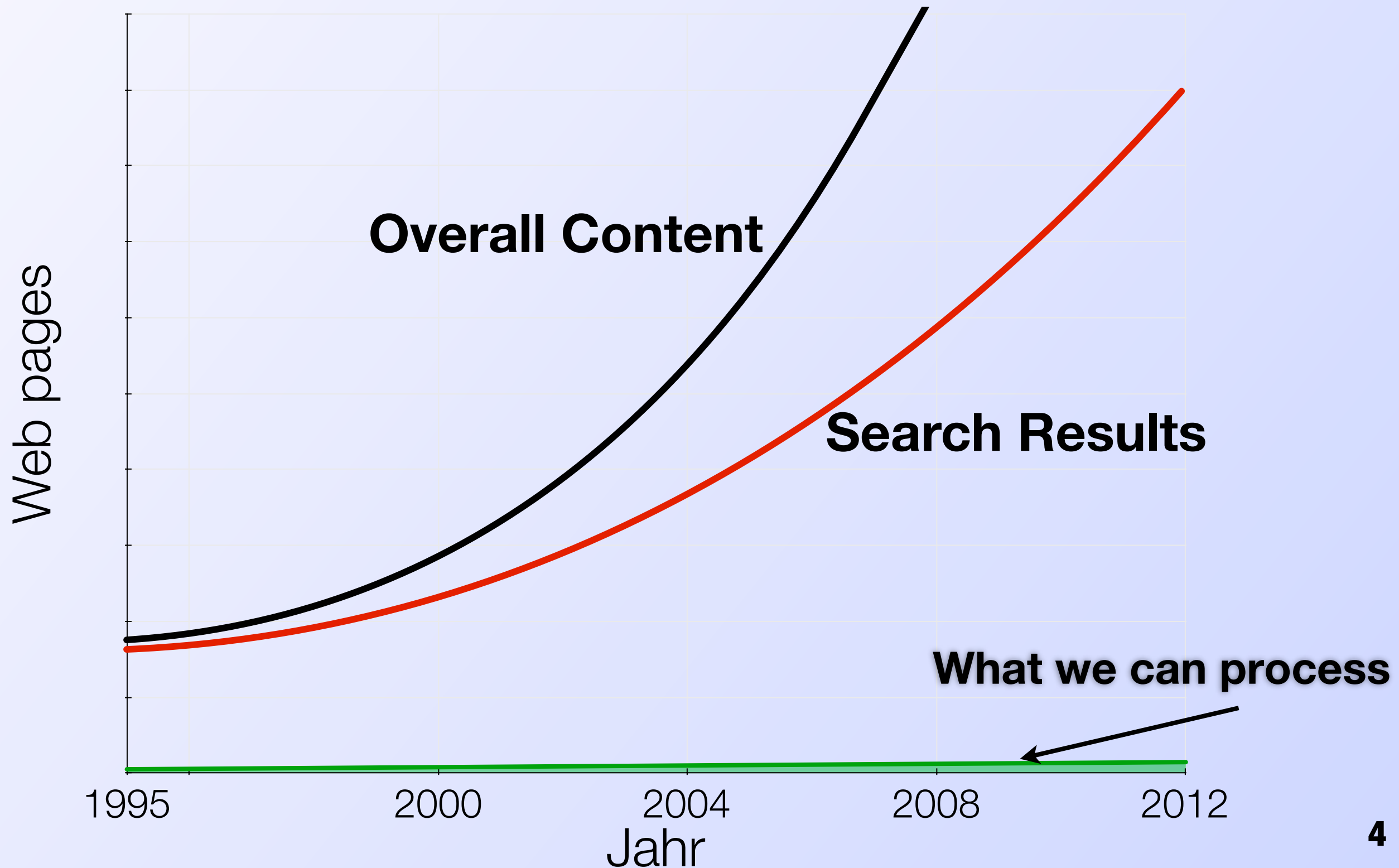


# Document search doesn't cut it (any more)





# Document search doesn't cut it (any more)





# Object Search Today @ Google

flat in oxford

Search

About 48,700,000 results (0.19 seconds)

[Advanced search](#)

## [Oxford Flats - Find Flats to Suit all Budgets | FindaProperty.com](#)

Ads

Updated Daily. Register for Alerts.

[Houses For Sale In Oxfordshire](#) - [Houses To Rent In Oxfordshire](#)

[www.findaproperty.com/flats](http://www.findaproperty.com/flats)

## [Flat In Oxford | TaylorWimpey.co.uk](#)

New **Flats** & Houses in **Oxford**. Starting from £157,995.

[www.taylorwimpey.co.uk/Oxford](http://www.taylorwimpey.co.uk/Oxford)

## [Flat In Oxford | Primelocation.com](#)

Search over 650,000 Luxury UK **Flats** from the Comfort of your Armchair!

[Houses For Sale In Oxfordshire](#) - [Houses To Rent In Oxfordshire](#)

[www.primelocation.com/flats](http://www.primelocation.com/flats)

## [Property to rent in Oxford, Oxfordshire](#)

Results 1 - 20 of 582 – Review houses, **flats** and homes to rent in **Oxford** or try the ...

- Parking Space to rent in **Oxford** – £120 pcm – unfurnished – 0.32 miles
- Garage to rent in **Oxford** – £150 pcm – unfurnished – 2 additional photos
- House Share to rent in **Oxford** – £315 pcm – Per Person furnished – 3 additional ...

[www.findaproperty.com/searchresults.aspx?edid=00...1...](http://www.findaproperty.com/searchresults.aspx?edid=00...1...) - [Cached](#) - [Similar](#)

## [Flats, flatshare rentals, Oxford - find a flatshare online](#)

Find a e.g. BMW, 2 bed **flat**, sofa; in e.g. Portslade ... 1388 ads in **Oxford**, Flatshare, Rooms to Rent Subscribe to email alerts ... East **Oxford**Date wanted: 20 Sep ...

[Wanted - Flatshare in Oxford offered - Short Term](#)

[www.gumtree.com/flatshare/oxford](http://www.gumtree.com/flatshare/oxford) - [Cached](#)

## [Flats / Houses to Rent, Oxford : Rent a house online](#)

677 ads in **Oxford**, **Flats** & Houses for Rent Subscribe to email alerts ...

[www.gumtree.com/flats-and-houses-for-rent/oxford](http://www.gumtree.com/flats-and-houses-for-rent/oxford) - [Cached](#)



# Object Search Today @ Google

flat in oxford

Search

About 48,700,000 results (0.19 seconds)

[Advanced search](#)

[Oxford Flats - Find Flats to Suit all Budgets | FindaProperty.com](#)

Ads

Updated Daily. Register for Alerts.

[Houses For Sale In Oxfordshire](#) - [Houses To Rent In Oxfordshire](#)

[www.findaproperty.com/flats](http://www.findaproperty.com/flats)

[Flat In Oxford | TaylorWimpey.co.uk](#)

New **Flats** & Houses in **Oxford**. Starting from £157,995.

[www.taylorwimpey.co.uk/Oxford](http://www.taylorwimpey.co.uk/Oxford)

[Flat In Oxford | Primelocation.com](#)

Search over 650,000 Luxury UK **Flats** from the Comfort of your Armchair!

[Houses For Sale In Oxfordshire](#) - [Houses To Rent In Oxfordshire](#)

[www.primelocation.com/flats](http://www.primelocation.com/flats)

[Property to rent in Oxford, Oxfordshire](#)

Results 1 - 20 of 582 – Review houses, **flats** and homes to

• Parking Space to rent in **Oxford** – £120 pcm – unfurnished –

• Garage to rent in **Oxford** – £150 pcm – unfurnished – 2 addit

• House Share to rent in **Oxford** – £315 pcm – Per Person furn

[www.findaproperty.com/searchresults.aspx?edid=00...1...](http://www.findaproperty.com/searchresults.aspx?edid=00...1...)

[Flats, flatshare rentals, Oxford - find a flatshare](#)

Find a e.g. BMW, 2 bed **flat**, sofa; in e.g. Portslade ... 138

to Rent Subscribe to email alerts ... East **Oxford**Date wanted

[Wanted - Flatshare in Oxford offered - Short Term](#)

[www.gumtree.com/flatshare/oxford](http://www.gumtree.com/flatshare/oxford) - [Cached](#)

[Flats / Houses to Rent, Oxford : Rent a house](#)

677 ads in **Oxford**, **Flats** & Houses for Rent Subscribe to

[www.gumtree.com/flats-and-houses-for-rent/oxford](http://www.gumtree.com/flats-and-houses-for-rent/oxford) - [Cached](#)

- doesn't understand **entity type**
- structured attributes and type keywords
- favors “big” aggregators & news sites
- with poor quality results



# Object Search Today @ Google

flat in oxford

Search

About 48,700,000 results (0.19 seconds)

[Advanced search](#)

**[Oxford Flats - Find Flats to Suit all Budgets | FindaProperty.com](#)**

Ads

Updated Daily. Register for Alerts.

[Houses For Sale In Oxfordshire](#) - [Houses To Rent In Oxfordshire](#)

[www.findaproperty.com/flats](http://www.findaproperty.com/flats)

**[Flat In Oxford | TaylorWimpey.co.uk](#)**

New **Flats** & Houses in **Oxford**. Starting from £157,955

[www.taylorwimpey.co.uk/Oxford](http://www.taylorwimpey.co.uk/Oxford)

**[Flat In Oxford | Primelocation.com](#)**

Search over 650,000 Luxury UK **Flats** from the Comfort of your Armchair!

[Houses For Sale In Oxfordshire](#) - [Houses To Rent In Oxfordshire](#)

[www.primelocation.com/flats](http://www.primelocation.com/flats)

**[Property to rent in Oxford, Oxfordshire](#)**

Results 1 - 20 of 582 – Review houses, **flats** and homes to rent in **Oxford** or try the ...

- Parking Space to rent in **Oxford** – £120 pcm – unfurnished – 0.32 miles
- Garage to rent in **Oxford** – £150 pcm – unfurnished – 2 additional photos
- House Share to rent in **Oxford** – £315 pcm – Per Person furnished – 3 additional ...

[www.findaproperty.com/searchresults.aspx?medid=00...1...](http://www.findaproperty.com/searchresults.aspx?medid=00...1...) - [Cached](#) - [Similar](#)

**[Flats, flatshare rentals, Oxford - find a flatshare online](#)**

Find a e.g. BMW, 2 bed **flat**, sofa; i.e.g. Portslade ... 1388 ads in **Oxford**, Flatshare, Rooms to Rent Subscribe to email alerts ... East **Oxford**Date wanted: 20 Sep ...

[Wanted - Flatshare in Oxford shared - Short Term](#)

[www.gumtree.com/flatshare/oxford](http://www.gumtree.com/flatshare/oxford) - [Cached](#)

**[Flats / Houses to Rent, Oxford : Rent a house online](#)**

677 ads in **Oxford**, **Flats** & Houses for Rent Subscribe to email alerts ...

[www.gumtree.com/flats-and-houses-for-rent/oxford](http://www.gumtree.com/flats-and-houses-for-rent/oxford) - [Cached](#)



# Object Search Today @ Google

flat in oxford, energy efficient, no stairs

Search

About 1,020,000 results (0.19 seconds)

[Advanced search](#)

## [OXFORD IS MY WORLD | Energy Home Energy Use](#)

**Oxford** is my world Your – Guide to **saving** the planet! ... who wants to improve the **energy efficiency** of their **house** or save **energy** at home there is ... Our 'Very Easy' **steps** show you how much **energy** you can save ... **without** spending a penny! ...

[www.oxfordismyworld.org/home\\_energy.html](http://www.oxfordismyworld.org/home_energy.html) - [Cached](#) - [Similar](#)

## [Escalator - Wikipedia, the free encyclopedia](#)

Escalator step widths and **energy** usage ..... This device actually consisted of **flat**, moving **stairs**, **not** unlike the escalators of .... the increased **efficiency** of each operator due to the elimination of **stair** climbing. .... <sup>2</sup>" The **Oxford** English Dictionary. ...

[en.wikipedia.org/wiki/Escalator](http://en.wikipedia.org/wiki/Escalator) - [Cached](#) - [Similar](#)

## [\[PDF\] THE EFFECTIVENESS OF FEEDBACK ON ENERGY CONSUMPTION](#)

File Format: PDF/Adobe Acrobat - [Quick View](#)

by S Darby - 2006 - [Cited by 148](#) - [Related articles](#)

The focus is on how people change their behaviour, **not** on the .... recognition that **energy efficiency** alone is inadequate to achieve the aims of a ..... **House**. Environmental Change Institute, University of **Oxford**, UK. Brandon G & Lewis A ...

[www.eci.ox.ac.uk/research/energy/.../smart-metering-report.pdf](http://www.eci.ox.ac.uk/research/energy/.../smart-metering-report.pdf) - [Similar](#)

## [\[PDF\] The Oxford Solar House - TVE](#)

File Format: PDF/Adobe Acrobat - [Quick View](#)

The **Oxford Solar House** is the first low **energy house** in the United Kingdom ... reduced by using all available **energy saving** technologies but **without** impairing ... service duct, **stairs** to the first floor and a hallway to the entry porch. ...

[www.tve.org/ho/series1/reports\\_7-12/reports.../theoxfordsolarhouse.pdf](http://www.tve.org/ho/series1/reports_7-12/reports.../theoxfordsolarhouse.pdf)

## [Gordon & Erika Wilson - Pre-fabricated energy-saving homes from ...](#)

**Saving energy** and the environment ... We went and knocked on the door of the neighbouring **house** there and then and asked if ... **Not** least so by the **energy efficiency**. ... To the right is a hallway leading to the **stairs**, and beyond to the study. .... +++ Planning permission granted for new build in **Oxford** +++ VIEW NEW videos ...

[www.hanse-haus.co.uk/our\\_projects/.../gordon\\_erika\\_wilson.html](http://www.hanse-haus.co.uk/our_projects/.../gordon_erika_wilson.html) - [Cached](#)



# Object Search Today @ Google

flat in oxford, energy efficient, no stairs

Search

About 1,020,000 results (0.19 seconds)

[Advanced search](#)

## [OXFORD IS MY WORLD | Energy Home Energy Use](#)

**Oxford** is my world Your – Guide to **saving** the planet! ... who wants to improve the **energy efficiency** of their **house** or save **energy** at home there is ... Our 'Very Easy' **steps** show you how much **energy** you can save ... **without** spending a penny! ...

[www.oxfordismyworld.org/home\\_energy.html](http://www.oxfordismyworld.org/home_energy.html) - [Cached](#) - [Similar](#)

## [Escalator - Wikipedia, the free encyclopedia](#)

Escalator step widths and **energy** usage ..... This device actually consisted of **flat**, moving **stairs**, **not** unlike the escalators of .... the increased **efficiency** of each operator due to the elimination of **stair** climbing. .... 2<sup>nd</sup> The **Oxford** English Dictionary. ...

[en.wikipedia.org/wiki/Escalator](http://en.wikipedia.org/wiki/Escalator) - [Cached](#) - [Similar](#)

## [\[PDF\] THE EFFECTIVENESS OF FEEDBACK ON ENERGY CONSUMPTION](#)

File Format: PDF/Adobe Acrobat - [Quick View](#)

by S Darby - 2006 - [Cited by 148](#) - [Related articles](#)

The focus is on how people change their behaviour, **not** on the .... recognition that **energy efficiency** alone is inadequate to achieve the aims of a ..... **House**. Environmental Change Institute, University of **Oxford**, UK. Brandon G & Lewis A ...

[www.eci.ox.ac.uk/research/energy/.../smart-metering-report.pdf](http://www.eci.ox.ac.uk/research/energy/.../smart-metering-report.pdf) - [Similar](#)

## [\[PDF\] The Oxford Solar House - TVE](#)

File Format: PDF/Adobe Acrobat - [Quick View](#)

The **Oxford Solar House** is the first low **energy house** in the UK using all available **energy saving** technologies but **without** impairing the first floor and a hallway to the entry porch. ...

[www.tve.org/ho/series1/reports\\_7-12/reports.../theoxfordsolarhouse](http://www.tve.org/ho/series1/reports_7-12/reports.../theoxfordsolarhouse)

## [Gordon & Erika Wilson - Pre-fabricated energy-saving](#)

**Saving energy** and the environment ... We went and knocked on a **house** there and then asked if ... **Not** least so by the **energy** ... a hallway leading to the **stairs**, and beyond to the study. .... +++ F for new build in **Oxford** +++ VIEW NEW videos ...

[www.hanse-haus.co.uk/our\\_projects/.../gordon\\_erika\\_wilson.html](http://www.hanse-haus.co.uk/our_projects/.../gordon_erika_wilson.html) - [Cached](#)



gets **worse** the more I know



doesn't understand primary object



lacks "attributes"



# Object Search Today @ Google

flat in oxford, energy efficient, no stairs

Search

About 1,020,000 results (0.19 seconds)

[Advanced Search](#)

## [OXFORD IS MY WORLD | Energy Home Energy Use](#)

**Oxford** is my world Your – Guide to **saving** the planet! ... who wants to improve the **energy efficiency** of their **house** or save **energy** at home there is ... Our 'Very Easy' **steps** show you how much **energy** you can save ... **without** spending a penny! ...

[www.oxfordismyworld.org/home\\_energy.html](http://www.oxfordismyworld.org/home_energy.html) - [Cached](#) - [Similar](#)

## [Escalator - Wikipedia, the free encyclopedia](#)

Escalator step widths and **energy** usage ..... This device actually consisted of **flat**, moving **stairs**, **not** unlike the escalators of .... the increased **efficiency** of each operator due to the elimination of **stair** climbing. .... 2<sup>nd</sup> The **Oxford** English Dictionary ...

[en.wikipedia.org/wiki/Escalator](http://en.wikipedia.org/wiki/Escalator) - [Cached](#) - [Similar](#)

## [\[PDF\] THE EFFECTIVENESS OF FEEDBACK ON ENERGY CONSUMPTION](#)

File Format: PDF/Adobe Acrobat - [Quick View](#)

by S Darby - 2006 - [Cited by 148](#) - [Related articles](#)

The focus is on how people change their behaviour, **not** on the .... recognition that **energy efficiency** alone is inadequate to achieve the aims of a ..... **House**. Environmental Change Institute, University of **Oxford**, UK. Brandon G & Lewis A ...

[www.eci.ox.ac.uk/research/energy/.../smart-metering-report.pdf](http://www.eci.ox.ac.uk/research/energy/.../smart-metering-report.pdf) - [Similar](#)

## [\[PDF\] The Oxford Solar House - TVE](#)

File Format: PDF/Adobe Acrobat - [Quick View](#)

The **Oxford Solar House** is the first low **energy house** in the United Kingdom ... reduced by using all available **energy saving** technologies but **without** impairing ... service duct, **stairs** to the first floor and a hallway to the entry porch. ...

[www.tve.org/ho/series1/reports\\_7-12/reports.../theoxfordsolarhouse.pdf](http://www.tve.org/ho/series1/reports_7-12/reports.../theoxfordsolarhouse.pdf)

## [Gordon & Erika Wilson - Pre-fabricated energy-saving homes from ...](#)

**Saving energy** and the environment ... We went and knocked on the door of the neighbouring **house** there and then and asked if ... **Not** least so by the **energy efficiency**. ... To the right is a hallway leading to the **stairs**, and beyond to the study. .... +++ Planning permission granted for new build in **Oxford** +++ VIEW NEW videos ...

[www.hanse-haus.co.uk/our\\_projects/.../gordon\\_erika\\_wilson.html](http://www.hanse-haus.co.uk/our_projects/.../gordon_erika_wilson.html) - [Cached](#)



# Microsoft Bing: “Model Every Object on the Planet”

## Google, Alert: Bing Wants “To Model Every Object On The Planet,” Reinvent Search

BY E.B. BOYD | 05-10-2012 | 3:03 AM

The future of search is not about page rank. It's about creating giant databases of every person, place, and thing on the planet.



Today [Bing](#) is announcing [a revamp of its front end](#), to make its search results more useful for users. But what's much more interesting is what's happening on the back end, underneath the hood, as Microsoft re-architects how the data used for search results is collected, stored, and repurposed.



Microsoft Bing:

## “Model Every Object on the Planet”

### Google, Alert: Bing Wants “To Model Every Object On The Planet,” Reinvent Search

BY E.B. BOYD | 05-10-2012 | 3:03 AM

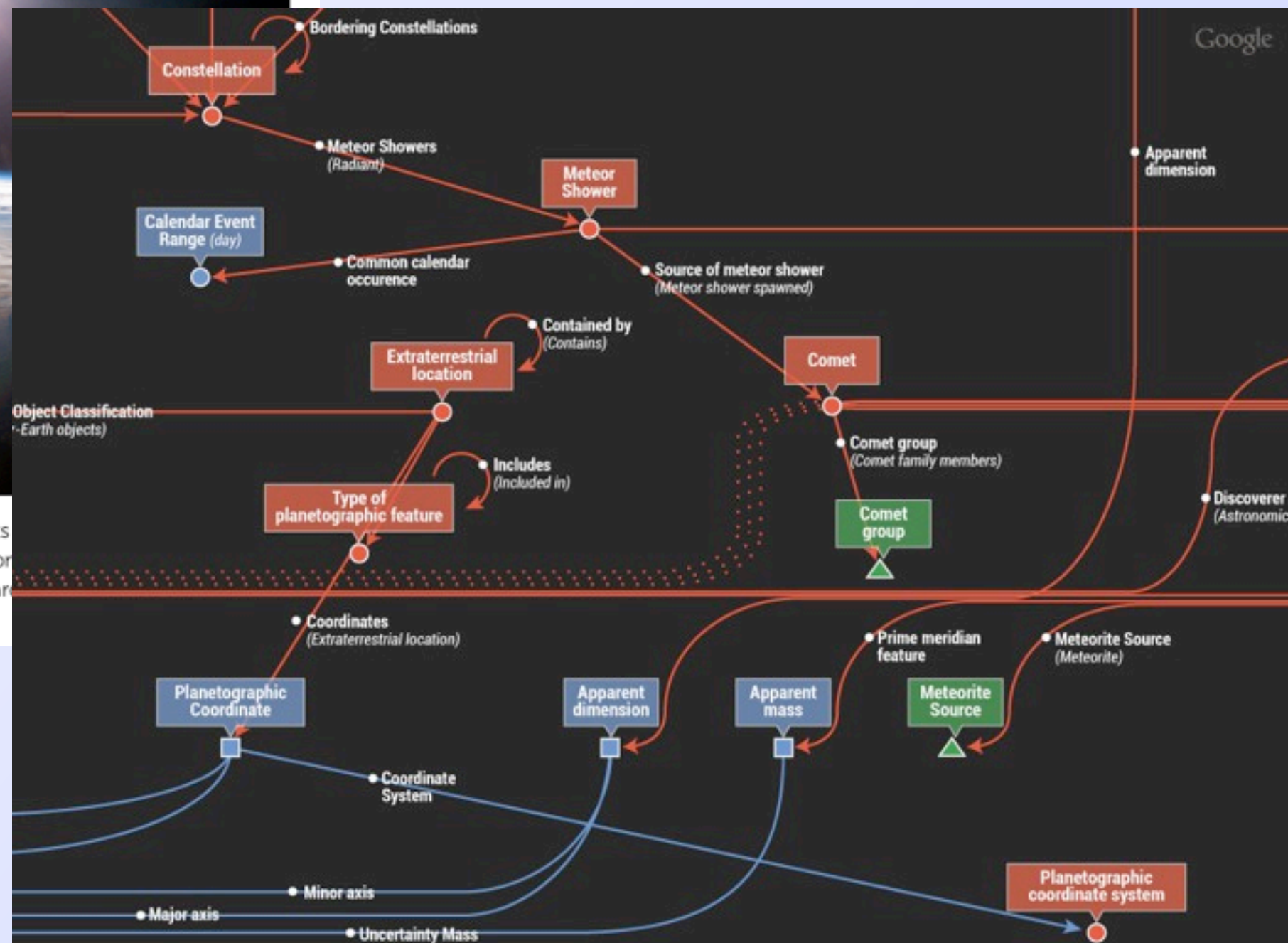
The future of search is not about page rank. It's about creating giant databases of every person, place, and thing on the planet.



Today [Bing](#) is announcing [a revamp of its front end](#), to make its users. But what's much more interesting is what's happening on the hood, as Microsoft re-architects how the data used for search and repurposed.

Google:

## “Knowledge Graph: things, not strings”





Microsoft Bing:

“Model Every Object on the Planet”

## Google, Alert: Bing Wants “To Model Every Object On The Planet,” Reinvent Search

BY E.B. BOYD | 05-10-2012 | 3:03 AM

The future of search is not about page rank. It's about creating giant databases of every person, place, and thing on the planet.

The screenshot shows a Bing search results page for 'Marie Curie'. The top navigation bar includes links for News, Gmail, Documents, Calendar, and More. A search bar with a microphone icon and a blue search button is visible, with the user's email 'avidbikar80@gmail.com' displayed to the right. Below the search bar, it indicates '10,000 other results (0.32 seconds)'. The main content area features a snippet from Wikipedia about Marie Curie, followed by a list of related links including 'the free encyclopedia', 'Nobelprize.org', and 'Science of Radioactivity'. At the bottom, there is a section titled 'Report images' with five small thumbnail images of Marie Curie and related scientific figures.

Google:

“Knowledge Graph: things, not strings”

### Marie Curie



es.wikipedia.org

Marie Skłodowska-Curie was a French-Polish physicist and chemist famous for her pioneering research on radioactivity. She was the first person honored with two Nobel Prizes—in physics and chemistry. [Wikipedia](#)

**Born:** November 7, 1867, [Warsaw](#)

**Died:** July 4, 1934, [Sancellemoz](#)

**Spouse:** [Pierre Curie](#) (m. 1895–1906)

**Children:** [Irène Joliot-Curie](#), [Ève Curie](#)

**Discovered:** [Radium](#), [Polonium](#)

**Education:** [École Supérieure de Physique et de Chimie Industrielles de la Ville de Paris](#), [University of Paris](#)

### People also search for



[Albert Einstein](#)



[Pierre Curie](#)



[Ernest Rutherford](#)



[Louis Pasteur](#)



[John Dalton](#)

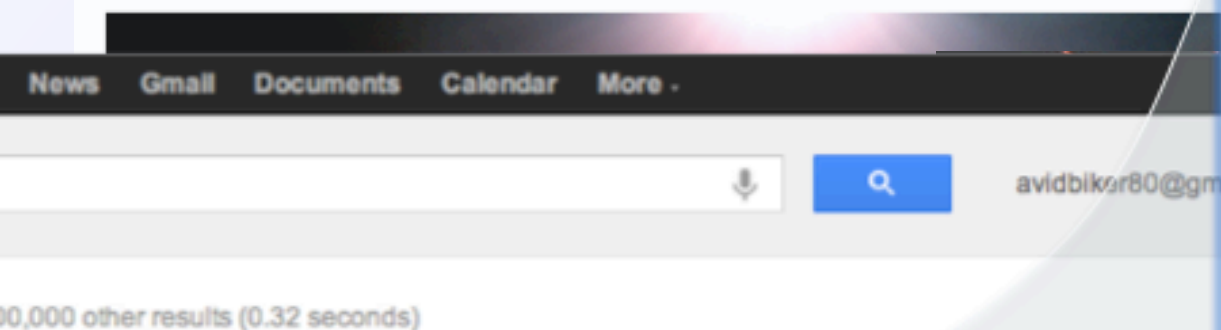


# Microsoft Bing: “Model Every Object on the Planet”

## Google, Alert: Bing Wants “To Model Every Object On The Planet,” Reinvent Search

BY E.B. BOYD | 05-10-2012 | 3:03 AM

The future of search is not about page rank. It's about creating giant databases of every person, place, and thing on the planet.



[the free encyclopedia](#)

[Curie](#)

(November 1867 – 4 July 1934) was a French-Polish physicist and chemist famous for her pioneering research on radioactivity.

[Pierre Curie](#) - [Irène Joliot-Curie](#) - [Radioactive decay](#)

[- Nobelprize.org](#)

[prizes/physics/.../marie-curie-bio.html](#)

tion that awards the Nobel Prize.

[- Nobelprize.org](#)

[prizes/chemistry/.../marie-curie-bio.html](#)

owska, was born in Warsaw on November 7, 1867, the youngest of four children. She received a general education ...

[Report images](#)



[Science of Radioactivity](#)

the AIP Center for History of Physics. Text by Naomi Robbins describes Curie's contributions to the science of ...

Google:

“Knowledge Graph:  
things, not strings”

### Marie Curie



Marie Skłodowska-Curie was a French-Polish physicist and chemist famous for her pioneering research on radioactivity. She was the first person honored with two Nobel Prizes—in physics and chemistry. [Wikipedia](#)

**Born:** November 7, 1867, [Warsaw](#)

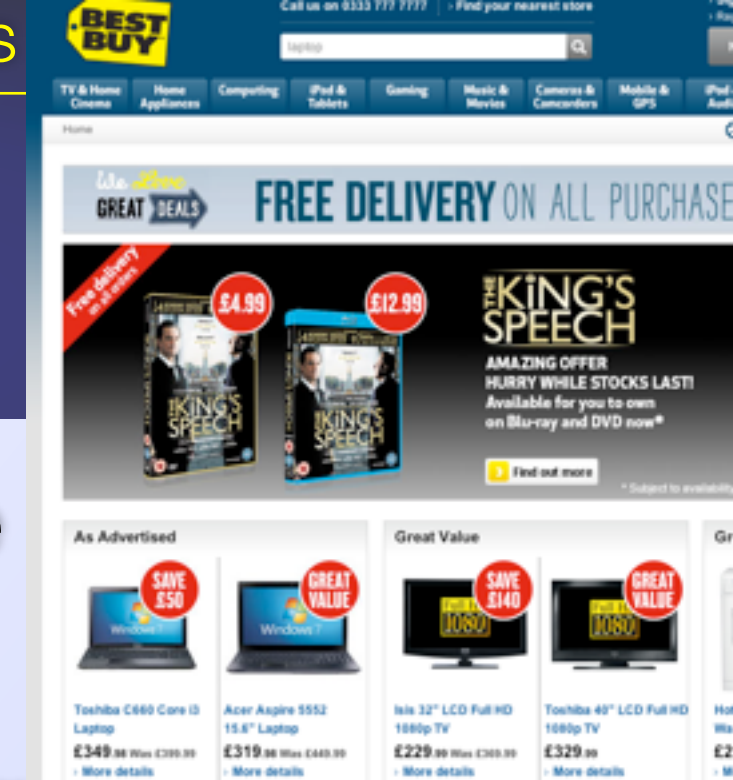
**Died:** July 4, 1934, [Sancellemoz](#)

- common sense, **static facts**
- high degree of content redundancy
- same information on many sites
- not suitable for **dynamic**, product data



# Scenario ①: Electronics retailer

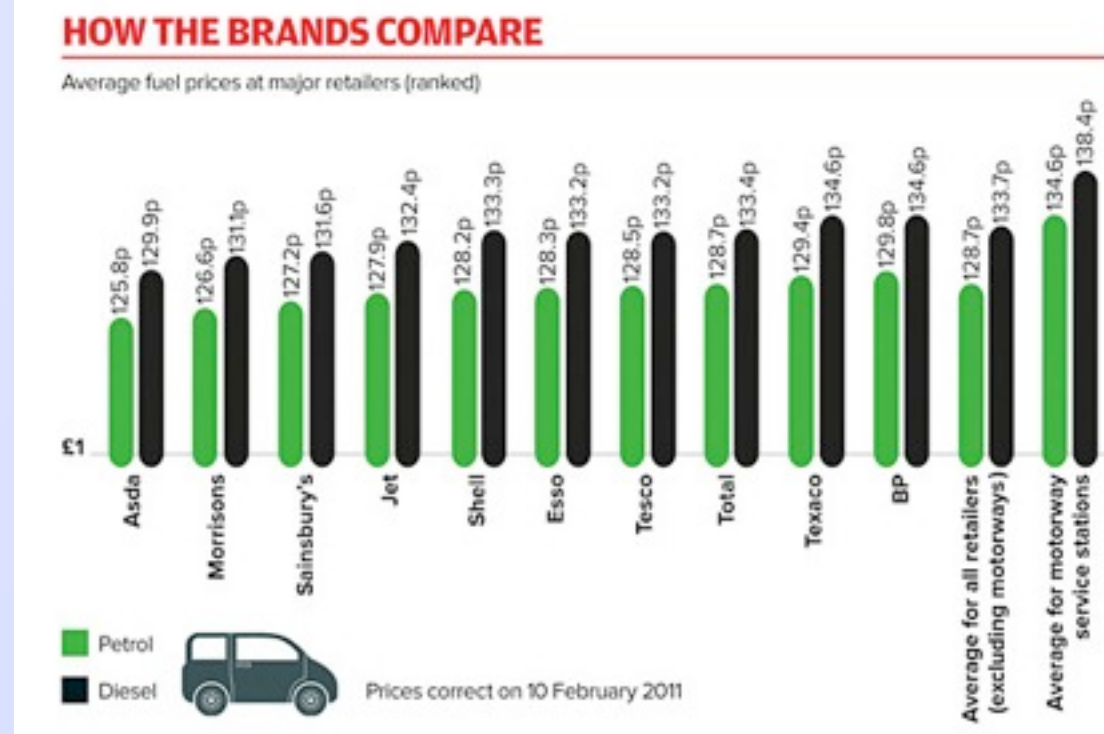
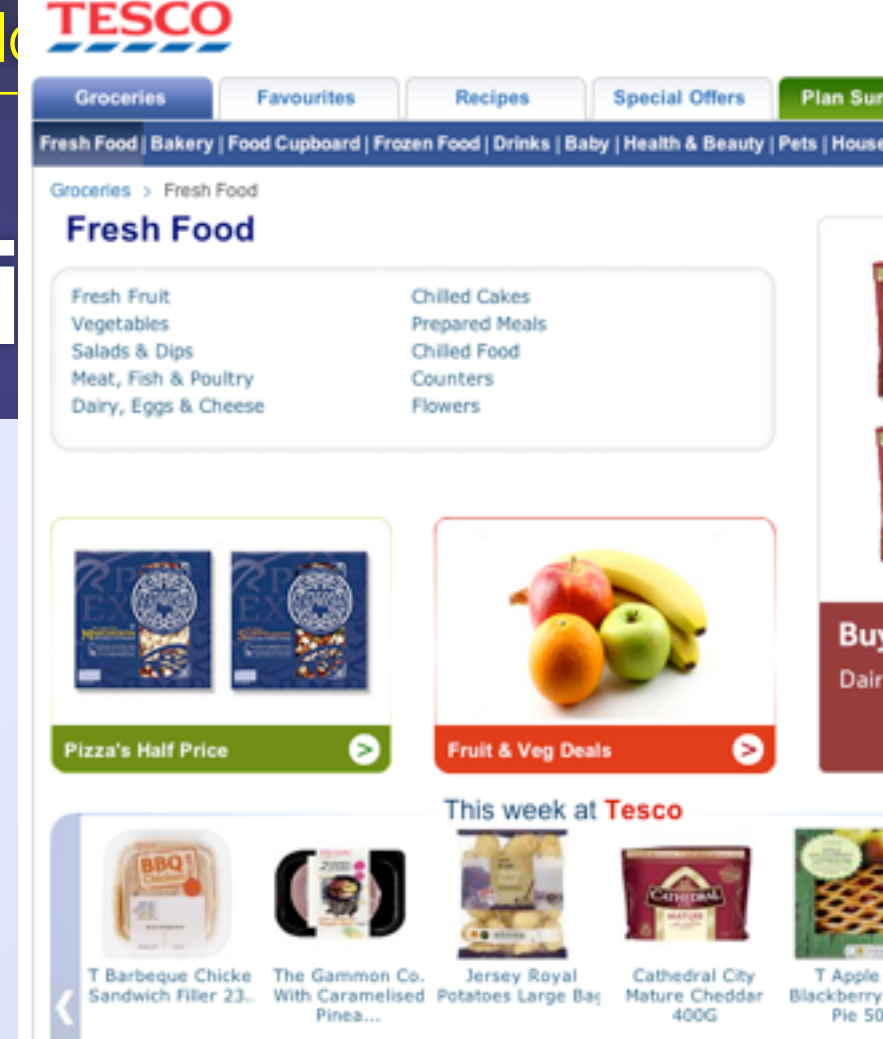
- electronics retailer: **online market intelligence**
  - comprehensive overview of the market
    - daily information on price, shipping costs, trends, product mix
    - by product, geographical region, or competitor
  - thousands of **products**
  - hundreds of competitors
- nowadays: specialised companies
  - mostly manual, interpolation
  - large cost





# Scenario ②: Supermarket chain

- supermarket chain
  - competitors' **product** prices
  - special offer or promotion (time sensitive)
  - new products, product formats & packaging





# Scenario ③: Hotel Agency

- online travel agency
- best price guarantee
- prices of competing agencies
- average market price

The screenshot shows the hotel.de website interface. On the left, there's a map of Oxford with various locations marked. On the right, under the 'Compare hotels' section, three hotels are listed:

- Ethos Hotel**: 59 Western Road, OX1 4LF Oxford, United Kingdom. Rating: 6.4 of 10 (4 reviews).
- Hawkwell House Hotel**: Church Way, OX4 4DZ Oxford, United Kingdom. Rating: 7.6 of 10 (12 reviews).
- Holiday Inn Express Oxford-Kassam Stadium**: Grenoble Road, OX4 4XP Oxford, United Kingdom. Rating: 7.3 of 10 (31 reviews).

Below the map, the 'Your search:' section shows 'Destination: Oxford' and 'Arrival date: M/D/YYYY'.

The screenshot shows the hotel.de website interface with a detailed price comparison table. The table lists various hotels and their prices, including columns for Date, HMID, Hotel, ES, Source, Arrival, Nights/PAX/Room, Cancellation, Breakfast, Price, and Diff.

Date	HMID	Hotel	ES	Source	Arrival	Nights/PAX/Room	Cancellation	Breakfast	Price	Diff.
17.04.2008	102521	artgenossen Landarhotel Umler - DE	1	hotel.de	11.05.2008	1 N. / 2 Adult / Doppelzimmer	OT	U +7.80	7	86.00
17.04.2008	144745	Insel Hotel Bonn - DE	1	hotel.de	11.05.2008	1 N. / 2 Adult / Doppelzimmer	NO_MATCH	UF	7	120.00
17.04.2008	55480	Hotel Berlin Sindelfingen - DE	2	hotel.de	11.05.2008	1 N. / 2 Adult / Design Doppelzimmer	OT	UF	7	100.00
17.04.2008	214437	Hotel One Twenty	1	hotel.de	11.05.2008	1 N. / 2 Adult / Doppelzimmer	OT	UF	7	65.00



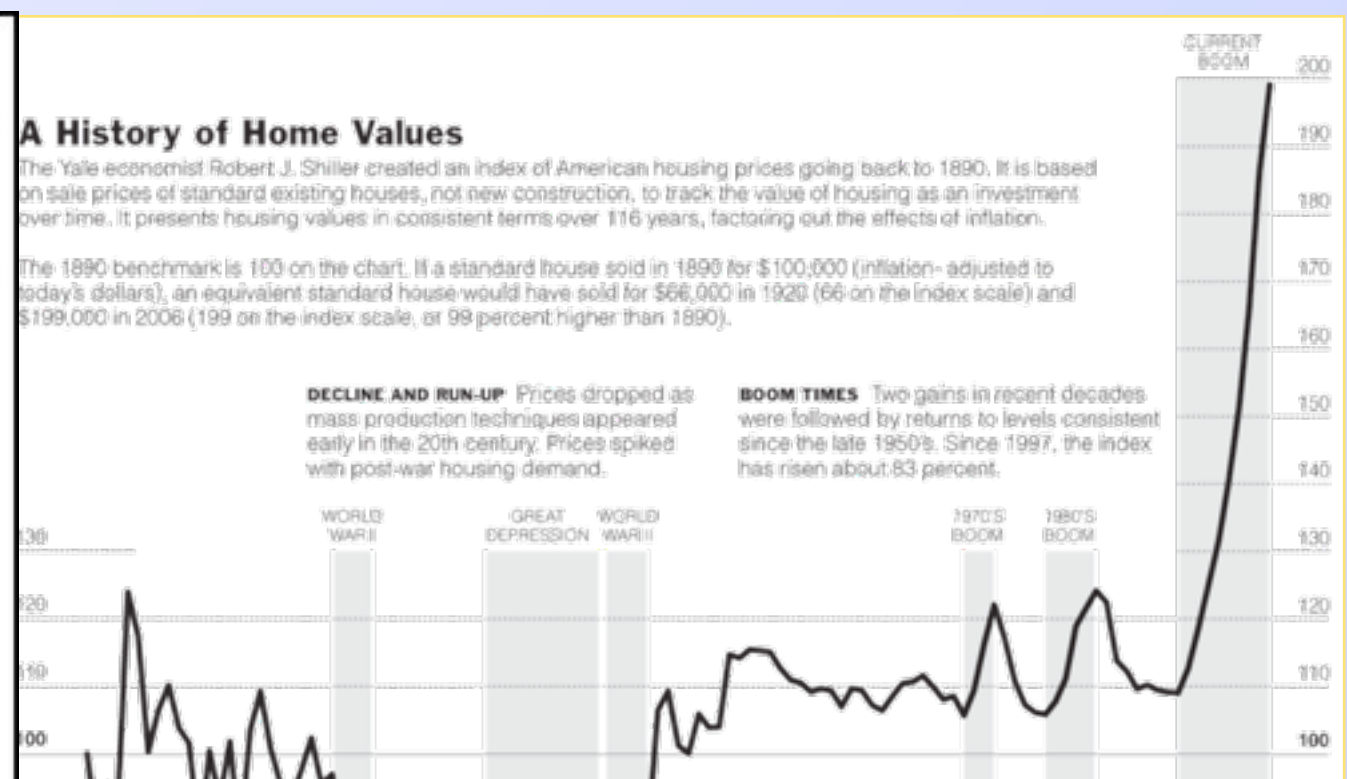
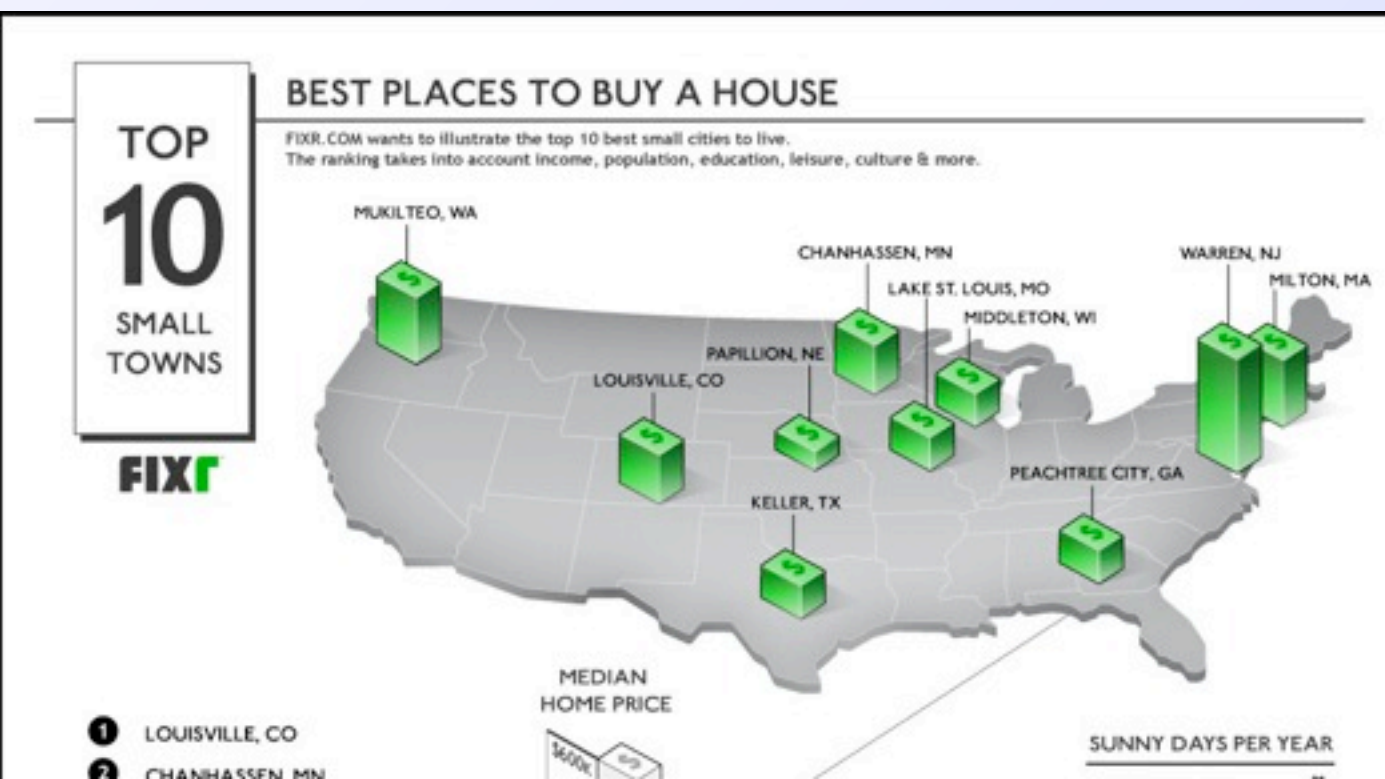


Hedge Fund Solutions

Blackstone Alternative Asset Management

# Scenario ④: Hedge Fund

- house price index
  - published in regular intervals by national statistics agency
  - affects share values of various industries
- **hedge fund:**
  - online market intelligence to **predict** the house **price** index





# Scenario ⑤: Construction

- **tenders** from all over the world
  - existing aggregators
    - expensive, often incomplete
  - yet need to be published (online) by law in most countries



GlobalTenders.com - A premium source of international business tenders

GlobalTenders

Home / Categories / Search / Publish Tenders / Directory / TP

Notices, national and international competitive bids, contracts, RFPs, RFQs

Website Marketing  
Effective SEO  
Improve Search Engine Ranking

**SUSHI COUPONS\***  
in Oxford  
\*Example of an upcoming Deal

**REGION-WISE TENDERS**

- African Tenders
- American Tenders
- Asian Tenders
- Australian & Oceania
- Caribbean Tenders

**Global Tenders - world's latest tender information**

Global Tenders has one of the largest

All Tenders Live Tenders Closed Tenders				
Within Search : <input type="text"/> --Select-- Search				
No.	Tender ID & Tender Brief	State	Tender Value	Due Date
1	TR 2104776 (I) Improvements to Road from Tayakanahalli to Chowtayanahatti (KM 0.00 to 2.50) II) Improvements to Road from Kadekola to Nelabommanahalli (KM 0.00 to 2.20) in Kudligi Taluk. (Package No.: BLR-KDL-I)	KARNATAKA	1,21,00,000.00	15/11/2011
2	TR 2386052 Repair and Maintenance of Main Road from Badam Pahar Corner to Jhamadih CISF Colony Gate at CTPS.			
3	TR 2377019 Request for Qualification for Short-Listing of Concessionaire for Development and Operation of Basti-Mehndawal-Tamkuhi Road Section on PPP basis.			
4	TR 2391770 Improvement and Widening to Shetphal Medha Papans Road MDR-18 km 0/00 to 28/00 Tal. Madha.			
5	TR 2396806 Improvements to Sangola-Pandharpur-Shetphal-Kurduwadi-Paranda Rd.Km. 54/00 to 55/500 (Pro.BM,Carpet,Sealcoat)Tal.Mohol Dist.Solapur.			
6	TR 2396810 Improvements to Pandharpur-Shetphal-Kurduwadi Rd.Km.63/00 to 65/500(Pro.Carpet,sealcoat).			
7	TR 2396815 Improvements to Pandharpur Shetphal Kurduwadi Road Km.54/00 to 57/00 (Pro.Carpet,Sealcoat)Tal.Mohol Dist.Solapur.			
8	TR 2396817 Improvements to Pandharpur-Kurul Road Km.24/00 to 25/700 (Pro.BM,Carpet,Sealcoat)Tal.Mohol Dist.Solapur.			

## International Tenders

Start today to get instant access to the latest [International Tenders](#).

We deliver thousands of live Tender Notices every month. You'll receive regular Tenders from a wide range of companies and public sector organisations.

Start to enjoy this invaluable source of new sales opportunities today with our [4 weeks FREE Trial](#).

**FREE Trial**  
for 4 weeks  
[click here](#)

**CONTINUE**



# Data Extraction



# Goal:

## Domain database



Whole Domain  
Single schema  
Rich attributes



# What we have:

**Product  
provider**



Single agency  
Few attributes



# What we have:

**Product  
provider**



Single agency  
Few attributes





# What we have:

**Product  
provider**



Single agency  
Few attributes





# What we have:

**Product  
provider**



Single agency  
Few attributes





# What we have:

**Product  
provider**



Single agency  
Few attributes





# What we have:

**Product  
provider**



Single agency  
Few attributes



**>15000** in the UK alone



# Aggregators vs. Long Tail

- For many object attributes (phones, ISBN, Homepage)
- significant long tail effect
  - >1000 sites to get above 80% coverage
  - long tail objects **more** valuable to users than aggregator objects
    - measured by demand (number of visits) vs. existing reviews
    - high value to a small amount of users

*for many kinds of information one may have to extract from thousands of sites in order to build a comprehensive database, even when we restrict to a given domain with known popular top sites*



**Product  
provider**



Single agency  
Few attributes



scottfraser



**>15000** in the UK alone



# Product provider

Structured API  
(XML/JSON)

Semantic API  
(RDF)



.....>  
template

HTML  
interface

The screenshot shows a real estate search interface with a purple background. At the top, there are tabs for 'Search' (selected), 'For sale', 'To rent', 'Research', 'Current values', and 'Sold prices'. Below the tabs is a search input field containing 'e.g. Oxford or NW3'. There are several filter sections: 'Min price' (No min), 'Max price' (No max), 'Type' (Show all), and 'Bedrooms' (No min). A 'Distance from location' dropdown is set to 'This area only'. There are also 'Added' (Any time) and 'Sort by' (Highest price) dropdowns. A 'Keywords' input field is present. Under the 'Include' section, there are four checked checkboxes: 'New homes', 'Shared ownership homes', 'Retirement homes', and 'Under offer or sold STC'. A red 'Search for sale' button is on the right. At the bottom right, there is a link 'Browse by area'.





# Product provider

Structured API  
(XML/JSON)

Semantic API  
(RDF)



.....>  
template

HTML  
interface

The screenshot shows a real estate search interface with a purple background. At the top, there are tabs for 'Search' (with sub-tabs 'For sale' and 'To rent') and 'Research' (with sub-tabs 'Current values' and 'Sold prices'). Below these is a search input field containing 'e.g. Oxford or NW3'. Further down are filters for 'Min price' (No min), 'Max price' (No max), 'Type' (Show all), and 'Bedrooms' (No min), with a 'Fewer options' link. There is also a 'Distance from location' dropdown set to 'This area only', an 'Added' dropdown set to 'Any time', and a 'Sort by' dropdown set to 'Highest price'. A 'Keywords' input field is present. Under the 'Include' section, there are four checked checkboxes: 'New homes', 'Shared ownership homes', 'Retirement homes', and 'Under offer or sold STC'. An orange 'Search for sale' button is on the right. At the bottom right, there is a 'Browse by area' link.





HTML  
interface

Search **For sale** To rent Research **Current values** Sold prices

e.g. Oxford or NW3

Min price No min Max price No max Type Show all Bedrooms No min [Fewer options](#)

Distance from location This area only

Added Any time Sort by Highest price

Keywords

Include

- ☒ New homes
- ☒ Shared ownership homes
- ☒ Retirement homes
- ☒ Under offer or sold STC

[Search for sale](#)

[Browse by area](#)

Form filling

**WHITE WALLS**  
people not just property

Home About us Sales Lettings/Management Free valuation EPC Careers Links Contact

## Property in Oxford search results

[Go back to postcode area map](#) 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 [NEXT](#)

**White Road, East Hendred** **£1,499,000**

Located within East Hendred, this impressive fifteen bedroom property offers fantastic business and development potential. A five bedroom family house is accompanied by a purpose built ten bedroom guest house extension. The five [...] [Read more](#)

Bedrooms: **15** | Bathrooms: **13** | Area: **East Hendred**

[View all 10 photos](#)

[View full details](#) [360° virtual tour](#) [Arrange a viewing](#) [Map view](#)

**Bagley Wood Rd, Kennington OX1** **£1,295,000**

A substantial detached Edwardian family house with scope for updating and improvement, set within delightful gardens and wooded grounds extending to over five acres located on the edge of Bagley Wood. This fine [...] [Read more](#)

Bedrooms: **6** | Bathrooms: **3** | Area: **Kennington**

[View all 16 photos](#)

[View full details](#) [360° virtual tour](#) [Arrange a viewing](#) [Map view](#)



HTML  
interface

Search For sale To rent Research Current values Sold prices

e.g. Oxford or NW3

Min price No min Max price No max Type Show all Bedrooms No min Fewer options

Distance from location This area only

Added Any time Sort by Highest price

Keywords

Include

☒ New homes ☒ Shared ownership homes

☒ Retirement homes ☒ Under offer or sold STC

Search for sale

Browse by area

Form filling

WHITE WALLS people not just property

Home About us Sales Lettings/Management Free valuation EPC Careers Links Contact

Property in Oxford search results

« Go back to postcode area map 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 NEXT 1 -

**White Road, East Hendred** £1,499,000 New price

Located within East Hendred, this impressive fifteen bedroom property offers fantastic business and development potential. A five bedroom family house is accompanied by a purpose built ten bedroom guest house extension. The five [...] [Read more](#)

Bedrooms: 15 | Bathrooms: 13 | Area: East Hendred

[View all 10 photos](#) [View full details](#) [360° virtual tour](#) [Arrange a viewing](#) [Map view](#)


**Bagley Wood Rd, Kennington OX1** £1,295,000

A substantial detached Edwardian family house with scope for updating and improvement, set within delightful gardens and wooded grounds extending to over five acres located on the edge of Bagley Wood. This fine [...] [Read more](#)

Bedrooms: 6 | Bathrooms: 3 | Area: Kennington

[View all 16 photos](#) [View full details](#) [360° virtual tour](#) [Arrange a viewing](#) [Map view](#)





Bagley Wood. This fine [...] [Read more](#)

Bedrooms: 6 | Bathrooms: 3 | Area: **Kennington**


[View all 16 photos](#)

[View full details](#)
[360° virtual tour](#)
[Arrange a viewing](#)
[Map view](#)

Object  
identification

## White Road, East Hendred

£1,499,950



[Arrange a viewing](#)
[Image gallery](#)
[360° virtual tour](#)
[Floorplan](#)
[Aerial view](#)
[Map view](#)
[Printable brochure](#)
[Email a friend](#)
[EPC](#)
[Free mortgage advice](#)


Sales or Lettings: <b>Sales</b>	Property Type: <b>House</b>	Postcode Area: <b>OX12</b>
Bedrooms: <b>15</b>	Bathrooms: <b>13</b>	Furnishing: <b>N/A</b>
Area: <b>East Hendred</b>	Tenure: <b>Freehold</b>	Sq feet: <b>6648</b>

Located within East Hendred, this impressive fifteen bedroom property offers fantastic business and development potential. A five bedroom family house is accompanied by a purpose built ten bedroom guest house extension.

The five bedroom family house benefits from generously proportioned living space with modern interiors, conservatory, dining room with feature bow window and stunning first floor views over substantial private gardens. This is further accompanied an attached recently purpose built AA 4\* rated ten bedroom guest house/bed and breakfast wing.

In addition, subject to planning, further development potential is possible.


The property is situated within its own grounds of approximately 2 acres, Local transport links are available nearby, providing access to Oxford and Wantage. Road users have easy access to the A34 (M40) towards London.



**0%** Coming soon to OX2

[Pre-register now](#)





Bagley Wood. This fine [...] [Read more](#)

Bedrooms: 6 | Bathrooms: 3 | Area: Kennington

[View all 16 photos](#)

[View full details](#)
[360° virtual tour](#)
[Arrange a viewing](#)
[Map view](#)

Object identification

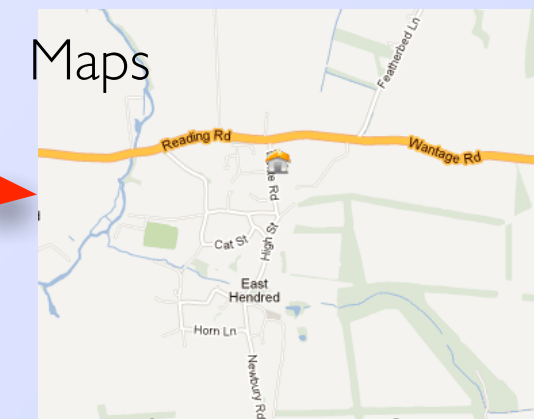
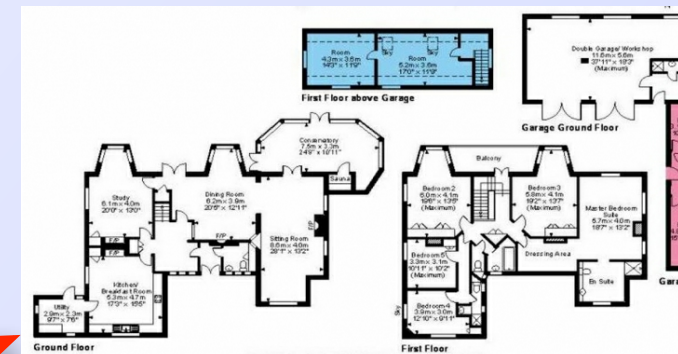
## White Road, East Hendred

3

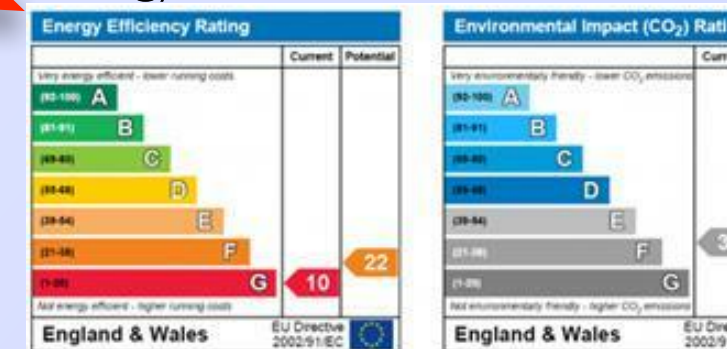
£1,499,950



- [Arrange a viewing](#)
- [Image gallery](#)
- [360° virtual tour](#)
- [3D Floorplan](#)
- [Aerial view](#)
- [Map view](#)
- [Printable brochure](#)
- [Email a friend](#)
- [EPC](#)
- [Free mortgage advice](#)



Energy Performance Chart



Tables

Sales or Lettings: <b>Sales</b>	Property Type: <b>House</b>	Postcode Area: <b>OX12</b>
Bedrooms: <b>15</b>	Bathrooms: <b>13</b>	Furnishing: <b>N/A</b>
Area: <b>East Hendred</b>	Tenure: <b>Freehold</b>	Sq feet: <b>6648</b>

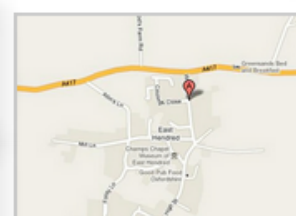
Flat Text

Located within East Hendred, this impressive fifteen bedroom property offers fantastic business and development potential. A five bedroom family house is accompanied by a purpose built ten bedroom guest house extension.

The five bedroom family house benefits from generously proportioned living space with modern interiors, conservatory, dining room with feature bow window and stunning first floor views over substantial private gardens. This is further accompanied an attached recently purpose built AA 4\* rated ten bedroom guest house/bed and breakfast wing.

In addition, subject to planning, further development potential is possible.

The property is situated within its own grounds of approximately 2 acres, Local transport links are available nearby, providing access to Oxford and Wantage. Road users have easy access to the A34 (M40) towards London.



0% Coming soon to OX2

[Pre-register now](#)



Product provider

Structured API (XML/JSON)

Semantic API (RDF)



template

HTML interface

Search For sale To rent Research Current values Sold prices

e.g. Oxford or NW3

Min price No min Max price No max Type Show all Bedrooms No min Fewer options

Distance from location This area only

Added Any time Sort by Highest price

Keywords

Include ☒ New homes ☒ Shared ownership homes ☒ Retirement homes ☒ Under offer or sold STC

Search for sale

Form filling

WHITE WALLS people not just property

Home About us Sales Lettings/Management Free valuation EPC Careers Links Contact us

Property in Oxford search results

Go back to postcode area map 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 NEXT 1 - 10 of properties

**White Road, East Hendred** £1,499,950 New price

Located within East Hendred, this impressive fifteen bedroom property offers fantastic business and development potential. A five bedroom family house is accompanied by a purpose built ten bedroom guest house extension. The five [...] Read more

Bedrooms: 15 | Bathrooms: 13 | Area: East Hendred

View all 10 photos View full details 360° virtual tour Arrange a viewing Map view

**Bagley Wood Rd, Kennington OX1** £1,295,000 Sold

A substantial detached Edwardian family house with scope for updating and improvement, set within delightful gardens and wooded grounds extending to over five acres located on the edge of Bagley Wood. This fine [...] Read more

Bedrooms: 6 | Bathrooms: 3 | Area: Kennington

View all 16 photos View full details 360° virtual tour Arrange a viewing Map view

Domain database



Cleaning & integration

4

Tables

Flat Text

White Road, East Hendred 3 £1,499,950

Arrange a viewing Image gallery 360° virtual tour Floorplan Aerial view Map view Printable brochure Email a friend EPC Free mortgage advice

Sales or Lettings: Sales	Property Type: House	Postcode Area: OX12
Bedrooms: 15	Bathrooms: 13	Furnishing: N/A
Area: East Hendred	Tenure: Freehold	Sq feet: 6648

Located within East Hendred, this impressive fifteen bedroom property offers fantastic business and development potential. A five bedroom family house is accompanied by a purpose built ten bedroom guest house extension.

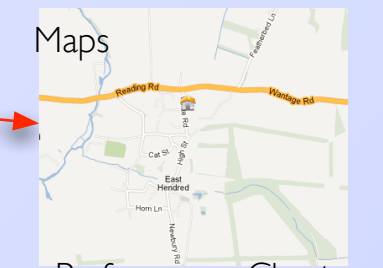
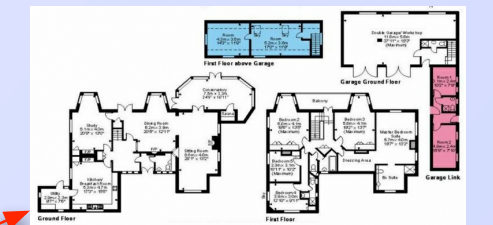
The five bedroom family house benefits from generously proportioned living space with modern interiors, conservatory, dining room with feature bow window and stunning first floor views over substantial private gardens. This is further accompanied an attached recently purpose built AA 4\* rated ten bedroom guest house/bed and breakfast wing.

In addition, subject to planning, further development potential is possible.

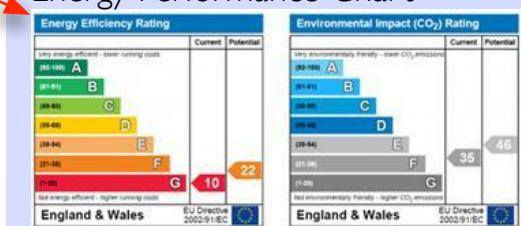
The property is situated within its own grounds of approximately 2 acres, Local transport links are available nearby, providing access to Oxford and Wantage. Road users have easy access to the A34 (M40) towards London.

Coming soon to OX2 Pre-register now

Object identification



Energy Performance Chart





Product provider

Structured API (XML/JSON)

Semantic API (RDF)



template

HTML interface

Search For sale To rent Research Current values Sold prices

e.g. Oxford or NW3

Min price No min Max price No max Type Show all Bedrooms No min Fewer options

Distance from location This area only

Added Any time Sort by Highest price

Keywords

Include New homes Shared ownership homes Retirement homes Under offer or sold STC

Search for sale

Form filling

WHITE WALLS people not just property

Home About us Sales Lettings/Management Free valuation EPC Careers Links Contact us

Property in Oxford search results

Go back to postcode area map 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 NEXT 1 - 10 of properties

**White Road, East Hendred** £1,499,950 New price

Located within East Hendred, this impressive fifteen bedroom property offers fantastic business and development potential. A five bedroom family house is accompanied by a purpose built ten bedroom guest house extension. The five [...] Read more

Bedrooms: 15 | Bathrooms: 13 | Area: East Hendred

View all 10 photos View full details 360° virtual tour Arrange a viewing Map view

**Bagley Wood Rd, Kennington OX1** £1,295,000 Sold

A substantial detached Edwardian family house with scope for updating and improvement, set within delightful gardens and wooded grounds extending to over five acres located on the edge of Bagley Wood. This fine [...] Read more

Bedrooms: 6 | Bathrooms: 3 | Area: Kennington

View all 16 photos View full details 360° virtual tour Arrange a viewing Map view

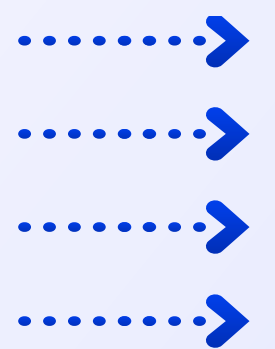
Domain database

Other providers



Cleaning & integration

4



Other Provider

Other Provider

Other Provider

Object identification

White Road, East Hendred 3 £1,499,950

Arrange a viewing Image gallery 360° virtual tour Floorplan Aerial view Map view Printable brochure Email a friend EPC Free mortgage advice

Sales or Lettings: Sales Property Type: House Postcode Area: OX12 Bedrooms: 15 Bathrooms: 13 Furnishing: N/A Area: East Hendred Tenure: Freehold Sq feet: 6648

Located within East Hendred, this impressive fifteen bedroom property offers fantastic business and development potential. A five bedroom family house is accompanied by a purpose built ten bedroom guest house extension.

The five bedroom family house benefits from generously proportioned living space with modern interiors, conservatory, dining room with feature bow window and stunning first floor views over substantial private gardens. This is further accompanied an attached recently purpose built AA 4\* rated ten bedroom guest house/bed and breakfast wing.

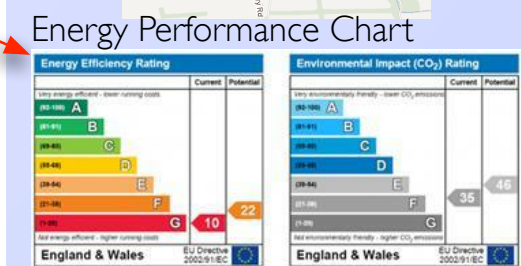
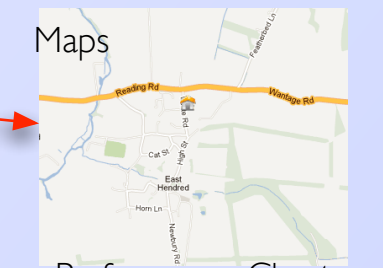
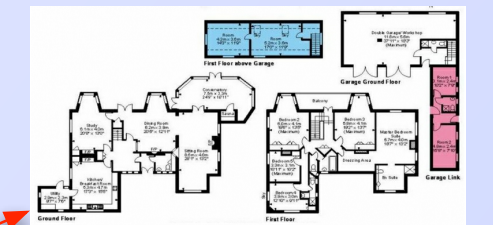
In addition, subject to planning, further development potential is possible.

The property is situated within its own grounds of approximately 2 acres, Local transport links are available nearby, providing access to Oxford and Wantage. Road users have easy access to the A34 (M40) towards London.

Coming soon to OX2 Pre-register now

Tables

Flat Text





# 4 Steps to Data Extraction

- **full-fledged data extraction system** covers
  - web **navigation and interaction**
  - **wrapper induction** or generation and **execution**
  - **scheduling**
  - data **transformation** and **integration**
- in this tutorial, we focus on the first two points
  - but if you are interested in the others
    - ask at the Lixto demo



## Lectures

- [Data Models and Query Languages for Linked Geospatial Data](#) (Manolis Koubarakis, Manos Karpathiotakis, Kostis Kyzirakos, Babis Nikolaou, [Michael Sioutis](#))
- [Semantic Wikis: Approaches, Applications, and Perspectives](#) (François Bry, [Sebastian Schaffert](#), [Denny Vrandečić](#), [Klara Weiland](#))
- [OWL 2 Profiles: An Introduction to Lightweight Ontology Languages](#) (Markus Krötzsch)
- [Argumentation and the Web](#) (Francesca Toni)
- [Federation and Navigation in SPARQL 1.1](#) (Marcelo Arenas, [Jorge Pérez](#))
- [Reasoning with Uncertain and Inconsistent Ontologies on the Semantic Web](#) (Guilin Qi, [Jianfeng Du](#))
- [Linked Data Stream Processing](#) (Manfred Hauswirth, [Danh Le Phuoc](#), [Josiane Xavier Parreira](#))
- [Datalog and Its Extensions for Semantic Web Databases](#) (Georg Gottlob, [Giorgio Orsi](#), [Andreas Pieris](#), [Mantas Šimkus](#))
- [Reasoning and Query Answering in Description Logics](#) (Magdalena Ortiz, [Mantas Šimkus](#))
- [Reasoning and Ontologies in Data Extraction](#) (Sergio Flesca, [Tim Furche](#), [Ermelinda Oro](#))

## Program

SEPTEMBER 02 (SUNDAY)			
Time	What	Who	Where
16:30 - 18:00	Summer School Registration and Poster Setup		<a href="#">Info Desk and Zeichensaal 13&amp;14, main building, TU Vienna</a>
SEPTEMBER 03 (MONDAY)			
Time	What	Who	Where
08:00 - 09:00	Summer School Registration and Poster Setup		<a href="#">Info Desk and Zeichensaal 13&amp;14, main building, TU Vienna</a>
09:00 - 09:15	Summer School Opening		
09:15 - 10:30	<a href="#">Reasoning and Query Answering in Description Logics I</a>	<a href="#">Magdalena Ortiz</a> , <a href="#">Mantas Šimkus</a>	<a href="#">Lecture room HS 8, main building, TU Vienna</a>
10:30	Coffee break		



### Platinum Sponsors





Lectures

- [Data Models and Query Languages for Linked Geospatial Data](#) (Manolis Koubarakis, Manos Karpathiotakis, Kostis Kyzirakos, Babis Nikolaou, [Michael Sioutis](#))
- [Semantic Wikis: Approaches, Applications, and Perspectives](#) (François Bry, Sebastian Schaffert, Denny Vrandečić, Klara Weiland)
- [OWL 2 Profiles: An Introduction to Lightweight Ontology Languages](#) (Markus Krötzsch)
- [Argumentation and the Web](#) (Francesca Toni)
- [Federation and Navigation in SPARQL 1.1](#) (Marcelo Arenas, Jorge Pérez)
- [Reasoning with Uncertain and Inconsistent Ontologies on the Semantic Web](#) (Guilin Qi, Jianfeng Du)
- [Linked Data Stream Processing](#) (Manfred Hauswirth, Danh Le Phuoc, Josiane Xavier Parreira)
- [Datalog and Its Extensions for Semantic Web Databases](#) (Georg Gottlob, Giorgio Orsi, Andreas Pieris, Mantas Šimkus)
- [Reasoning and Query Answering in Description Logics](#) (Magdalena Ortiz, Mantas Šimkus)
- [Reasoning and Ontologies in Data Extraction](#) (Sergio Flesca, Tim Furche, Ermelinda Oro)

Program

SEPTEMBER 02 (SUNDAY)			
Time	What	Who	Where
16:30 - 18:00	Summer School Registration and Poster Setup		<a href="#">Info Desk and Zeichensaal 13&amp;14, main building, TU Vienna</a>

SEPTEMBER 03 (MONDAY)			
Time	What	Who	Where
08:00 - 09:00	Summer School Registration and Poster Setup		<a href="#">Info Desk and Zeichensaal 13&amp;14, main building, TU Vienna</a>
09:00 - 09:15	Summer School Opening		
09:15 - 10:30	<a href="#">Reasoning and Query Answering in Description Logics I</a>	<a href="#">Magdalena Ortiz, Mantas Šimkus</a>	<a href="#">Lecture room HS 8, main building, TU Vienna</a>
10:30 - 11:00	Coffee break		
11:00 - 12:30	<a href="#">Reasoning and Query Answering in Description Logics II</a>	<a href="#">Magdalena Ortiz, Mantas Šimkus</a>	
12:30 - 14:30	Lunch break		Vienna
14:30 - 16:00	<a href="#">Datalog and Its Extensions for Semantic Web Databases I</a>	<a href="#">Georg Gottlob, Giorgio Orsi, Andreas Pieris, Mantas Šimkus</a>	<a href="#">Lecture room HS 8, main building, TU Vienna</a>
16:00 - 16:30	Coffee break		
16:30 - 18:00	<a href="#">Datalog and Its Extensions for Semantic Web Databases II</a>	<a href="#">Georg Gottlob, Giorgio Orsi, Andreas Pieris, Mantas Šimkus</a>	

SEPTEMBER 04 (TUESDAY)			
Time	What	Who	Where
09:00 - 10:30	<a href="#">Semantic Wikis: Approaches, Applications, and Perspectives I</a>	<a href="#">François Bry, Sebastian Schaffert, Denny Vrandečić, Klara Weiland</a>	<a href="#">Lecture room HS 8, main building, TU Vienna</a>
10:30 - 11:00	Coffee break		
11:00 - 12:30	<a href="#">Semantic Wikis: Approaches, Applications, and Perspectives II</a>	<a href="#">François Bry, Sebastian Schaffert, Denny Vrandečić, Klara Weiland</a>	
12:30 - 14:30	Lunch break		Vienna
14:30 - 16:00	<a href="#">Semantic Wikis: Approaches, Applications, and Perspectives III</a>	<a href="#">François Bry, Sebastian Schaffert, Denny Vrandečić, Klara Weiland</a>	<a href="#">Lecture room HS 8, main building, TU Vienna</a>



Platinum Sponsors



Silver Sponsors



Hosted by





Lectures

- [Data Models and Query Languages for Linked Geospatial Data](#) (Manolis Koubarakis, Manos Karpathiotakis, Kostis Kyzirakos, Babis Nikolaou, [Michael Sioutis](#))
- [Semantic Wikis: Approaches, Applications, and Perspectives](#) (François Bry, Sebastian Schaffert, Denny Vrandečić, Klara Weiland)
- [OWL 2 Profiles: An Introduction to Lightweight Ontology Languages](#) (Markus Krötzsch)
- [Argumentation and the Web](#) (Francesca Toni)
- [Federation and Navigation in SPARQL 1.1](#) (Marcelo Arenas, Jorge Pérez)
- [Reasoning with Uncertain and Inconsistent Ontologies on the Semantic Web](#) (Guilin Qi, Jianfeng Du)
- [Linked Data Stream Processing](#) (Manfred Hauswirth, Danh Le Phuoc, Josiane Xavier Parreira)
- [Datalog and Its Extensions for Semantic Web Databases](#) (Georg Gottlob, Giorgio Orsi, Andreas Pieris, Mantas Šimkus)
- [Reasoning and Query Answering in Description Logics](#) (Magdalena Ortiz, Mantas Šimkus)
- [Reasoning and Ontologies in Data Extraction](#) (Sergio Flesca, Tim Furche, Ermelinda Oro)

Program

SEPTEMBER 02 (SUNDAY)	
Time	What
16:30 - 18:00	Summer School Registration and Poster Setup
SEPTEMBER 03 (MONDAY)	
Time	What
08:00 - 09:00	Summer School Registration and Poster Setup
09:00 - 09:15	Summer School Opening
09:15 - 10:30	<a href="#">Reasoning and Query Answering in Description Logics I</a>
10:30 - 11:00	Coffee break
11:00 - 12:30	<a href="#">Reasoning and Query Answering in Description Logics II</a>
12:30 - 14:30	Lunch break
14:30 - 16:00	<a href="#">Datalog and Its Extensions for Semantic Web Databases I</a>
16:00 - 16:30	Coffee break
16:30 - 18:00	<a href="#">Datalog and Its Extensions for Semantic Web Databases II</a>
SEPTEMBER 04 (TUESDAY)	
Time	What
09:00 - 10:30	<a href="#">Semantic Wikis: Approaches, Applications, and Perspectives I</a>
10:30 - 11:00	Coffee break
11:00 - 12:30	<a href="#">Semantic Wikis: Approaches, Applications, and Perspectives II</a>
12:30 - 14:30	Lunch break
14:30 - 16:00	<a href="#">Semantic Wikis: Approaches, Applications, and Perspectives III</a>



# Structured Data

Time	Person		Room
09:15–12:30	Magdalena Ortiz	Reasoning and Query Answering in Description Logics	HS 8
09:15–12:30	Mantas Šimkus	Reasoning and Query Answering in Description Logics	HS 8
14:30–18:00	Georg Gottlob	Datalog and Its Extensions for Semantic Web Databases	HS 8
14:30–18:00	Giorgio Orsi	Datalog and Its Extensions for Semantic Web Databases	HS 8
14:30–18:00	Andreas Pieris	Datalog and Its Extensions for Semantic Web Databases	HS 8
14:30–18:00	Mantas Šimkus	Datalog and Its Extensions for Semantic Web Databases	HS 8
09:00-12:30	François Bry	Semantic Wikis: Approaches, Applications, and Perspectives	HS 8

...

11:00 - 12:30	<a href="#">Semantic Wikis: Approaches, Applications, and Perspectives II</a>	<a href="#">François Bry</a> , <a href="#">Sebastian Schaffert</a> , <a href="#">Denny Vrandečić</a> , <a href="#">Klara Weiland</a>
12:30 - 14:30	Lunch break	Vienna
14:30 - 16:00	<a href="#">Semantic Wikis: Approaches, Applications, and Perspectives III</a>	<a href="#">François Bry</a> , <a href="#">Sebastian Schaffert</a> , <a href="#">Denny Vrandečić</a> , <a href="#">Klara Weiland</a>

TO  
WIEN

kbs

Lecture room HS 8, main building.



Lectures

- [Data Models and Query Languages for Linked Geospatial Data](#) (Manolis Koubarakis, Manos Karpathiotakis, Kostis Kyzirakos, Babis Nikolaou, [Michael Sioutis](#))
- [Semantic Wikis: Approaches, Applications, and Perspectives](#) (François Bry, Sebastian Schaffert, Denny Vrandečić, Klara Weiland)
- [OWL 2 Profiles: An Introduction to Lightweight Ontology Languages](#) (Markus Krötzsch)
- [Argumentation and the Web](#) (Francesca Toni)
- [Federation and Navigation in SPARQL 1.1](#) (Marcelo Arenas, Jorge Pérez)
- [Reasoning with Uncertain and Inconsistent Ontologies on the Semantic Web](#) (Guilin Qi, Jianfeng Du)
- [Linked Data Stream Processing](#) (Manfred Hauswirth, Danh Le Phuoc, Josiane Xavier Parreira)
- [Datalog and Its Extensions for Semantic Web Databases](#) (Georg Gottlob, Giorgio Orsi, Andreas Pieris, Mantas Šimkus)
- [Reasoning and Query Answering in Description Logics](#) (Magdalena Ortiz, Mantas Šimkus)
- [Reasoning and Ontologies in Data Extraction](#) (Sergio Flesca, Tim Furche, Ermelinda Oro)

Program

SEPTEMBER 02 (SUNDAY)	
Time	What
16:30 - 18:00	Summer School Registration and Poster Setup

SEPTEMBER 03 (MONDAY)	
Time	What
08:00 - 09:00	Summer School Registration and Poster Setup
09:00 - 09:15	Summer School Opening
09:15 - 10:30	<a href="#">Reasoning and Query Answering in Description Logics I</a>
10:30 - 11:00	Coffee break
11:00 - 12:30	<a href="#">Reasoning and Query Answering in Description Logics II</a>
12:30 - 14:30	Lunch break
14:30 - 16:00	<a href="#">Datalog and Its Extensions for Semantic Web Databases I</a>
16:00 - 16:30	Coffee break
16:30 - 18:00	<a href="#">Datalog and Its Extensions for Semantic Web Databases II</a>

SEPTEMBER 04 (TUESDAY)	
Time	What
09:00 - 10:30	<a href="#">Semantic Wikis: Approaches, Applications, and Perspectives I</a>
10:30 - 11:00	Coffee break

11:00 - 12:30	<a href="#">Semantic Wikis: Approaches, Applications, and Perspectives II</a>	<a href="#">François Bry, Sebastian Schaffert, Denny Vrandečić, Klara Weiland</a>
12:30 - 14:30	Lunch break	Vienna
14:30 - 16:00	<a href="#">Semantic Wikis: Approaches, Applications, and Perspectives II</a>	<a href="#">François Bry, Sebastian Schaffert, Denny Vrandečić, Klara Weiland</a>



# Structured Data

Time	Person		Room
09:15–12:30	Magdalena Ortiz	Reasoning and Query Answering in Description Logics	HS 8
09:15–12:30	Mantas Šimkus	Reasoning and Query Answering in Description Logics	HS 8
14:30–18:00	Georg Gottlob	Datalog and Its Extensions for Semantic Web Databases	HS 8
14:30–18:00	Giorgio Orsi	Datalog and Its Extensions for Semantic Web Databases	HS 8
14:30–18:00	Andreas Pieris	Datalog and Its Extensions for Semantic Web Databases	HS 8
14:30–18:00	Mantas Šimkus	Datalog and Its Extensions for Semantic Web Databases	HS 8
09:00-12:30	François Bry	Semantic Wikis: Approaches, Applications, and Perspectives	HS 8

...

≈ Data Extraction proper

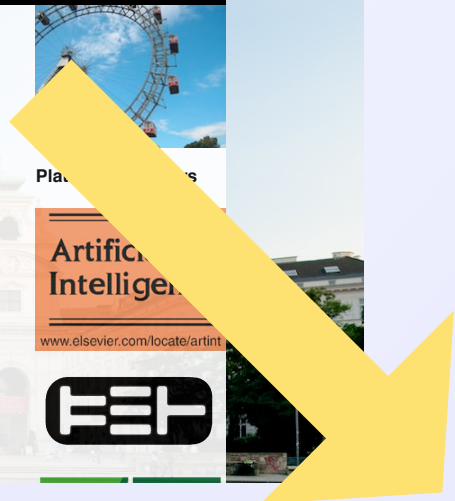


Lectures

- [Data Models and Query Languages for Linked Geospatial Data](#) (Manolis Koubarakis, Manos Karpathiotakis, Kostas Kyzirakos, Babis Nikolaou, [Michael Sioutis](#))
- [Semantic Wikis: Approaches, Applications, and Perspectives](#) (François Bry, Sebastian Schaffert, Denny Vrandečić, Klara Weiland)
- [OWL 2 Profiles: An Introduction to Lightweight Ontology Languages](#) (Markus Krötzsch)
- [Argumentation and the Web](#) (Francesca Toni)
- [Federation and Navigation in SPARQL 1.1](#) (Marcelo Arenas, Jorge Pérez)
- [Reasoning with Uncertain and Inconsistent Ontologies on the Semantic Web](#) (Guilin Qi, Jianfeng Du)
- [Linked Data Stream Processing](#) (Manfred Hauswirth, Danh Le Phuoc, Josiane Xavier Parreira)
- [Datalog and Its Extensions for Semantic Web Databases](#) (Georg Gottlob, Giorgio Orsi, Andreas Pieris, Mantas Šimkus)
- [Reasoning and Query Answering in Description Logics](#) (Magdalena Ortiz, Mantas Šimkus)
- [Reasoning and Ontologies in Data Extraction](#) (Sergio Flesca, Tim Furche, Ermelinda Oro)

Program

SEPTEMBER 02 (SUNDAY)	
Time	What
16:30 - 18:00	Summer School Registration and Poster Setup
SEPTEMBER 03 (MONDAY)	
Time	What
08:00 - 09:00	Summer School Registration and Poster Setup
09:00 - 09:15	Summer School Opening
09:15 - 10:30	<a href="#">Reasoning and Query Answering in Description Logics I</a>
10:30 - 11:00	Coffee break
11:00 - 12:30	<a href="#">Reasoning and Query Answering in Description Logics II</a>
12:30 - 14:30	Lunch break
14:30 - 16:00	<a href="#">Datalog and Its Extensions for Semantic Web Databases I</a>
16:00 - 16:30	Coffee break
16:30 - 18:00	<a href="#">Datalog and Its Extensions for Semantic Web Databases II</a>
SEPTEMBER 04 (TUESDAY)	
Time	What
09:00 - 10:30	<a href="#">Semantic Wikis: Approaches, Applications, and Perspectives I</a>
10:30 - 11:00	Coffee break
11:00 - 12:30	<a href="#">Semantic Wikis: Approaches, Applications, and Perspectives II</a>
12:30 - 14:30	Lunch break
14:30 - 15:00	<a href="#">Semantic Wikis: Approaches, Applications, and Perspectives III</a>



# Structured Data

Time	Person		Room
09:15–12:30	Magdalena Ortiz	Reasoning and Query Answering in Description Logics	HS 8
09:15–12:30	Mantas Šimkus	Reasoning and Query Answering in Description Logics	HS 8
14:30–18:00	Georg Gottlob	Datalog and Its Extensions for Semantic Web Databases	HS 8
14:30–18:00	Giorgio Orsi	Datalog and Its Extensions for Semantic Web Databases	HS 8
14:30–18:00	Andreas Pieris	Datalog and Its Extensions for Semantic Web Databases	HS 8
14:30–18:00	Mantas Šimkus	Datalog and Its Extensions for Semantic Web Databases	HS 8
09:00-12:30	François Bry	Semantic Wikis: Approaches, Applications, and Perspectives	HS 8

...

≈ **Data Extraction proper**  
≈ Table Extraction

TOUWIEN

kbs

11:00

[Semantic Wikis: Approaches, Applications, and Perspectives II](#)

[François Bry, Sebastian Schaffert, Denny Vrandečić, Klara Weiland](#)

12:30

Lunch break

Vienna

14:30

[Semantic Wikis: Approaches, Applications, and Perspectives III](#)

[François Bry, Sebastian Schaffert, Denny Vrandečić, Klara Weiland](#)

[Lecture room HS 8, main building, 8th floor](#)



# Variations I: Output



Lectures

- [Data Models and Query Languages for Linked Geospatial Data](#) (Manolis Koubarakis, Manos Karpathiotakis, Kostis Kyzirakos, Babis Nikolaou, [Michael Sioutis](#))
- [Semantic Wikis: Approaches, Applications, and Perspectives](#) (François Bry, Sebastian Schaffert, Denny Vrandečić, Klara Weiland)
- [OWL 2 Profiles: An Introduction to Lightweight Ontology Languages](#) (Markus Krötzsch)
- [Argumentation and the Web](#) (Francesca Toni)
- [Federation and Navigation in SPARQL 1.1](#) (Marcelo Arenas, Jorge Pérez)
- [Reasoning with Uncertain and Inconsistent Ontologies on the Semantic Web](#) (Guilin Qi, Jianfeng Du)
- [Linked Data Stream Processing](#) (Manfred Hauswirth, Danh Le Phuoc, Josiane Xavier Parreira)
- [Datalog and Its Extensions for Semantic Web Databases](#) (Georg Gottlob, Giorgio Orsi, Andreas Pieris, Mantas Šimkus)
- [Reasoning and Query Answering in Description Logics](#) (Magdalena Ortiz, Mantas Šimkus)
- [Reasoning and Ontologies in Data Extraction](#) (Sergio Flesca, Tim Furrche, Ermelinda Oro)

Program

SEPTEMBER 02 (SUNDAY)			
Time	What	Who	Where
16:30 - 18:00	Summer School Registration and Poster Setup		<a href="#">Info Desk and Zeichensaal 13&amp;14, main building, TU Vienna</a>

SEPTEMBER 03 (MONDAY)			
Time	What	Who	Where
08:00 - 09:00	Summer School Registration and Poster Setup		<a href="#">Info Desk and Zeichensaal 13&amp;14, main building, TU Vienna</a>
09:00 - 09:15	Summer School Opening		
09:15 - 10:30	<a href="#">Reasoning and Query Answering in Description Logics I</a>	<a href="#">Magdalena Ortiz</a> , <a href="#">Mantas Šimkus</a>	<a href="#">Lecture room HS 8, main building, TU Vienna</a>
10:30 - 11:00	Coffee break		
11:00 - 12:30	<a href="#">Reasoning and Query Answering in Description Logics II</a>	<a href="#">Magdalena Ortiz</a> , <a href="#">Mantas Šimkus</a>	
12:30 - 14:30	Lunch break		Vienna
14:30 - 16:00	<a href="#">Datalog and Its Extensions for Semantic Web Databases I</a>	<a href="#">Georg Gottlob</a> , <a href="#">Giorgio Orsi</a> , <a href="#">Andreas Pieris</a> , <a href="#">Mantas Šimkus</a>	<a href="#">Lecture room HS 8, main building, TU Vienna</a>
16:00 - 16:30	Coffee break		
16:30 - 18:00	<a href="#">Datalog and Its Extensions for Semantic Web Databases II</a>	<a href="#">Georg Gottlob</a> , <a href="#">Giorgio Orsi</a> , <a href="#">Andreas Pieris</a> , <a href="#">Mantas Šimkus</a>	

SEPTEMBER 04 (TUESDAY)			
Time	What	Who	Where
09:00 - 10:30	<a href="#">Semantic Wikis: Approaches, Applications, and Perspectives I</a>	<a href="#">François Bry</a> , <a href="#">Sebastian Schaffert</a> , <a href="#">Denny Vrandečić</a> , <a href="#">Klara Weiland</a>	<a href="#">Lecture room HS 8, main building, TU Vienna</a>
10:30 - 11:00	Coffee break		
11:00 - 12:30	<a href="#">Semantic Wikis: Approaches, Applications, and Perspectives II</a>	<a href="#">François Bry</a> , <a href="#">Sebastian Schaffert</a> , <a href="#">Denny Vrandečić</a> , <a href="#">Klara Weiland</a>	
12:30 -	Lunch break		Vienna



Platinum Sponsors



Silver Sponsors



Hosted by



Time	Speaker	Topic	Room
09:15–12:30	Magdalena Ortiz	Reasoning and Query Answering in Description Logics	HS 8
09:15–12:30	Mantas Šimkus	Reasoning and Query Answering in Description Logics	HS 8
14:30–18:00	Georg Gottlob	Datalog and Its Extensions for Semantic Web Databases	HS 8
14:30–18:00	Giorgio Orsi	Datalog and Its Extensions for Semantic Web Databases	HS 8
14:30–18:00	Andreas Pieris	Datalog and Its Extensions for Semantic Web Databases	HS 8
14:30–18:00	Mantas Šimkus	Datalog and Its Extensions for Semantic Web Databases	HS 8
09:00-12:30	François Bry	Semantic Wikis: Approaches, Applications, and Perspectives	HS 8

...



Lectures

- [Data Models and Query Languages for Linked Geospatial Data](#) (Manolis Koubarakis, Manos Karpathiotakis, Kostis Kyzirakos, Babis Nikolaou, [Michael Sioutis](#))
- [Semantic Wikis: Approaches, Applications, and Perspectives](#) (François Bry, Sebastian Schaffert, Denny Vrandečić, Klara Weiland)
- [OWL 2 Profiles: An Introduction to Lightweight Ontology Languages](#) (Markus Krötzsch)
- [Argumentation and the Web](#) (Francesca Toni)
- [Federation and Navigation in SPARQL 1.1](#) (Marcelo Arenas, Jorge Pérez)
- [Reasoning with Uncertain and Inconsistent Ontologies on the Semantic Web](#) (Guilin Qi, Jianfeng Du)
- [Linked Data Stream Processing](#) (Manfred Hauswirth, Danh Le Phuoc, Josiane Xavier Parreira)
- [Datalog and Its Extensions for Semantic Web Databases](#) (Georg Gottlob, Giorgio Orsi, Andreas Pieris, Mantas Šimkus)
- [Reasoning and Query Answering in Description Logics](#) (Magdalena Ortiz, Mantas Šimkus)
- [Reasoning and Ontologies in Data Extraction](#) (Sergio Flesca, Tim Furbche, Ermelinda Oro)

Program

SEPTEMBER 02 (SUNDAY)			
Time	What	Who	Where
16:30 - 18:00	Summer School Registration and Poster Setup		<a href="#">Info Desk and Zeichensaal 13&amp;14, main building, TU Vienna</a>

SEPTEMBER 03 (MONDAY)			
Time	What	Who	Where
08:00 - 09:00	Summer School Registration and Poster Setup		<a href="#">Info Desk and Zeichensaal 13&amp;14, main building, TU Vienna</a>
09:00 - 09:15	Summer School Opening		
09:15 - 10:30	<a href="#">Reasoning and Query Answering in Description Logics I</a>	<a href="#">Magdalena Ortiz</a> , <a href="#">Mantas Šimkus</a>	<a href="#">Lecture room HS 8, main building, TU Vienna</a>
10:30 - 11:00	Coffee break		
11:00 - 12:30	<a href="#">Reasoning and Query Answering in Description Logics II</a>	<a href="#">Magdalena Ortiz</a> , <a href="#">Mantas Šimkus</a>	
12:30 - 14:30	Lunch break		Vienna
14:30 - 16:00	<a href="#">Datalog and Its Extensions for Semantic Web Databases I</a>	<a href="#">Georg Gottlob</a> , <a href="#">Giorgio Orsi</a> , <a href="#">Andreas Pieris</a> , <a href="#">Mantas Šimkus</a>	<a href="#">Lecture room HS 8, main building, TU Vienna</a>
16:00 - 16:30	Coffee break		
16:30 - 18:00	<a href="#">Datalog and Its Extensions for Semantic Web Databases II</a>	<a href="#">Georg Gottlob</a> , <a href="#">Giorgio Orsi</a> , <a href="#">Andreas Pieris</a> , <a href="#">Mantas Šimkus</a>	

SEPTEMBER 04 (TUESDAY)			
Time	What	Who	Where
09:00 - 10:30	<a href="#">Semantic Wikis: Approaches, Applications, and Perspectives I</a>	<a href="#">François Bry</a> , <a href="#">Sebastian Schaffert</a> , <a href="#">Denny Vrandečić</a> , <a href="#">Klara Weiland</a>	<a href="#">Lecture room HS 8, main building, TU Vienna</a>
10:30 - 11:00	Coffee break		
11:00 - 12:30	<a href="#">Semantic Wikis: Approaches, Applications, and Perspectives II</a>	<a href="#">François Bry</a> , <a href="#">Sebastian Schaffert</a> , <a href="#">Denny Vrandečić</a> , <a href="#">Klara Weiland</a>	
12:30 -	Lunch break		Vienna



Platinum Sponsors



Silver Sponsors



Hosted by



# + Ontology/Schema



Time	Speaker	Topic	Room
09:15–12:30	Magdalena Ortiz	Reasoning and Query Answering in Description Logics	HS 8
09:15–12:30	Mantas Šimkus	Reasoning and Query Answering in Description Logics	HS 8
14:30–18:00	Georg Gottlob	Datalog and Its Extensions for Semantic Web Databases	HS 8
14:30–18:00	Giorgio Orsi	Datalog and Its Extensions for Semantic Web Databases	HS 8
14:30–18:00	Andreas Pieris	Datalog and Its Extensions for Semantic Web Databases	HS 8
14:30–18:00	Mantas Šimkus	Datalog and Its Extensions for Semantic Web Databases	HS 8
09:00-12:30	François Bry	Semantic Wikis: Approaches, Applications, and Perspectives	HS 8

...



Lectures

- [Data Models and Query Languages for Linked Geospatial Data](#) (Manolis Koubarakis, Manos Karpathiotakis, Kostis Kyzirakos, Babis Nikolaou, [Michael Sioutis](#))
- [Semantic Wikis: Approaches, Applications, and Perspectives](#) (François Bry, Sebastian Schaffert, Denny Vrandečić, Klara Weiland)
- [OWL 2 Profiles: An Introduction to Lightweight Ontology Languages](#) (Markus Krötzsch)
- [Argumentation and the Web](#) (Francesca Toni)
- [Federation and Navigation in SPARQL 1.1](#) (Marcelo Arenas, Jorge Pérez)
- [Reasoning with Uncertain and Inconsistent Ontologies on the Semantic Web](#) (Guilin Qi, Jianfeng Du)
- [Linked Data Stream Processing](#) (Manfred Hauswirth, Danh Le Phuoc, Josiane Xavier Parreira)
- [Datalog and Its Extensions for Semantic Web Databases](#) (Georg Gottlob, Giorgio Orsi, Andreas Pieris, Mantas Šimkus)
- [Reasoning and Query Answering in Description Logics](#) (Magdalena Ortiz, Mantas Šimkus)
- [Reasoning and Ontologies in Data Extraction](#) (Sergio Flesca, Tim Furrche, Ermelinda Oro)

Program

SEPTEMBER 02 (SUNDAY)			
Time	What	Who	Where
16:30 - 18:00	Summer School Registration and Poster Setup		<a href="#">Info Desk and Zeichensaal 13&amp;14, main building, TU Vienna</a>

SEPTEMBER 03 (MONDAY)			
Time	What	Who	Where
08:00 - 09:00	Summer School Registration and Poster Setup		<a href="#">Info Desk and Zeichensaal 13&amp;14, main building, TU Vienna</a>
09:00 - 09:15	Summer School Opening		
09:15 - 10:30	<a href="#">Reasoning and Query Answering in Description Logics I</a>	<a href="#">Magdalena Ortiz</a> , <a href="#">Mantas Šimkus</a>	<a href="#">Lecture room HS 8, main building, TU Vienna</a>
10:30 - 11:00	Coffee break		
11:00 - 12:30	<a href="#">Reasoning and Query Answering in Description Logics II</a>	<a href="#">Magdalena Ortiz</a> , <a href="#">Mantas Šimkus</a>	
12:30 - 14:30	Lunch break		Vienna
14:30 - 16:00	<a href="#">Datalog and Its Extensions for Semantic Web Databases I</a>	<a href="#">Georg Gottlob</a> , <a href="#">Giorgio Orsi</a> , <a href="#">Andreas Pieris</a> , <a href="#">Mantas Šimkus</a>	<a href="#">Lecture room HS 8, main building, TU Vienna</a>
16:00 - 16:30	Coffee break		
16:30 - 18:00	<a href="#">Datalog and Its Extensions for Semantic Web Databases II</a>	<a href="#">Georg Gottlob</a> , <a href="#">Giorgio Orsi</a> , <a href="#">Andreas Pieris</a> , <a href="#">Mantas Šimkus</a>	

SEPTEMBER 04 (TUESDAY)			
Time	What	Who	Where
09:00 - 10:30	<a href="#">Semantic Wikis: Approaches, Applications, and Perspectives I</a>	<a href="#">François Bry</a> , <a href="#">Sebastian Schaffert</a> , <a href="#">Denny Vrandečić</a> , <a href="#">Klara Weiland</a>	<a href="#">Lecture room HS 8, main building, TU Vienna</a>
10:30 - 11:00	Coffee break		
11:00 - 12:30	<a href="#">Semantic Wikis: Approaches, Applications, and Perspectives II</a>	<a href="#">François Bry</a> , <a href="#">Sebastian Schaffert</a> , <a href="#">Denny Vrandečić</a> , <a href="#">Klara Weiland</a>	
12:30 -	Lunch break		Vienna



Platinum Sponsors



Silver Sponsors



Hosted by



# + Ontology/Schema



Time	Speaker	Topic	Room
09:15–12:30	Magdalena Ortiz	Reasoning and Query Answering in Description Logics	HS 8
09:15–12:30	Mantas Šimkus	Reasoning and Query Answering in Description Logics	HS 8
14:30–18:00	Georg Gottlob	Datalog and Its Extensions for Semantic Web Databases	HS 8
14:30–18:00	Giorgio Orsi	Datalog and Its Extensions for Semantic Web Databases	HS 8
14:30–18:00	Andreas Pieris	Datalog and Its Extensions for Semantic Web Databases	HS 8
14:30–18:00	Mantas Šimkus	Datalog and Its Extensions for Semantic Web Databases	HS 8
09:00-12:30	François Bry	Semantic Wikis: Approaches, Applications, and Perspectives	HS 8

...

≈ Ontology Population

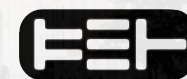


## Lectures

- [Data Models and Query Languages for Linked Geospatial Data](#) (Manolis Koubarakis, Manos Karpapathiotakis, Kostis Kyzirakos, Babis Nikolaou, [Michael Slioutis](#))
- [Semantic Wikis: Approaches, Applications, and Perspectives](#) (François Bry, Sebastian Schaffert, Denny Vrandečić, Klara Weiland)
- [OWL 2 Profiles: An Introduction to Lightweight Ontology Languages](#) (Markus Krötzsch)
- [Argumentation and the Web](#) (Francesca Toni)
- [Federation and Navigation in SPARQL 1.1](#) (Marcelo Arenas, Jorge Pérez)
- [Reasoning with Uncertain and Inconsistent Ontologies on the Semantic Web](#) (Guilin Qi, Jianfeng Du)
- [Linked Data Stream Processing](#) (Manfred Hauswirth, Danh Le Phuoc, Josiane Xavier Parreira)
- [Datalog and Its Extensions for Semantic Web Databases](#) (Georg Gottlob, Giorgio Orsi, Andreas Pieris, Mantas Šimkus)
- [Reasoning and Query Answering in Description Logics](#) (Magdalena Ortiz, Mantas Šimkus)
- [Reasoning and Ontologies in Data Extraction](#) (Sergio Flesca, Tim Furche, Ermelinda Oró)



## Platinum Sponsors



### Silver Sponsors



## Hosted by



## Program

SEPTEMBER 02 (SUNDAY)			
Time	What	Who	Where
16:30 - 18:00	Summer School Registration and Poster Setup		<a href="#">Info Desk and Zeichensaal 13&amp;14, main building, TU Vienna</a>

SEPTEMBER 03 (MONDAY)			
Time	What	Who	Where
08:00	Summer School		<a href="#">Info Desk and</a>
-	Registration and Poster		<a href="#">Zeichensaal 13&amp;14,</a>
09:00	Setup		<a href="#">main building, TU</a>
			<a href="#">Vienna</a>
09:00	Summer School		
-	Opening		
09:15			
09:15	<a href="#">Reasoning and Query</a>	<a href="#">Magdalena Ortiz,</a>	<a href="#">Lecture room HS 8,</a>
-	<a href="#">Answering in</a>	<a href="#">Mantas Šimkus</a>	<a href="#">main building, TU</a>
10:30	<a href="#">Description Logics I</a>		<a href="#">Vienna</a>
10:30	Coffee break		
-			
11:00			
11:00	<a href="#">Reasoning and Query</a>	<a href="#">Magdalena Ortiz,</a>	
-	<a href="#">Answering in</a>	<a href="#">Mantas Šimkus</a>	
12:30	<a href="#">Description Logics II</a>		
12:30	Lunch break		Vienna
-			
14:30			
14:30	<a href="#">Datalog and Its</a>	<a href="#">Georg Gottlob, Giorgio</a>	<a href="#">Lecture room HS 8,</a>
-	<a href="#">Extensions for</a>	<a href="#">Orsi, Andreas Pieris,</a>	<a href="#">main building, TU</a>
16:00	<a href="#">Semantic Web</a>	<a href="#">Mantas Šimkus</a>	<a href="#">Vienna</a>
	<a href="#">Databases I</a>		
16:00	Coffee break		
-			
16:30			
16:30	<a href="#">Datalog and Its</a>	<a href="#">Georg Gottlob, Giorgio</a>	
-	<a href="#">Extensions for</a>	<a href="#">Orsi, Andreas Pieris,</a>	
18:00	<a href="#">Semantic Web</a>	<a href="#">Mantas Šimkus</a>	
	<a href="#">Databases II</a>		

SEPTEMBER 04 (TUESDAY)			
Time	What	Who	Where
09:00	<a href="#">Semantic Wikis:</a>	<a href="#">François Bry,</a> <a href="#">Sebastian</a>	<a href="#">Lecture room HS</a>
-	<a href="#">Approaches, Applications,</a>	<a href="#">Schaffert,</a> <a href="#">Denny Vrandečić,</a>	<a href="#">8. main building,</a>
10:30	<a href="#">and Perspectives I</a>	<a href="#">Klara Weiland</a>	<a href="#">TU Vienna</a>
10:30	Coffee break		
-			
11:00			
11:00	<a href="#">Semantic Wikis:</a>	<a href="#">François Bry,</a> <a href="#">Sebastian</a>	
-	<a href="#">Approaches, Applications,</a>	<a href="#">Schaffert,</a> <a href="#">Denny Vrandečić,</a>	
12:30	<a href="#">and Perspectives II</a>	<a href="#">Klara Weiland</a>	
12:30	Lunch break		Vienna
-			

# Person

Magdalena Ortiz

Mantas Šimkus

Georg Gottlob

Giorgio Orsi

# Andreas Pieris

François Bry

## Topic

## Description Logics

## Datalog

# Semantic Wiki

# Room

HS 8



Lectures

- [Data Models and Query Languages for Linked Geospatial Data](#) (Manolis Koubarakis, Manos Karpathiotakis, Kostis Kyzirakos, Babis Nikolaou, [Michael Sioutis](#))
- [Semantic Wikis: Approaches, Applications, and Perspectives](#) (François Bry, Sebastian Schaffert, Denny Vrandečić, Klara Weiland)
- [OWL 2 Profiles: An Introduction to Lightweight Ontology Languages](#) (Markus Krötzsch)
- [Argumentation and the Web](#) (Francesca Toni)
- [Federation and Navigation in SPARQL 1.1](#) (Marcelo Arenas, Jorge Pérez)
- [Reasoning with Uncertain and Inconsistent Ontologies on the Semantic Web](#) (Guilin Qi, Jianfeng Du)
- [Linked Data Stream Processing](#) (Manfred Hauswirth, Danh Le Phuoc, Josiane Xavier Parreira)
- [Datalog and Its Extensions for Semantic Web Databases](#) (Georg Gottlob, Giorgio Orsi, Andreas Pieris, Mantas Šimkus)
- [Reasoning and Query Answering in Description Logics](#) (Magdalena Ortiz, Mantas Šimkus)
- [Reasoning and Ontologies in Data Extraction](#) (Sergio Flesca, Tim Furrer, Ermelinda Oro)

Program

SEPTEMBER 02 (SUNDAY)			
Time	What	Who	Where
16:30 - 18:00	Summer School Registration and Poster Setup		<a href="#">Info Desk and Zeichensaal 13&amp;14, main building, TU Vienna</a>

SEPTEMBER 03 (MONDAY)			
Time	What	Who	Where
08:00 - 09:00	Summer School Registration and Poster Setup		<a href="#">Info Desk and Zeichensaal 13&amp;14, main building, TU Vienna</a>
09:00 - 09:15	Summer School Opening		
09:15 - 10:30	<a href="#">Reasoning and Query Answering in Description Logics I</a>	<a href="#">Magdalena Ortiz</a> , <a href="#">Mantas Šimkus</a>	<a href="#">Lecture room HS 8, main building, TU Vienna</a>
10:30 - 11:00	Coffee break		
11:00 - 12:30	<a href="#">Reasoning and Query Answering in Description Logics II</a>	<a href="#">Magdalena Ortiz</a> , <a href="#">Mantas Šimkus</a>	
12:30 - 14:30	Lunch break		Vienna
14:30 - 16:00	<a href="#">Datalog and Its Extensions for Semantic Web Databases I</a>	<a href="#">Georg Gottlob</a> , <a href="#">Giorgio Orsi</a> , <a href="#">Andreas Pieris</a> , <a href="#">Mantas Šimkus</a>	<a href="#">Lecture room HS 8, main building, TU Vienna</a>
16:00 - 16:30	Coffee break		
16:30 - 18:00	<a href="#">Datalog and Its Extensions for Semantic Web Databases II</a>	<a href="#">Georg Gottlob</a> , <a href="#">Giorgio Orsi</a> , <a href="#">Andreas Pieris</a> , <a href="#">Mantas Šimkus</a>	

SEPTEMBER 04 (TUESDAY)			
Time	What	Who	Where
09:00 - 10:30	<a href="#">Semantic Wikis: Approaches, Applications, and Perspectives I</a>	<a href="#">François Bry</a> , <a href="#">Sebastian Schaffert</a> , <a href="#">Denny Vrandečić</a> , <a href="#">Klara Weiland</a>	<a href="#">Lecture room HS 8, main building, TU Vienna</a>
10:30 - 11:00	Coffee break		
11:00 - 12:30	<a href="#">Semantic Wikis: Approaches, Applications, and Perspectives II</a>	<a href="#">François Bry</a> , <a href="#">Sebastian Schaffert</a> , <a href="#">Denny Vrandečić</a> , <a href="#">Klara Weiland</a>	
12:30 -	Lunch break		Vienna



Platinum Sponsors



Silver Sponsors



Hosted by



# – Structure

# – Domain

## Person

Magdalena Ortiz

Mantas Šimkus

Georg Gottlob

Giorgio Orsi

Andreas Pieris

François Bry

## Topic

Description Logics

Datalog

Semantic Wiki

## Room

HS 8

≈ List/Entity Extraction



Lectures

- [Data Models and Query Languages for Linked Geospatial Data](#) (Manolis Koubarakis, Manos Karpathiotakis, Kostis Kyzirakos, Babis Nikolaou, [Michael Sioutis](#))
- [Semantic Wikis: Approaches, Applications, and Perspectives](#) (François Bry, Sebastian Schaffert, Denny Vrandečić, Klara Weiland)
- [OWL 2 Profiles: An Introduction to Lightweight Ontology Languages](#) (Markus Krötzsch)
- [Argumentation and the Web](#) (Francesca Toni)
- [Federation and Navigation in SPARQL 1.1](#) (Marcelo Arenas, Jorge Pérez)
- [Reasoning with Uncertain and Inconsistent Ontologies on the Semantic Web](#) (Guilin Qi, Jianfeng Du)
- [Linked Data Stream Processing](#) (Manfred Hauswirth, Danh Le Phuoc, Josiane Xavier Parreira)
- [Datalog and Its Extensions for Semantic Web Databases](#) (Georg Gottlob, Giorgio Orsi, Andreas Pieris, Mantas Šimkus)
- [Reasoning and Query Answering in Description Logics](#) (Magdalena Ortiz, Mantas Šimkus)
- [Reasoning and Ontologies in Data Extraction](#) (Sergio Flesca, Tim Furché, Ermelinda Oro)

Program

SEPTEMBER 02 (SUNDAY)			
Time	What	Who	Where
16:30 - 18:00	Summer School Registration and Poster Setup		<a href="#">Info Desk and Zeichensaal 13&amp;14, main building, TU Vienna</a>

SEPTEMBER 03 (MONDAY)			
Time	What	Who	Where
08:00 - 09:00	Summer School Registration and Poster Setup		<a href="#">Info Desk and Zeichensaal 13&amp;14, main building, TU Vienna</a>
09:00 - 09:15	Summer School Opening		
09:15 - 10:30	<a href="#">Reasoning and Query Answering in Description Logics I</a>	<a href="#">Magdalena Ortiz</a> , <a href="#">Mantas Šimkus</a>	<a href="#">Lecture room HS 8, main building, TU Vienna</a>
10:30 - 11:00	Coffee break		
11:00 - 12:30	<a href="#">Reasoning and Query Answering in Description Logics II</a>	<a href="#">Magdalena Ortiz</a> , <a href="#">Mantas Šimkus</a>	
12:30 - 14:30	Lunch break		Vienna
14:30 - 16:00	<a href="#">Datalog and Its Extensions for Semantic Web Databases I</a>	<a href="#">Georg Gottlob</a> , <a href="#">Giorgio Orsi</a> , <a href="#">Andreas Pieris</a> , <a href="#">Mantas Šimkus</a>	<a href="#">Lecture room HS 8, main building, TU Vienna</a>
16:00 - 16:30	Coffee break		
16:30 - 18:00	<a href="#">Datalog and Its Extensions for Semantic Web Databases II</a>	<a href="#">Georg Gottlob</a> , <a href="#">Giorgio Orsi</a> , <a href="#">Andreas Pieris</a> , <a href="#">Mantas Šimkus</a>	

SEPTEMBER 04 (TUESDAY)			
Time	What	Who	Where
09:00 - 10:30	<a href="#">Semantic Wikis: Approaches, Applications, and Perspectives I</a>	<a href="#">François Bry</a> , <a href="#">Sebastian Schaffert</a> , <a href="#">Denny Vrandečić</a> , <a href="#">Klara Weiland</a>	<a href="#">Lecture room HS 8, main building, TU Vienna</a>
10:30 - 11:00	Coffee break		
11:00 - 12:30	<a href="#">Semantic Wikis: Approaches, Applications, and Perspectives II</a>	<a href="#">François Bry</a> , <a href="#">Sebastian Schaffert</a> , <a href="#">Denny Vrandečić</a> , <a href="#">Klara Weiland</a>	
12:30 -	Lunch break		Vienna



Platinum Sponsors



Silver Sponsors

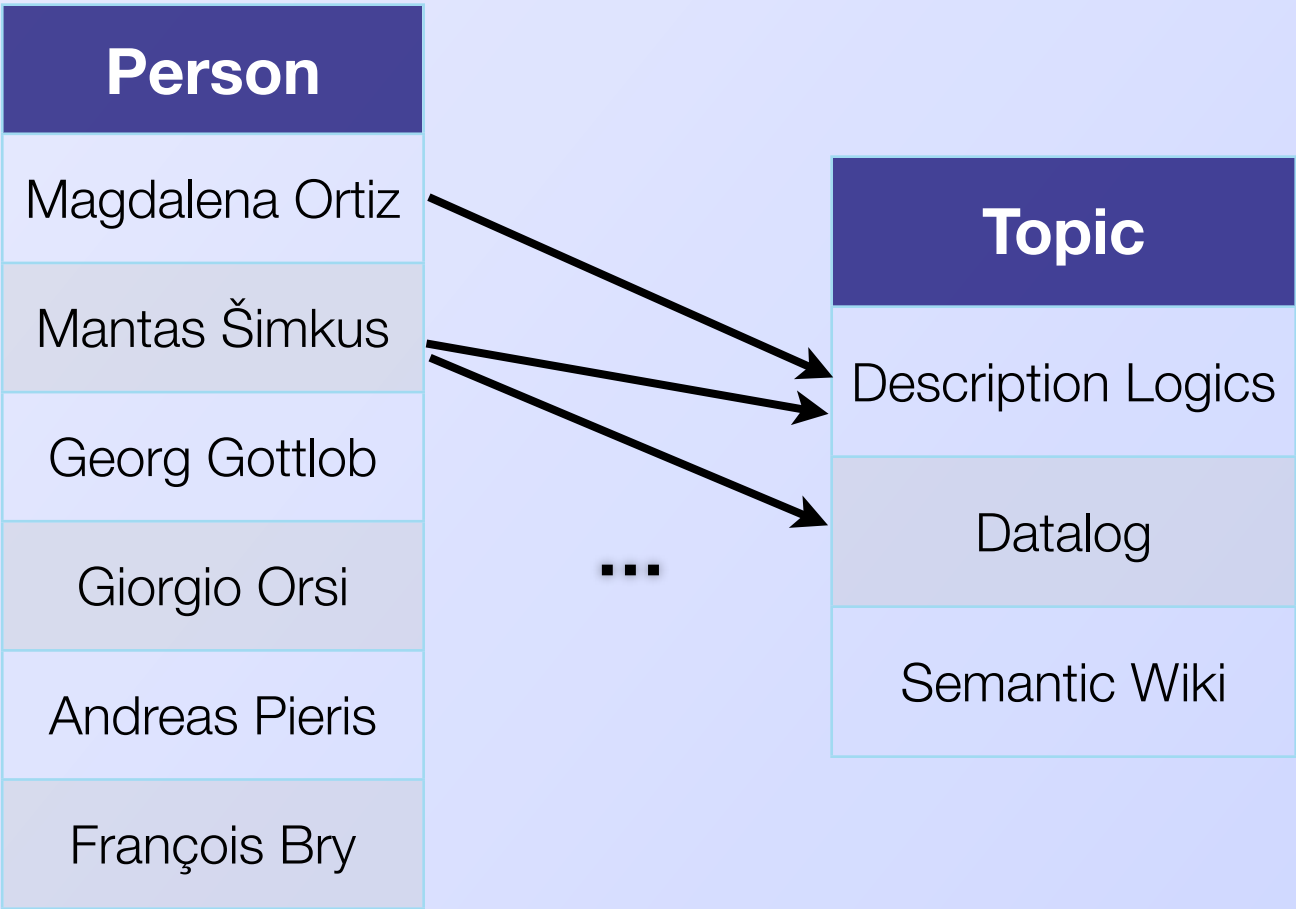


Hosted by



# + Structure

but less than in proper DE





Lectures

- [Data Models and Query Languages for Linked Geospatial Data](#) (Manolis Koubarakis, Manos Karpathiotakis, Kostis Kyzirakos, Babis Nikolaou, [Michael Sioutis](#))
- [Semantic Wikis: Approaches, Applications, and Perspectives](#) (François Bry, Sebastian Schaffert, Denny Vrandečić, Klara Weiland)
- [OWL 2 Profiles: An Introduction to Lightweight Ontology Languages](#) (Markus Krötzsch)
- [Argumentation and the Web](#) (Francesca Toni)
- [Federation and Navigation in SPARQL 1.1](#) (Marcelo Arenas, Jorge Pérez)
- [Reasoning with Uncertain and Inconsistent Ontologies on the Semantic Web](#) (Guilin Qi, Jianfeng Du)
- [Linked Data Stream Processing](#) (Manfred Hauswirth, Danh Le Phuoc, Josiane Xavier Parreira)
- [Datalog and Its Extensions for Semantic Web Databases](#) (Georg Gottlob, Giorgio Orsi, Andreas Pieris, Mantas Šimkus)
- [Reasoning and Query Answering in Description Logics](#) (Magdalena Ortiz, Mantas Šimkus)
- [Reasoning and Ontologies in Data Extraction](#) (Sergio Flesca, Tim Furché, Ermelinda Oro)

Program

SEPTEMBER 02 (SUNDAY)			
Time	What	Who	Where
16:30 - 18:00	Summer School Registration and Poster Setup		<a href="#">Info Desk and Zeichensaal 13&amp;14, main building, TU Vienna</a>

SEPTEMBER 03 (MONDAY)			
Time	What	Who	Where
08:00 - 09:00	Summer School Registration and Poster Setup		<a href="#">Info Desk and Zeichensaal 13&amp;14, main building, TU Vienna</a>
09:00 - 09:15	Summer School Opening		
09:15 - 10:30	<a href="#">Reasoning and Query Answering in Description Logics I</a>	<a href="#">Magdalena Ortiz</a> , <a href="#">Mantas Šimkus</a>	<a href="#">Lecture room HS 8, main building, TU Vienna</a>
10:30 - 11:00	Coffee break		
11:00 - 12:30	<a href="#">Reasoning and Query Answering in Description Logics II</a>	<a href="#">Magdalena Ortiz</a> , <a href="#">Mantas Šimkus</a>	
12:30 - 14:30	Lunch break		Vienna
14:30 - 16:00	<a href="#">Datalog and Its Extensions for Semantic Web Databases I</a>	<a href="#">Georg Gottlob</a> , <a href="#">Giorgio Orsi</a> , <a href="#">Andreas Pieris</a> , <a href="#">Mantas Šimkus</a>	<a href="#">Lecture room HS 8, main building, TU Vienna</a>
16:00 - 16:30	Coffee break		
16:30 - 18:00	<a href="#">Datalog and Its Extensions for Semantic Web Databases II</a>	<a href="#">Georg Gottlob</a> , <a href="#">Giorgio Orsi</a> , <a href="#">Andreas Pieris</a> , <a href="#">Mantas Šimkus</a>	

SEPTEMBER 04 (TUESDAY)			
Time	What	Who	Where
09:00 - 10:30	<a href="#">Semantic Wikis: Approaches, Applications, and Perspectives I</a>	<a href="#">François Bry</a> , <a href="#">Sebastian Schaffert</a> , <a href="#">Denny Vrandečić</a> , <a href="#">Klara Weiland</a>	<a href="#">Lecture room HS 8, main building, TU Vienna</a>
10:30 - 11:00	Coffee break		
11:00 - 12:30	<a href="#">Semantic Wikis: Approaches, Applications, and Perspectives II</a>	<a href="#">François Bry</a> , <a href="#">Sebastian Schaffert</a> , <a href="#">Denny Vrandečić</a> , <a href="#">Klara Weiland</a>	
12:30 -	Lunch break		Vienna



Platinum Sponsors



Silver Sponsors



Hosted by

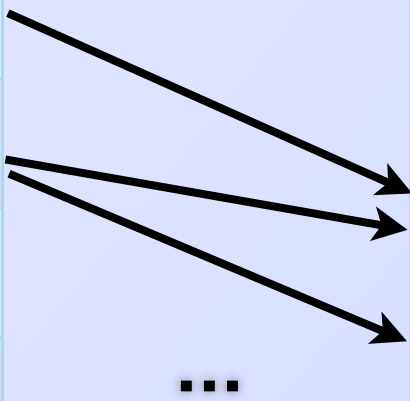


# + Structure

but less than in proper DE

Person
Magdalena Ortiz
Mantas Šimkus
Georg Gottlob
Giorgio Orsi
Andreas Pieris
François Bry

Topic
Description Logics
Datalog
Semantic Wiki



# ≈ Relation Extraction



# Variations II: Input



Lectures

- [Data Models and Query Languages for Linked Geospatial Data](#) (Manolis Koubarakis, Manos Karpapthiotakis, Kostas Kyzirakos, Babis Nikolaou, [Michael Sioutis](#))
- [Semantic Wikis: Approaches, Applications, and Perspectives](#) (François Bry, Sebastian Schaffert, Denny Vrandečić, Klara Weiland)
- [OWL 2 Profiles: An Introduction to Lightweight Ontology Languages](#) (Markus Krötzsch)
- [Argumentation and the Web](#) (Francesca Toni)
- [Federation and Navigation in SPARQL 1.1](#) (Marcelo Arenas, Jorge Pérez)
- [Reasoning with Uncertain and Inconsistent Ontologies on the Semantic Web](#) (Guilin Qi, Jianfeng Du)
- [Linked Data Stream Processing](#) (Manfred Hauswirth, Danh Le Phuoc, Josiane Xavier Parreira)
- [Datalog and Its Extensions for Semantic Web Databases](#) (Georg Gottlob, Giorgio Orsi, Andreas Pieris, Mantas Šimkus)
- [Reasoning and Query Answering in Description Logics](#) (Magdalena Ortiz, Mantas Šimkus)
- [Reasoning and Ontologies in Data Extraction](#) (Serjio Plesca, Tim Fuchs, Ermelinda Orto)

Program

SEPTEMBER 02 (SUNDAY)			
Time	What	Who	Where
16:30 - 18:00	Summer School Registration and Poster Setup		<a href="#">Info Desk and Zeichensaal 13&amp;14, main building, TU Vienna</a>
SEPTEMBER 03 (MONDAY)			
Time	What	Who	Where
08:00 - 09:00	Summer School Registration and Poster Setup		<a href="#">Info Desk and Zeichensaal 13&amp;14, main building, TU Vienna</a>
09:00 - 09:15	Summer School Opening		
09:15 - 10:30	<a href="#">Reasoning and Query Answering in Description Logics I</a>	<a href="#">Magdalena Ortiz, Mantas Šimkus</a>	<a href="#">Lecture room HS 8, main building, TU Vienna</a>
10:30 - 11:00	Coffee break		
11:00 - 12:30	<a href="#">Reasoning and Query Answering in Description Logics II</a>	<a href="#">Magdalena Ortiz, Mantas Šimkus</a>	
12:30 - 14:30	Lunch break		Vienna
14:30 - 16:00	<a href="#">Datalog and Its Extensions for Semantic Web Databases I</a>	<a href="#">Georg Gottlob, Giorgio Orsi, Andreas Pieris, Mantas Šimkus</a>	<a href="#">Lecture room HS 8, main building, TU Vienna</a>
16:00 - 16:30	Coffee break		
16:30 - 18:00	<a href="#">Datalog and Its Extensions for Semantic Web Databases II</a>	<a href="#">Georg Gottlob, Giorgio Orsi, Andreas Pieris, Mantas Šimkus</a>	
SEPTEMBER 04 (TUESDAY)			
Time	What	Who	Where
09:00 - 10:30	<a href="#">Semantic Wikis: Approaches, Applications, and Perspectives I</a>	<a href="#">François Bry, Sebastian Schaffert, Denny Vrandečić, Klara Weiland</a>	<a href="#">Lecture room HS 8, main building, TU Vienna</a>
10:30 - 11:00	Coffee break		
11:00 - 12:30	<a href="#">Semantic Wikis: Approaches, Applications, and Perspectives II</a>	<a href="#">François Bry, Sebastian Schaffert, Denny Vrandečić, Klara Weiland</a>	
12:30 - 14:30	Lunch break		Vienna
14:30 - 16:30	<a href="#">Semantic Wikis: Approaches, Applications, and Perspectives III</a>	<a href="#">François Bry, Sebastian Schaffert, Denny Vrandečić, Klara Weiland</a>	<a href="#">Lecture room HS 8, main building, TU Vienna</a>
16:30 - 18:00	Student Poster Session, incl. coffee		
SEPTEMBER 05 (WEDNESDAY)			
Time	What	Who	Where
09:00 - 10:30	<a href="#">Federation and Navigation in SPARQL 1.1 I</a>	<a href="#">Marcelo Arenas, Jorge Pérez</a>	<a href="#">Lecture room HS 8, main building, TU Vienna</a>
10:30 - 11:00	Coffee break		



Platinum Sponsors



Silver Sponsors



Hosted by





Lectures

- [Data Models and Query Languages for Linked Geospatial Data](#) (Manolis Kox Karpachiotakis, Kostas Kyzirakis, Babis Nikolaou, [Michael Sioutis](#))
- [Semantic Wikis: Approaches, Applications, and Perspectives](#) (François Br. Schaffert, Denny Vrandečić, Klara Weiland)
- [OWL 2 Profiles: An Introduction to Lightweight Ontology Languages](#) (Markus Argumentation and the Web (Francesca Toni)
- [Federation and Navigation in SPARQL 1.1](#) (Marcelo Arenas, Jorge Pérez)
- [Reasoning with Uncertain and Inconsistent Ontologies on the Semantic Web](#) Jianfeng Du)
- [Linked Data Stream Processing](#) (Manfred Hauswirth, Danh Le Phuoc, Josiah Datablog and Its Extensions for Semantic Web Databases (Georg Gottlob, G. Andreas Pieris, Mantas Simkus)
- [Reasoning and Query Answering in Description Logics](#) (Magdalena Ortiz, M. Reasoning and Ontologies in Data Extraction (Gennio Plescia, Tim Fuchs, Er

Program

SEPTEMBER
Time
16:30 - 18:00
SEPTEMBER
Time
08:00
09:00
09:00 - 09:15
09:15 - 10:30
10:30 - 11:00
11:00 - 12:30
12:30 - 14:30
14:30 - 16:00
16:00 - 16:30
16:30 - 18:00
SEPTEMBER
Time
09:00 - 10:30
10:30 - 11:00
11:00 - 12:30
12:30 - 14:30
14:30 - 16:30
16:30 - 18:00
SEPTEMBER
Time
09:00 - 10:30
10:30 - 11:00
11:00 - 12:30
12:30 - 14:30
14:30 - 16:30
16:30 - 18:00



GENERAL INFORMATION ▾ PARTICIPANTS ▾ CONTRIBUTORS ▾ NEWS CONTACT SEARCH



VLDB 2012 Program



Home

Contact

GENERAL

- Previous Conferences
- Photo Gallery
- Co-located Events

CONTRIBUTORS

- Accepted Papers
- Conference Programme
- Workshop Programme
- Call for Papers
- Paper registration/submission
- Formatting guidelines
- LNCS/LNBIP copyright
- Journal: LNCS Transactions Subline

PARTICIPANTS

- Registration of Participation
- Accommodation
- Conference Venue
- Social Events
- Fees



Conference Programme

23rd International Conference on Database and Expert Systems Applications (DEXA 2012 )

2012-09-03 [10:30 - 12:00] : Session 1A: XML Queries and Labeling I

Session Chair : Ruiming Tang

- [A Hybrid Approach for General XML Query Processing](#) Huiyu Wu, Ruiming Tang, Tok Wang Ling, Yong Zeng, Stéphane Bressan
- [SCOOTER: A Compact and Scalable Dynamic Labeling Scheme for XML Updates](#) Martin O'Connor, Mark Roantree
- [Reuse the Deleted Labels for Vector Order-based Dynamic XML Labeling Schemes](#) Canwei Zhuang, Shaorong Feng



2012-09-03 [10:30 - 12:00] : Session 1B: Computational Efficiency

- [Towards an Efficient Flash-Based Mid-Tier Cache](#) Yi Ou, Jianliang Xu, Theo Härder
- [Evacuation Planning of Large Buildings using Ladders](#) Alka Bhushan, Nandlal Sarda, Rami Reddy
- [A Write Efficient PCM-Aware Sort](#) Venkata Vamsikrishna Meduri, Zhan Su, Kian-Lee Tan



2012-09-03 [12:00 - 13:30] : Lunch

2012-09-03 [13:30 - 15:00] : Session 2A: XML Queries

Session Chair : Zeng Yong

- [Performance Evaluation for Reasoning about XML Keys](#) Flavio Ferrarotti, Sven Hartmann, Sebastian Link, Mauricio Marin, Emir Muñoz
- [Finding top-K Correct XPath Queries of User's Incorrect XPath Query](#) Ikeda Kosetsu, Nobutaka Suzuki
- [Analyzing Plan Diagrams of XQuery Optimizers](#) Bruhathi H S, Jayant Haritsa



2012-09-03 [13:30 - 15:00] : Session 2B: Data Extraction

Session Chair : Damir Vadic

- [Spreadsheet Metadata Extraction: A Layout-Based Approach](#) Somchai Chaivichienchai
- [Automated Extraction of Semantic Concepts From Semi-Structured Data: Supporting Computer-based Education Through the Analysis of Lecture notes](#) Thushari Atapattu, Katrina Pal
- [A Confidence-Weighted Metric for Unsupervised Ontology Population from Web Texts](#) Roberto Quaresima, Alessandro Delfino, Stefano Ferraresi, Paolo Frosini, Roberto Costantini



2012-09-03 [15:00 - 15:30] : Coffee Break

2012-09-03 [15:30 - 17:00] : Session 3A: Personalization, Preferences, and Ranking

Session Chair : John Debenham

- [Situation-Aware User's Interests prediction for query enrichment](#) Imen Ben Sassi, Chiraz Trabelsi, Amel Bouzeghoub, Sadok Ben Yahia
- [The Effective Relevance Link between a Document and a Query](#) Karam Abdulahad, Jean-Pierre Chevallet, Catherine Bernut
- [Incremental Computation of Skyline Queries with Dynamic Preferences](#) Tassadit Bouadi, Marie-Odile Cordier, René Quiniou



2012-09-03 [15:30 - 17:00] : Session 3B: Databases and Schemas

Invited Talks

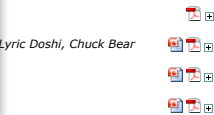
More information about the invited talks can be found [here](#).



Sponsors



Adrian van A	Ballroom 1	Saturn
twani		
M) in the Era of Self-Organizing, Self-		
Tutorial7: eHealth	Demo3: Information Retrieval, Web, and Mobility	
Tutorial8: Privacy in Cloud	Demo1: MapReduce, Big Data Systems, and Crowdsourcing	
Tutorial9: Graph synopses	Demo2: Query Pricing, Processing, and Optimization	PhD3: Query Processing



Huge scale:  
many Sites, many domains











## Property in Oxford search results

Save search

« [Back to postcode area map](#)

1 2 3 4 5 6 NEXT

1 - 10 of 144 properties



### White Road, East Hendred

**£1,499,950**

This impressive property offers an ideal opportunity to live the dream and run your own business from home. A five bedroom family house for you, and a ten bedroom, AA 4\* rated ten bedroom guest house/B&B wing. The family house benefits from generously proportioned living space with modern interiors, conservatory, dining room [...]

Bedrooms: **15** Bathrooms: **13** Area: **East Hendred** Type: **House**

[View full property details »](#)



### Bagley Wood Rd, Kennington OX1

**£1,295,000**

A substantial detached Edwardian family house with scope for updating and improvement, set within delightful gardens and wooded grounds extending to over five acres located on the edge of Bagley Wood. This fine Edwardian detached family house believed to date from about 1903, is set within its own delightful gardens and grounds including [...]

Bedrooms: **6** Bathrooms: **3** Area: **Kennington** Type: **House**

[View full property details »](#)



### Jackson Drive, Kennington OX1

**£1,100,000**

Probably one of the most beautifully decorated and designed houses in Oxford. This six bedroom Edwardian family house is the ultimate lifestyle property. Arranged over three floors, the house offers six bedrooms, three bathrooms [...]

Bedrooms: **6** Bathrooms: **3** Area: **Kennington** Type: **House**

[View full property details »](#)



### Silver Trees, Bagley Wood Rd OX1


**£949,950**

Silver Trees is a lovely family home a stone's throw from Oxford, but surrounded by woods that give the house its name. Built originally for an Oxford Don in the early 1900's, this Edwardian house still has many original features. The accommodation comprises eat-in kitchen with granite worktops and range cooker, dining [...]



# Property in Oxford search results

« [Back to postcode area map](#) 1 2 3 4 5 6 NEXT 1 - 10 of 144 properties



**White Road, East Hendred** **£1,499,950**

This impressive property offers an ideal opportunity to live the dream and run your own business from home. A five bedroom family house for you, and a ten bedroom, AA 4\* rated ten bedroom guest house/B&B wing. The family house benefits from generously proportioned living space with modern interiors, conservatory, dining room [...]

Bedrooms: 15 Bathrooms: 13 Area: **East Hendred** Type: **House**

[View full property details »](#)




**Bagley Wood Rd, Kennington OX1** **£1,295,000**

A substantial detached Edwardian family house with scope for updating and improvement, set within delightful gardens and wooded grounds extending to over five acres located on the edge of Bagley Wood. This fine Edwardian detached family house believed to date from about 1903, is set within its own delightful gardens and grounds including [...]

Bedrooms: 6 Bathrooms: 3 Area: **Kennington** Type: **House**

[View full property details »](#)




**Jackson Drive, Kennington OX1** **£1,100,000**

Probably one of the most beautifully decorated and designed houses in Oxford. This six bedroom Edwardian family house is the ultimate lifestyle property. Arranged over three floors, the house offers six bedrooms, three bathrooms [...]

Bedrooms: 6 Bathrooms: 3 Area: **Kennington** Type: **House**

[View full property details »](#)



**Silver Trees, Bagley Wood Rd OX1** **£949,950**

Silver Trees is a lovely family home a stone's throw from Oxford, but surrounded by woods that give the house its name. Built originally for an Oxford Don in the early 1900's, this Edwardian house still has many original features. The accommodation comprises eat-in kitchen with granite worktops and range cooker, dining [...]

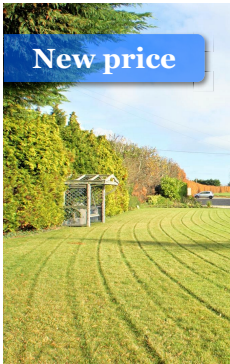
Page scale:  
single page, repeated structure



## Property

« [Back to postcode area](#)

New price



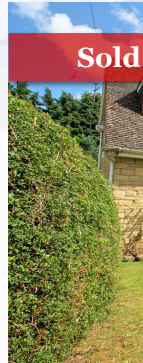
Sold



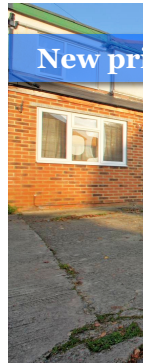
## Prope

« [Back to postc](#)

Sold



New pri



## Property in Oxford search results

Save search

« [Back to postcode area map](#)

PREVIOUS

4

5

6

7

8

9

10

11

12

13

14

NEXT

81 - 90 of 144 properties



### Hemmingway Drive, Bicester **OX26** £265,000

Spacious, detached family house with four double bedrooms, great storage, three reception rooms, modern eat-in kitchen, secure private garden and off-street parking. Arranged over two floors, the property features a large reception room, dining room, useful study room, modern family kitchen providing access to the garden, four double bedrooms, bathroom, ground [...]

Bedrooms: 4 Bathrooms: 1 Area: **Oxfordshire** Type: **House**

[View full property details »](#)



New instruction. Photos coming soon

### Lytton Road, Florence Park **OX4** £265,000

Totally refurbished four bedroomed mid-terraced family house located along a quiet residential street within Florence Park, off-street parking and large private garden. The property further benefits from a bright front reception, stylish new central kitchen, second rear reception providing access to garden with side-access, four well proportioned bedrooms and modern family bathroom. Located on [...]

Bedrooms: 4 Bathrooms: 1 Area: **Florence Park** Type: **House**

[View full property details »](#)



Sold

### Howard Street, East Oxford **OX4** £265,000

Offering modernisation potential with added extension possibility, this Impressive three double bedroomed terraced house has a spacious rear garden backing onto the allotments. The property comprises two reception rooms, kitchen with access to large private garden, three well proportioned bedrooms, family bathroom and shed at end of garden. Located on a popular residential [...]

Bedrooms: 3 Bathrooms: 1 Area: **East Oxford** Type: **House**

[View full property details »](#)



## White Road East Hendred

New price

£1,499,950

This impressive property offers an ideal opportunity to live the dream and run your own business from home. A five bedroom family house for you, and a ten bedroom, AA 4\* rated ten bedroom guest house/B&B wing.

The family house benefits from generously proportioned living space with modern interiors, conservatory, dining room with feature bow window and stunning first floor views over substantial private gardens.

With approximately 2 acres in total of grounds, there is further development potential subject to planning.

The picturesque village of East Hendred nestles at the foot of a chain of hills known as The Downs in the historic Vale of White Horse in the heart of England. The village boasts a pre-school, 2 primary schools, 3 pubs, a sports hall and village store including Post Office.

Local transport links are available nearby, providing access to Oxford and Wantage. Road users have easy access to the A34 (M40) towards London.

Reception: 24' x 12'10" / Conservatory: 10'9" x 10'8"  
Dining Room: 16'10" x 12'7" / Family Room: 15'11" x 9'7"  
Kitchen: 13'11" x 12'7" / Master Bedroom: 16' x 9'9"  
2nd Bedroom: 13'1" x 11'9" / 3rd Bedroom: 18'8" x 11'9"  
4th Bedroom: 18'7" x 9'9" / 5th Bedroom: 12'2" x 10'7"



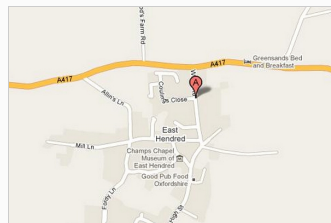
Gallery 360° virtual tour Floorplan Aerial view Map view Printable brochure Email a friend EPC Mortgage advice

### Key Details

Property Type	House
Postcode Area	OX12
Bedrooms	15
Bathrooms	13
Area	East Hendred
Tenure	Freehold
Sq feet	6648

### Property Location

Click the map image below for a larger interactive view.



### Live Postcode Intelligence

Applicants looking to buy in OX12 through White Walls Agency

93	1 bed+	0
	2 bed+	93
	3 bed+	
	Various	

Save property »

Sell for £555 in OX2 »

My property value?

Interested in this property?

Site scale:  
single site, many pages

Edward Parkinson

Mon to Fri 8am – 7pm

01865 246 100

15 Hollybush Row

Oxford

OX1 1JH

Arrange a viewing »

## search results

Save search

US 4 5 6 7 8 9 10 11 12 13 14 NEXT 81 - 90 of 144 properties

### Hemmingway Drive, Bicester OX26 £265,000

Spacious, detached family house with four double bedrooms, great storage, three reception rooms, modern eat-in kitchen, secure private garden and off-street parking. Arranged over two floors, the property features a large reception room, dining room, useful study room, modern family kitchen providing access to the garden, four double bedrooms, bathroom, ground [...]

Bedrooms: 4 Bathrooms: 1 Area: Oxfordshire Type: House

View full property details »

### Lytton Road, Florence Park OX4 £265,000

Totally refurbished four bedroomed mid-terraced family house located along a quiet residential street within Florence Park, off-street parking and large private garden. The property further benefits from a bright front reception, stylish new central kitchen, second rear reception providing access to garden with side-access, four well proportioned bedrooms and modern family bathroom. Located on [...]

Bedrooms: 4 Bathrooms: 1 Area: Florence Park Type: House

View full property details »

### Howard Street, East Oxford OX4 £265,000

Offering modernisation potential with added extension possibility, this Impressive three double bedroomed terraced house has a spacious rear garden backing onto the allotments. The property comprises two reception rooms, kitchen with access to large private garden, three well proportioned

View full property details »



# Property in Oxford search results

Save search

« [Back to postcode area map](#)

1 2 3 4 5 6 NEXT

1 - 10 of 144 properties



New price

## White Road, East Hendred £1,499,950

This impressive property offers an ideal opportunity to live the dream and run your own business from home. A five bedroom family house for you, and a ten bedroom, AA 4\* rated ten bedroom guest house/B&B wing. The family house benefits from generously proportioned living space with modern interiors, conservatory, dining room [...]

Bedrooms: 15 Bathrooms: 13 Area: East Hendred Type: House

[View full property details »](#)



Sold

## Bagley Wood Rd, Kennington OX1 £1,295,000

A substantial detached Edwardian family house with scope for updating and improvement, set within delightful gardens and wooded grounds extending to over five acres located on the edge of Bagley Wood. This fine Edwardian detached family house believed to date from about 1903, is set within its own delightful gardens and grounds including [...]

Bedrooms: 6 Bathrooms: 3 Area: Kennington Type: House

[View full property details »](#)



## Jackson Drive, Kennington OX1 £1,100,000

Probably one of the most beautifully decorated and designed houses in Oxford. This six bedroom Edwardian family house is the ultimate lifestyle property. Arranged over three floors, the house offers six bedrooms, three bathrooms [...]

Bedrooms: 6 Bathrooms: 3 Area: Kennington Type: House

[View full property details »](#)



## Silver Trees, Bagley Wood Rd OX1 £949,950

Silver Trees is a lovely family home a stone's throw from Oxford, but surrounded by woods that give the house its name. Built originally for an Oxford Don in the early 1900's, this Edwardian house still has many original features. The accommodation comprises eat-in kitchen with granite worktops and range cooker, dining [...]

Bedrooms: 5 Bathrooms: 2 Area: Kennington Type: House

[View full property details »](#)



## Carey Close, North Oxford OX2 £895,000

A fantastic detached family house situated within a quiet close in always desirable North Oxford. Immaculately presented throughout, the property is ideally laid out for modern family living. Central to this is a stunning family kitchen with dining area and fully retractable doors leading to a raised decking area and [...]

Bedrooms: 5 Bathrooms: 3 Area: North Oxford Type: House



# Property in Ox

« [Back to postcode area map](#)



Find property for sale in OX1

Draw a Search

Focus your search on *exactly* where you want to live with [Draw-a-Search](#)

Change location and criteria

Location:

Radius:

Property:

Beds:  to

Price (£):  to

Added:

☐ Include Sold STC

Update Results

Create an alert for OX1 or [save this search](#)

- Filter your results
- Property type:
- [Houses](#) (93)
- [Detached houses](#) (50)
- [Semi-detached houses](#) (19)
- [Terraced houses](#) (21)
- [Boats](#) (1)
- [Flats / Apartments](#) (25)
- [Bungalows](#) (18)
- [Land](#) (1)
- [Commercial Property](#) (1)
- Property style:
- [Character](#) (1)
- Preowned & new homes:
- [Pre-owned homes](#) (118)
- [Brand new homes](#) (1)
- Outside space:
- [Garden](#) (92)
- [Parking](#) (88)
- Ways to buy:
- [Shared ownership](#) (2)
- [Non-Shared ownership](#) (117)
- Retirement properties:
- [Retirement](#) (3)
- [Non-retirement](#) (116)

Need a Remover?

Use our quotes service to find local removal companies

Click to get quotes

Sold House Prices

List of properties Properties on map

119 properties for sale found Sort by: [Lowest price](#) [Highest price](#)

FEATURED AGENT	FEATURED PROPERTY	FEATURED PROPERTY
 <b>Local know how better results</b> Call 01865 246100 for a FREE no obligation valuation  <small>Terms and Conditions Apply</small>	 NEW LISTING Guide Price <b>£1,500,000</b> 5 bed detached house for sale <a href="#">Foxcombe Road, Boars Hill, Oxford, OX1</a> 	 <b>£390,000</b> 2 bed flat for sale <a href="#">The Lion Brewery Oxford City Centre</a> 

Over 60,000 properties in over 65 countries Find your dream home overseas with [rightmoveoverseas](#)

**Guide Price £3,850,000**  
**5 bedroom house for sale**  
[Boars Hill, Oxford, Oxfordshire](#)  
Tudor style mansion within minutes of the City of Oxford - An elegant private country house built by Millgate Homes with spacious and luxurious interiors, garage complex, guest /staff cottage and more than 1.7 acres of south facing gardens, including a swimming pool and summerhouse. Th...  
[More details, 14 photos, floorplan and brochure](#) | [Save property](#) | [Contact agent](#) | [Upgrade listing](#)  
Marketed by Knight Frank, Oxford. Telephone: 0843 314 8227 [BT 4p/min](#)

**PREMIUM LISTING**  
**£3,500,000**  
**6 bedroom detached house for sale**  
[Old Boars Hill, Oxford](#)  
Boars Hill Heath is a Victorian family home located at the end of a private lane in landscaped grounds of approximately 3.4 acres. There are 4 principle reception rooms, a Mark Wilkinson desigend kitchen/breakfast room; striking orangery opening onto a terrace which runs the length of the...  
[More details, 9 photos, floorplan and brochure](#) | [Save property](#) | [Contact agent](#)  
Marketed by Thomas Merrifield, Oxford . Telephone: 0843 315 0432 [BT 4p/min](#)

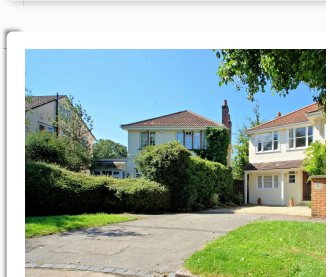
**£3,000,000**  
**6 bedroom detached house for sale**  
[Sandy Lane, Boars Hill, Oxford, OX1](#)  
A substantial family house with equestrian facilities in an elevated position overlooking its garden and grounds with spectacular views over the Vale of White Horse. (contd...)  
[More details, 10 photos, floorplans and brochure](#) | [Save property](#) | [Contact agent](#) | [Upgrade listing](#)  
Marketed by John D Wood & Co. Sales, Oxford. Telephone: 0843 313 1202 [BT 4p/min](#)

**Guide Price £2,500,000**  
**4 bedroom semi-detached house for sale**  
[Paradise Street, Central Oxford, OX1 1LD](#)



## Property in Oxford

[« Back to postcode area map](#)



For Sale To Let

ALL RESIDENTIAL COMMERCIAL DEVELOPMENT WOODLANDS SPORTING AGRICULTURAL NEWEST

You have selected: For Sale > 235 properties currently displayed

### FILTER PROPERTIES

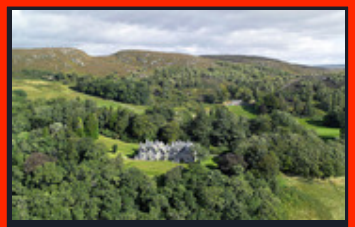
- By location: ▶
- By price range: ▶
- By offer type: ▶
- By property type: ▶
- By size: ▶



**eProperties**  
Sign up and receive regular eProperties updates by email  
[Find out more...](#)



**Contact**  
Want to know more or have an enquiry? Send us a message...  
[Find out more...](#)



**Tressady Estate, Rogart, By Lairg, Sutherland**  
Location: **Highlands**  
Guide Price: **£3,300,000**  
Bedrooms: **10**

[Find out more...](#)



**Mains of Rhynie Farm, Rhynie, Aberdeenshire AB54 4HX**  
Location: **Aberdeenshire**  
Offers Over: **£1,950,000**  
Bedrooms: **5**

[Find out more...](#)



**Craigengillan North Forest, Carsphairn, Dumfries & Galloway**  
Location: **Dumfries & Galloway**  
Guide Price: **£1,950,000**

[Find out more...](#)



**Hillhead Farm, Hillhead Road, Bieldside, Aberdeen**  
Location: **Aberdeenshire**  
Offers Over: **£1,575,000**  
Bedrooms: **4**

[Find out more...](#)



**Drumfork, Bridge of Cally, Perthshire**  
Location: **Perth & Kinross**  
Offers Over: **£1,500,000**  
Bedrooms: **11**

[Find out more...](#)



**West Garty Estate, Helmsdale, Sutherland**  
Location: **Highlands**  
Offers Over: **£1,500,000**  
**Under Offer**

[Find out more...](#)



**Ancarraig Lodges, Bunloit, Drumnadrochit**  
Location: **Highlands**  
Offers Over: **£800,000**

[Find out more...](#)



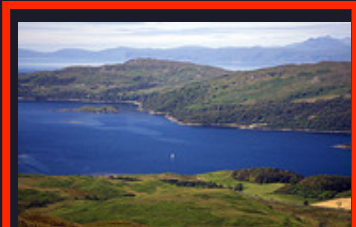
**Sealasdaire, Ardentallen, By Oban**  
Location: **Argyll & Bute**  
Offers Over: **£650,000**  
Bedrooms: **4**

[Find out more...](#)



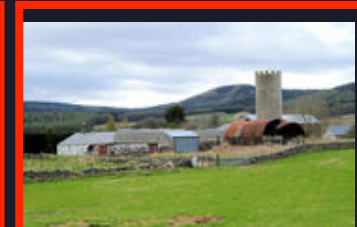
**Gairn Farm, Blacktop, Kingswells, Aberdeen**  
Location: **Aberdeenshire**  
Offers Over: **£595,000**  
Bedrooms: **2**

[Find out more...](#)



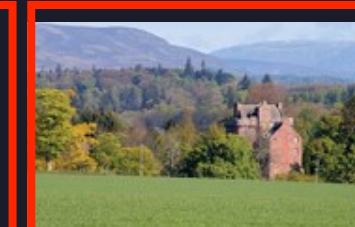
**Land at Kilmelford, Kilmelford, Argyll**  
Location: **Argyll & Bute**  
Offers Over: **£575,000**  
**Under Offer**

[Find out more...](#)



**Easter Bleaton, Bridge of Cally**  
Location: **Perth & Kinross**  
Offers Over: **£560,000**

[Find out more...](#)



**Inverquhar Castle, By Kirriemuir, Angus**  
Location: **Dundee & Angus**  
Offers Over: **£550,000**  
Bedrooms: **4**

[Find out more...](#)



## Property in Ox

[« Back to postcode area map](#)



For Sale To Let

ALL RESIDENTIAL

You have selected: For Sale

FILTER PROPERTIES

By location: ▶

By price range: ▶

By offer type: ▶

By property type: ▶

By size: ▶



eProperties

Sign up and receive regular eProperties updates by email

Find out more...



Contact

Want to know more or have an enquiry? Send us a message...

Find out more...

Domain scale:  
many site, many pages,  
same domain

Irvine

1,689 results. 7 unmapped.

- ☒ For Sale (1153)
- ☒ For Rent (455)
- ☒ Make Me Move (89)
- ☐ Recently Sold (7368)

Price

Min — Max

Beds

0+ ▶

Baths

0+ ▶

[Save search](#) [More filters](#)

Save this search and get email alerts for new listings.

bing

Sort by: **Featured** · Days on Zillow · Price · more ▼



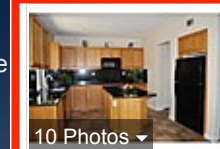
**40 Christamon S # 59, Irvine, CA** (Northwood)

🏠 **Condo For Sale: \$459,500**  
Zestimate®: \$415,125  
Est. Mortgage: \$2,048/mo  
Zillow Special Offer: **\$3,000**

Beds: 3  
Baths: 2.5  
Sqft: 1,383  
Lot: --

Days on Zillow: 24  
Built: 1984  
Condo  
Price/sqft: \$332

[Maps](#)  
[Save](#)  
[Comps](#)



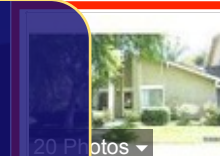
**111 Waterman, Irvine, CA** (Lower Peters Canyon)

🏠 **House For Sale: \$598,800**  
Zestimate®: --  
Est. Mortgage: \$2,669/mo  
[See current rates](#)

Beds: 4  
Baths: 3.0  
Sqft: 1,728  
Lot: --

Days on Zillow: 2  
Built: 2001  
Single Family  
Price/sqft: \$346

[Maps](#)  
[Save](#)  
[Comps](#)



**11 Denver # 85, Irvine, CA** (El Camino Real)

🏠 **Condo For Sale: \$389,000**  
Zestimate®: \$331,696  
Est. Mortgage: \$1,734/mo  
[See current rates](#)

Beds: 3  
Baths: 1.5  
Sqft: 1,223  
Lot: --

Days on Zillow: 2  
Built: 1977  
Condo  
Price/sqft: \$318

[Maps](#)  
[Save](#)  
[Comps](#)



**16 Tahoe, Irvine, CA** (University Park)

🏠 **House For Sale: \$539,500**

Beds: 2

Days on Zillow: 2

[Maps](#)



# Domain-specific data extraction







## White Road East Hendred

New price

£1,499,950

This impressive property offers an ideal opportunity to live the dream and run your own business from home. A five bedroom family house for you, and a ten bedroom, AA 4\* rated ten bedroom guest house/B&B wing.

The family house benefits from generously proportioned living space with modern interiors, conservatory, dining room with feature bow window and stunning first floor views over substantial private gardens.

With approximately 2 acres in total of grounds, there is further development potential subject to planning.

The picturesque village of East Hendred nestles at the foot of a chain of hills known as The Downs in the historic Vale of White Horse in the heart of England. The village boasts a pre-school, 2 primary schools, 3 pubs, a sports hall and village store including Post Office.

Local transport links are available nearby, providing access to Oxford and Wantage. Road users have easy access to the A34 (M40) towards London.

Reception: 24' x 12'10" / Conservatory: 10'9" x 10'8"  
Dining Room: 16'10" x 12'7" / Family Room: 15'11" x 9'7"  
Kitchen: 13'11" x 12'7" / Master Bedroom: 16' x 9'9"  
2nd Bedroom: 13'1" x 11'9" / 3rd Bedroom: 18'8" x 11'9"  
4th Bedroom: 18'7" x 9'9" / 5th Bedroom: 12'2" x 10'7"



## search results

Save search

US 4 5 6 7 8 9 10 11 12 13 14 NEXT 81 - 90 of 144 properties

### Hemmingway Drive, Bicester OX26 £265,000

Spacious, detached family house with four double bedrooms, great storage,

# supervised, site-specific data extraction

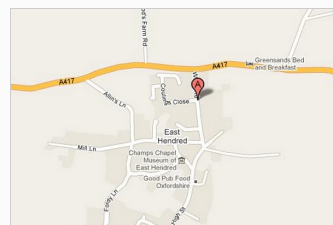
Gallery 360° virtual tour Floorplan Aerial view Map view Printable brochure Email a friend

### Key Details

Property Type	House
Postcode Area	OX12
Bedrooms	15
Bathrooms	13
Area	East Hendred
Tenure	Freehold
Sq feet	6648

### Property Location

Click the map image below for a larger interactive view.



### Live Post

Applicants looking to buy in OX12 through White Walls Agency

93	1 bed+	0
	2 bed+	93
	3 bed+	
	Various	

### Lytton Road, Florence Park OX4 £265,000

Totally refurbished four bedroomed mid-terraced family house located along a quiet residential street within Florence Park, off-street parking and large private garden. The property further benefits from a bright front reception, stylish new central kitchen, second rear reception providing access to garden with side-access, four well proportioned bedrooms and modern family bathroom. Located on [...]

Bedrooms: 4 Bathrooms: 1 Area: Florence Park Type: House

View full property details »

### Howard Street, East Oxford OX4 £265,000

Offering modernisation potential with added extension possibility, this Impressive three double bedroomed terraced house has a spacious rear garden backing onto the allotments. The property comprises two reception rooms, kitchen with access to large private garden, three well proportioned bedrooms, family bathroom and shed at end of garden. Located on a popular residential [...]

Bedrooms: 3 Bathrooms: 1 Area: East Oxford Type: House

View full property details »

Interested in this property?

Site scale:  
single site, many pages



Edward Parkinson

Mon to Fri 8am - 7pm

01865 246 100

15 Hollybush Row

Oxford

OX1 1JH

My property value?

Arrange a viewing »

in Share

f Like 1

Twitter Tweet 1

+1 0

Share

IMPORTANT NOTICE

1. White Walls Agency Limited for itself and for the vendor(s) or lessor(s) of this property whose agents they are, give notice that 1) These particulars do not constitute any part of an offer or contract 2) None of the statements contained in these particulars are to be relied upon as statements or representations of fact. 3) Any intending purchaser or lessee must satisfy himself by



## Property in Oxford

[« Back to postcode area map](#)



For Sale To Let

ALL RESIDENTIAL

You have selected: For Sale

FILTER PROPERTIES

By location: ▶

By price range: ▶

By offer type: ▶

By property type: ▶

By size: ▶



eProperties

Sign up and receive regular eProperties updates by email

Find out more...



Contact

Want to know more or have an enquiry? Send us a message...

Find out more...

**Domain scale:**  
many site, many pages,  
same domain

Irvine

1,689 results. 7 unmapped.

- ☒ For Sale (1153)
- ☒ For Rent (455)
- ☒ Make Me Move (89)
- ☐ Recently Sold (7368)

Price

Min — Max

Beds

0+ —

Baths

0+ —

[Save search](#) [More filters](#)

Save this search and get email alerts for new listings.

bing

Sort by: **Featured** · Days on Zillow · Price · more ▼



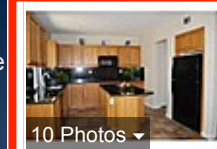
**40 Christamon S # 59, Irvine, CA** (Northwood)

**Condo For Sale: \$459,500**  
Zestimate®: \$415,125  
Est. Mortgage: \$2,048/mo  
Zillow Special Offer: **\$3,000**

Beds: 3  
Baths: 2.5  
Sqft: 1,383  
Lot: --

Days on Zillow: 24  
Built: 1984  
Condo  
Price/sqft: \$332

[Maps](#)  
[Save](#)  
[Comps](#)



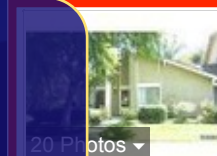
**111 Waterman, Irvine, CA** (Lower Peters Canyon)

**House For Sale: \$598,800**  
Zestimate®: --  
Est. Mortgage: \$2,669/mo  
[See current rates](#)

Beds: 4  
Baths: 3.0  
Sqft: 1,728  
Lot: --

Days on Zillow: 2  
Built: 2001  
Single Family  
Price/sqft: \$346

[Maps](#)  
[Save](#)  
[Comps](#)



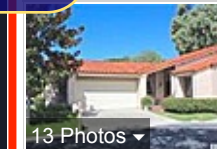
**11 Denver # 85, Irvine, CA** (El Camino Real)

**Condo For Sale: \$389,000**  
Zestimate®: \$331,696  
Est. Mortgage: \$1,734/mo  
[See current rates](#)

Beds: 3  
Baths: 1.5  
Sqft: 1,223  
Lot: --

Days on Zillow: 2  
Built: 1977  
Condo  
Price/sqft: \$318

[Maps](#)  
[Save](#)  
[Comps](#)



**16 Tahoe, Irvine, CA** (University Park)

**House For Sale: \$539,500**  
Zestimate®: \$541,458  
Est. Mortgage: \$2,405/mo  
[See current rates](#)

Beds: 2  
Baths: 2.0  
Sqft: 1,540  
Lot: 3,600

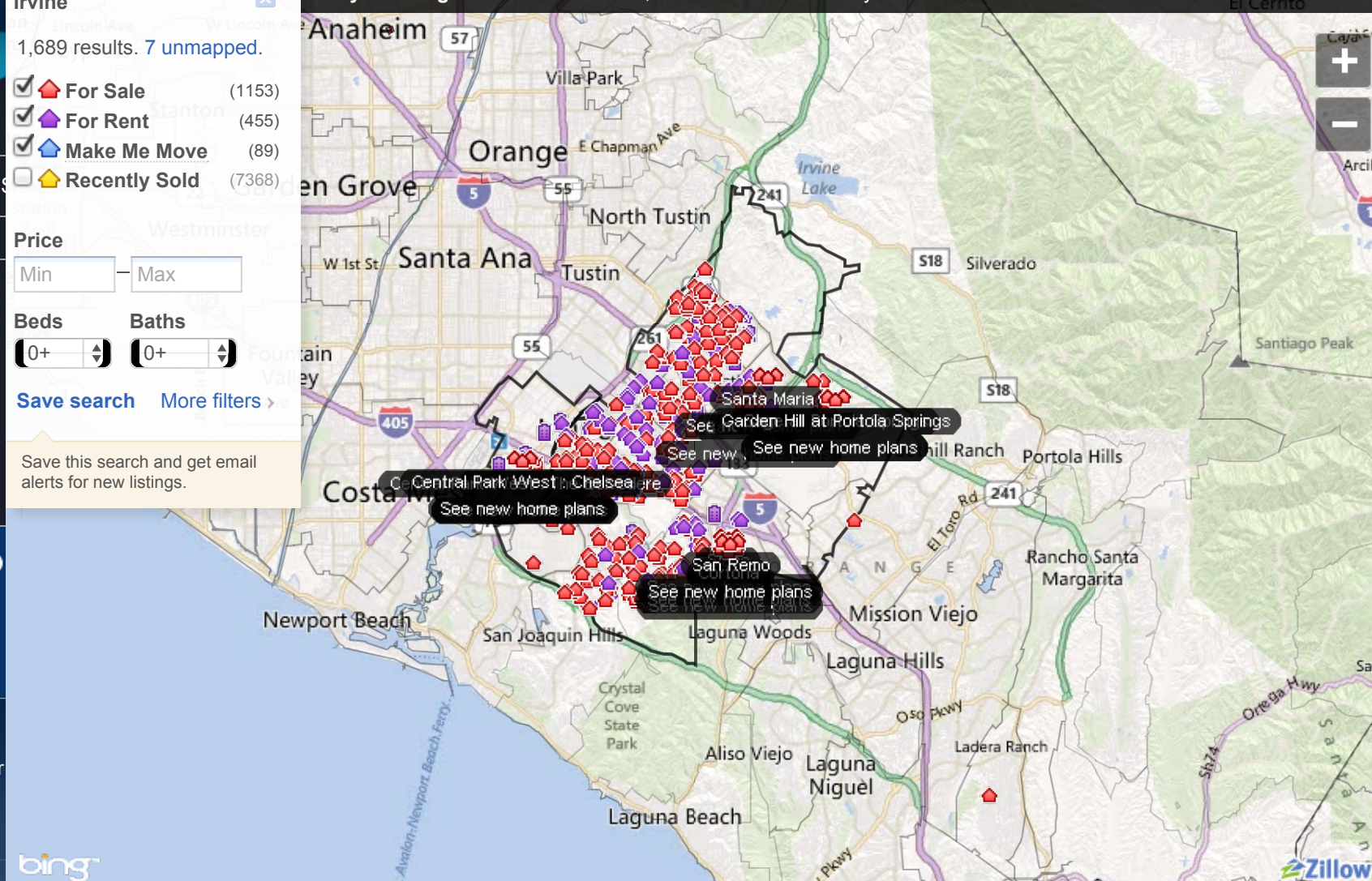
Days on Zillow: 2  
Built: 1976  
Single Family  
Price/sqft: \$350

[Maps](#)  
[Save](#)  
[Comps](#)



**30 Blue Heron, Irvine, CA** (Shady Canyon)

Only showing 500 homes. Zoom in, or use filters to narrow your search.



[Save this search](#) [Share](#) [Map](#)



## Property in Ox

[« Back to postcode area map](#)

New price

Sold

**Domain scale:**  
many site, many pages,  
same domain

For Sale To Let

ALL RESIDENTIAL

You have selected: For Sale

FILTER PROPERTIES

By location: ▶

By price range: ▶

By offer type: ▶

By property type: ▶

By size: ▶

**Bell  
Ingram**

eProperties

Sign up and receive regular

Irvine

1,689 results. 7 unmapped.

- ☒ For Sale (1153)
- ☒ For Rent (455)
- ☒ Make Me Move (89)
- ☐ Recently Sold (7368)

Price

Min — Max

Beds

0+ ▶

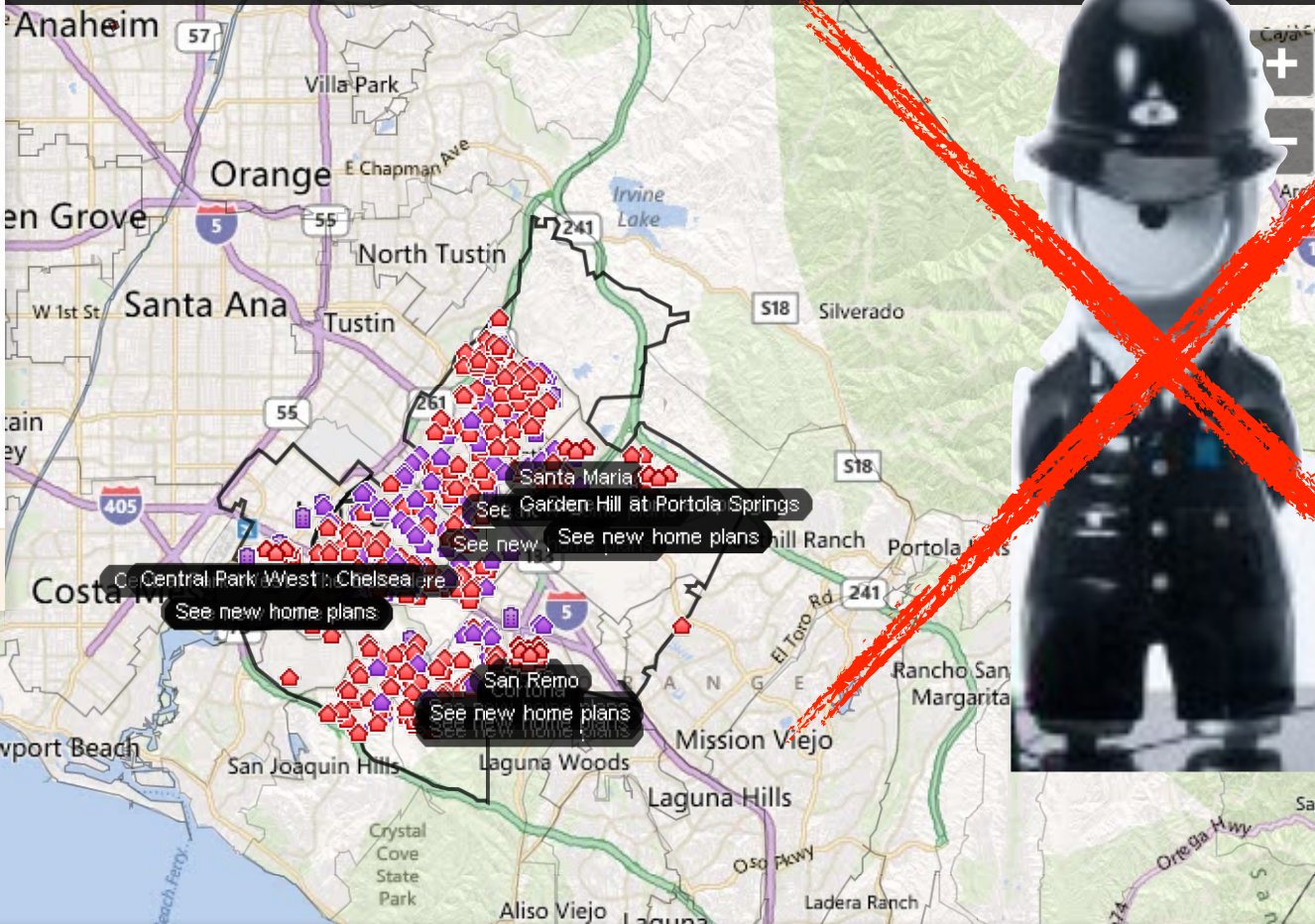
Baths

0+ ▶

[Save search](#) [More filters](#) ▶

Save this search and get email alerts for new listings.

Only showing 500 homes. Zoom in, or use filters to narrow your search.

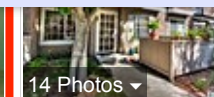


# Unsupervised Domain-specific data extraction

Contact

Want to know more or have an enquiry? Send us a message...

Find out more...

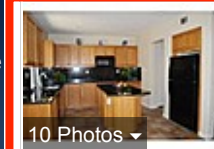


Zestimate®: \$415,125  
Est. Mortgage: \$2,048/mo  
Zillow Special Offer: **\$3,000**

Baths: 2.5  
Sqft: 1,383  
Lot: --

Built: 1984  
Condo  
Price/sqft: \$332

[Save](#)  
[Comps](#)



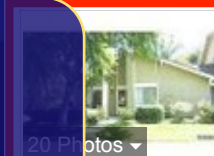
**111 Waterman, Irvine, CA** (Lower Peters Canyon)  
 **House For Sale: \$598,800**  
Zestimate®: --  
Est. Mortgage: \$2,669/mo  
[See current rates](#)

Beds: 4  
Baths: 3.0  
Sqft: 1,728  
Lot: --

Days on Zillow: 2  
Built: 2001  
Single Family  
Price/sqft: \$346

[Maps](#)  
[Save](#)  
[Comps](#)

FEATURED



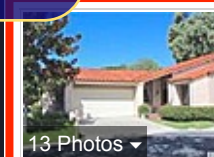
**11 Denver # 85, Irvine, CA** (El Camino Real)  
 **Condo For Sale: \$389,000**  
Zestimate®: \$331,696  
Est. Mortgage: \$1,734/mo  
[See current rates](#)

Beds: 3  
Baths: 1.5  
Sqft: 1,223  
Lot: --

Days on Zillow: 2  
Built: 1977  
Condo  
Price/sqft: \$318

[Maps](#)  
[Save](#)  
[Comps](#)

FEATURED  
COLDWELL BANKER



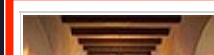
**16 Tahoe, Irvine, CA** (University Park)  
 **House For Sale: \$539,500**  
Zestimate®: \$541,458  
Est. Mortgage: \$2,405/mo  
[See current rates](#)

Beds: 2  
Baths: 2.0  
Sqft: 1,540  
Lot: 3,600

Days on Zillow: 2  
Built: 1976  
Single Family  
Price/sqft: \$350

[Maps](#)  
[Save](#)  
[Comps](#)

FEATURED  
FIRSTTEAM



**30 Blue Heron, Irvine, CA** (Shady Canyon)

FEATURED



	scope	input	output	supervision
—	page scope (single page)	structured page	table or ontology	yes
<b>supervised, site-specific data extraction</b>	site scope (multiple pages from same site)	structured page	table or ontology	yes
<b>unsupervised, domain-specific data extraction</b>	domain scope (multiple sites from same domain)	structured page	table or ontology	no
<b>information extraction</b>	web scope	flat text	entities or entities and relations	no



# Current Challenges in Data Extraction

- Effective **use of visual information**
  - layout carries significant semantic information, yet rarely exploited
- Adapting to **scripted, interactive web** applications
  - interaction is more than form filling
- Effective use of **semantic knowledge**
  - to improve robustness and automation of DE
- The role of **domain-specific knowledge**
  - everyone agrees it would be useful
  - but many see domain-dependence as too high a price
  - **goal:** minimize effort for creating and maintaining domain knowledge



# Current Challenges in Data Extraction

- Effective **use of visual information**
  - layout carries significant semantic information, yet rarely exploited
- Adapting to **scripted, interactive web** applications
  - interaction is more than form filling
- Effective use of **semantic knowledge**
  - to improve robustness and automation of DE
- The role of
  - everyone
  - but many
  - goal:** mir

*Because data on the **Web** is about everything, any approach that attempts to leverage it cannot rely on building a model of the data ahead of time but on **domain-independent methods** instead.*

Cafarella, Halevy, and Madhavan. 2011. **Structured data on the web.** *Comm. ACM* 54, 2.



	Knowledge			
	– (site-specific)	Schema	Instances	Patterns
Visual	<div>SXPath</div> <div>XPath + Visual Relations</div>	<div>SCRAP</div> <div>Robust wrappers by schema</div>	<div>WRAPPO</div> <div>Visual patterns and annotations</div>	
	<div>OXPath</div> <div>XPath + Actions</div>	<div>Lixto</div> <div>Visual designer + expressive extraction language Elog</div>	<div>DIADEM</div> <div>Form understanding, object identification, and exploration driven by domain knowledge</div>	
Interaction				

40



# Part I:

# Supervised, site-specific



	Knowledge			
	– (site-specific)	Schema	Instances	Patterns
Visual	<div>SXPath</div> <div>XPath + Visual Relations</div>	<div>SCRAP</div> <div>Robust wrappers by schema</div>	<div>WRAPPO</div> <div>Visual patterns and annotations</div>	<div>DIADEM</div> <div>Form understanding, object identification, and exploration driven by domain knowledge</div>
Interaction	<div>OXPath</div> <div>XPath + Actions</div>	<div>Lixto</div> <div>Visual designer + expressive extraction language Elog</div>		



# Part I: Introduction

## **Wrapper Induction**



# Wrapper Induction

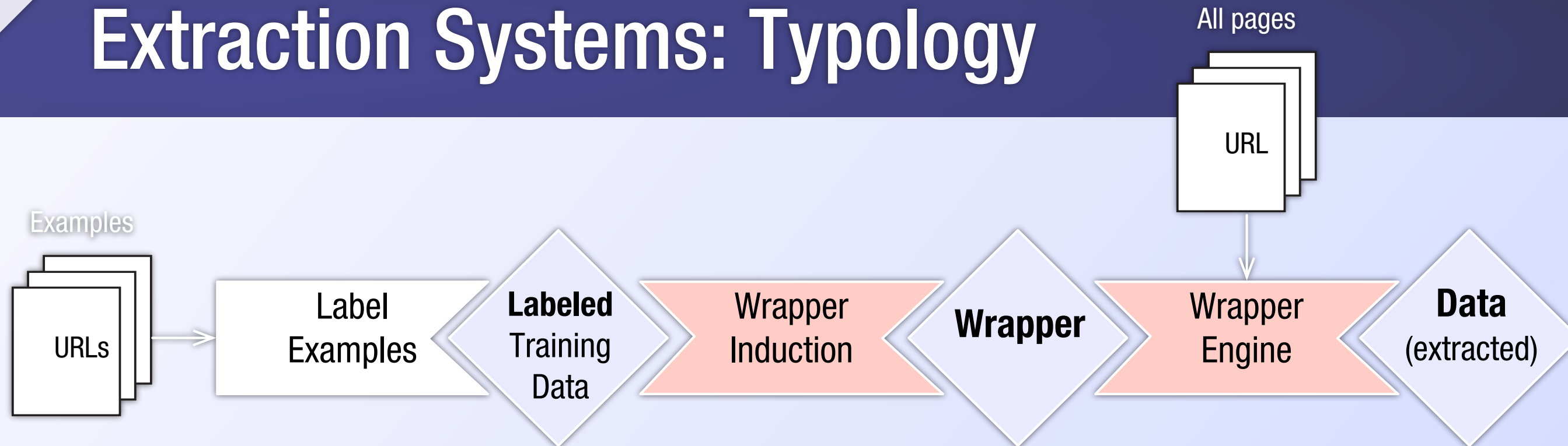
- **Three types** of tasks in wrapper induction:
  - finding the relevant **pages** interaction (forms)
  - identifying the relevant **objects** segmentation
  - extracting the relevant **attributes** alignment
- In all cases: **derive patterns from examples**



**Problem:**  
How to derive  
those patterns?



# Extraction Systems: Typology



- **manual:** (e.g., Web Harvest)
  - user writes the wrapper, sometimes using wrapping libraries
- **supervised/semi-supervised:** (e.g., Lixto)
  - **user** provides examples (per site)
  - **wrapper induction:** supervised to some degree
- **unsupervised:** entirely automated (e.g., DIADEM)
  - automatically guess examples & automatic induction



# Extraction Languages

- Wrapper ind. systems differ in supervision, but also language
- Typical wrapper languages
  - regular expressions on a canonical textual representation
    - e.g., suffix, prefix in HTML
  - XPath: by now the most common choice
  - relational algebra or logical expressions as in Elog or DIADEM



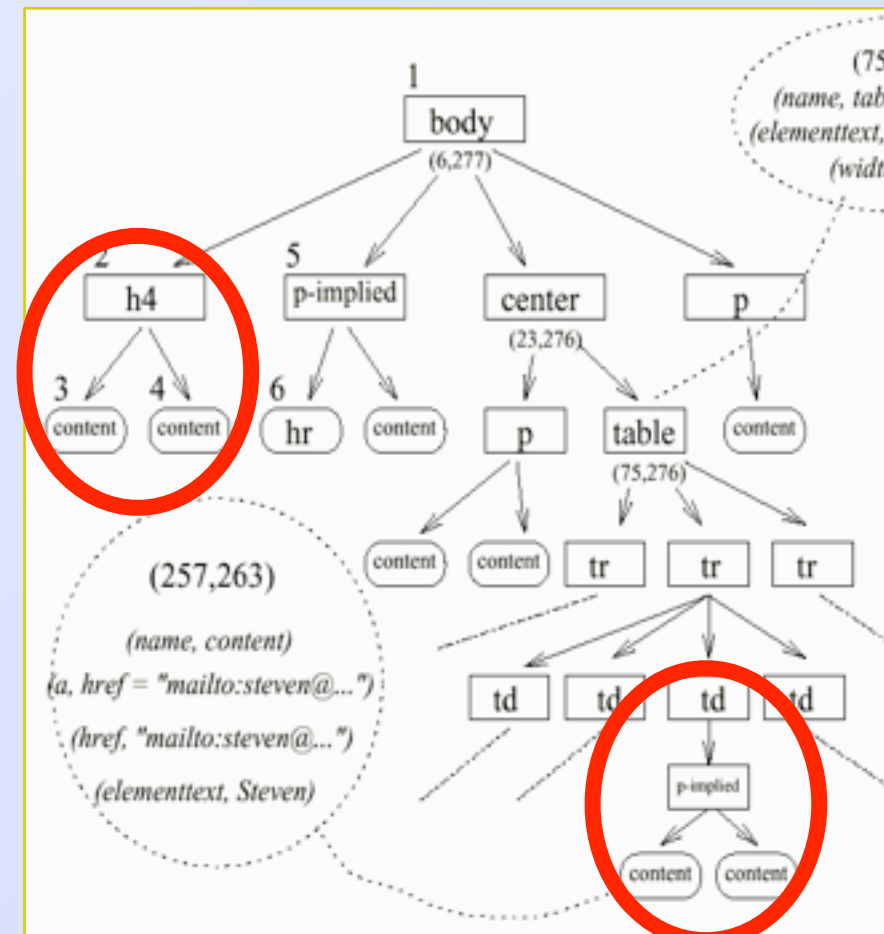
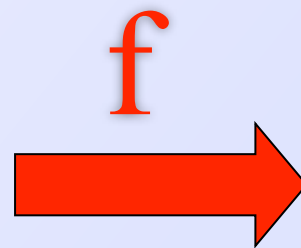
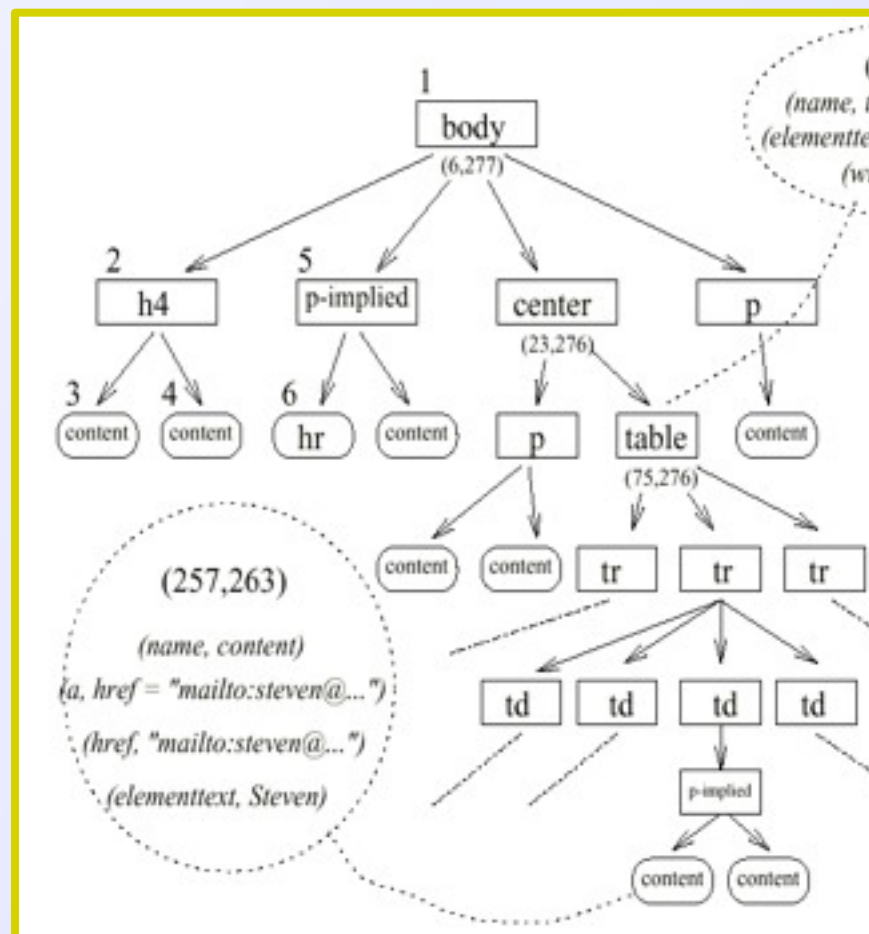
# Desiderata of a Web Wrapping Language

- Solid and **well understood theoretical foundation**
  - wrappers run much longer than DB queries
  - thus optimization and verification is even more important
- Good trade-off between **complexity** and the number of **practical wrappers**
- Easy to use as a wrapper programming language
  - use of **familiar** web query languages such as XPath a plus
- Suitable for incorporation into **visual tools**
  - where constructs of a wrapping language are realized through corresponding visual primitives
- Allows the formulation of **robust wrappers**



# Extraction from Tree Structure

- Wrapper as **function f**: HTML DOM Tree  $\rightarrow$  Subtrees
  - leaves of subtrees are among leaves of original tree
  - simplest case: subtrees always single leaves
  - extracted subtrees may be rearranged into new output tree or graph





# XPath (XML Path Language)

- Queries an XML document based on its DOM representation
- Notation inspired by file system path
- Foundational language for XQuery, XSLT
- Used extensively with JavaScript/web applications



# XPath Example

- On this document:

```
<?xml version="1.0"?>
<people>
  <person>
    <familyName>Obama</familyName>
    <givenName>Barack</givenName>
  </person>
  <person>
    <familyName>Bush</familyName>
    <givenName>George</givenName>
  </person>
</people>
```

- ... the following XPath expression

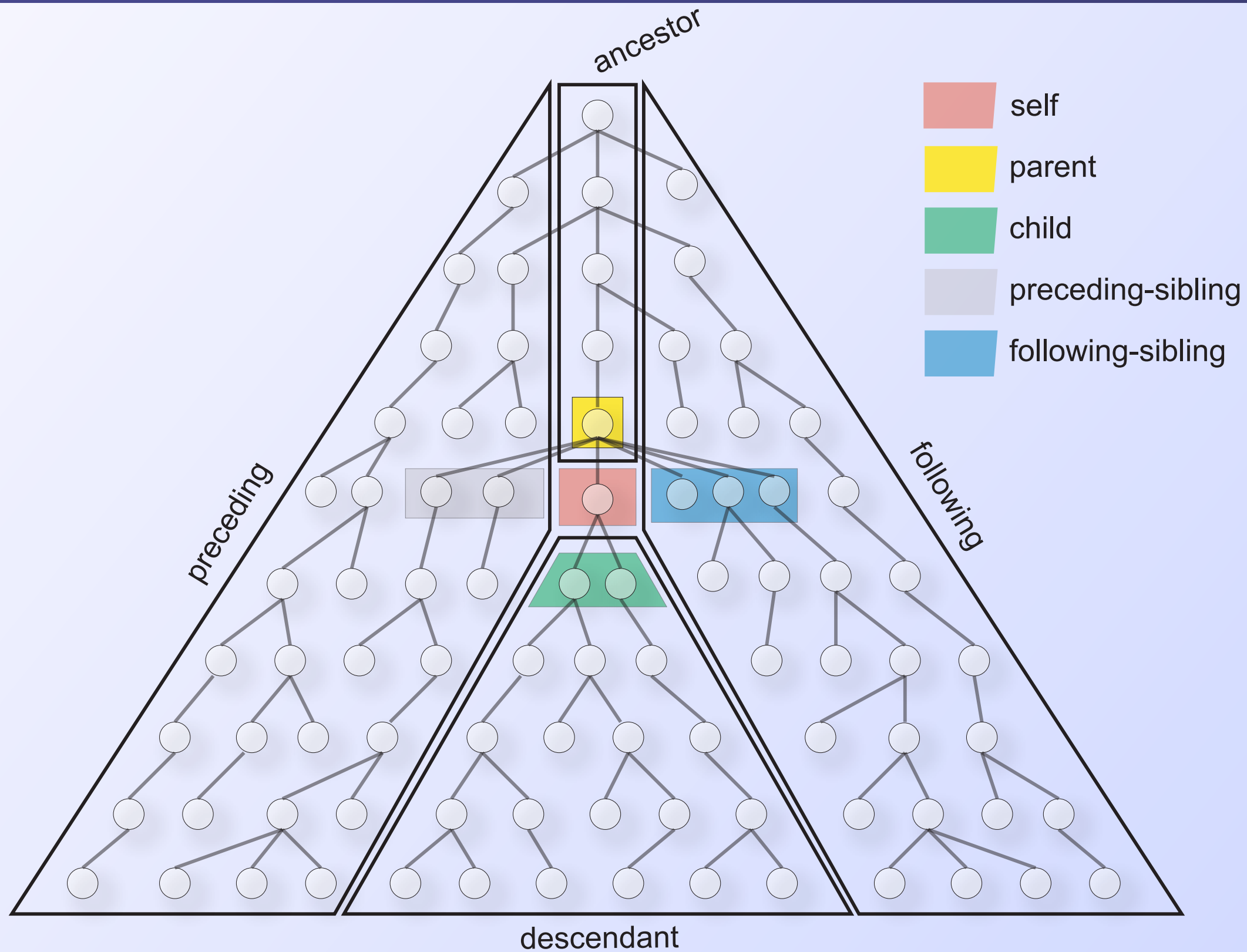
```
/child::people/child::person[1]/familyName
```

- ... returns

```
<familyName>Obama</familyName>
```



# XPath Axes





	Knowledge			
	– (site-specific)	Schema	Instances	Patterns
Visual	<div>SXPath</div> <div>XPath + Visual Relations</div>	<div>SCRAP</div> <div>Robust wrappers by schema</div>	<div>WRAPPO</div> <div>Visual patterns and annotations</div>	<div>DIADEM</div> <div>Form understanding, object identification, and exploration driven by domain knowledge</div>
Interaction	<div>OXPath</div> <div>XPath + Actions</div>	<div>Lixto</div> <div>Visual designer + expressive extraction language Elog</div>		



# **XPath**

## **Actions in XPath**



# OXPath

Furche, Gottlob, Grasso, Schallhart and Sellers. **OXPath: A Language for Scalable, Memory-efficient Data Extraction from Web Applications.** *VLDB*, 2011

Furche, Gottlob, Grasso, Schallhart, and Sellers. **OXPATH: A Language for Scalable Data Extraction, Automation, and Crawling on the Deep Web.** In *VLDB J Special Issue*. To appear 2013.



# **XPath**

## **Actions in XPath**



$$\text{XPath} = \text{XPath} + 4$$

action

extraction

style

iteration



KAYAK

SEARCH ONE AND DONE

Flights

Hotels

Cars

Holidays

Explore

More

☒ Return ☐ One-way ☐ Multi-city

From

☐ Add nearby airports [custom](#)

To

☐ Add nearby airports [custom](#)



[My dates are flexible](#)

Depart

01/10/2011  Anytime   
Sat, 1 Oct 2011

Return

04/10/2011  Anytime   
Tue, 4 Oct 2011

1 traveller  Economy  ☐ Non-stops Only


☐ Find Hotels (in new window)

Compare hundreds of travel sites at once.

Search

Flexible Dates

Best Fares: LON to SEA

Oct 2011 

Mon	Tue	Wed	Thu	Fri	Sat	Sun
					1 £522	2 £508
3 £508	4 £528	5	6 £538	7	8	9
10 £488	11	12 £538	13 £538	14 £618	15 £528	16
17	18	19	20	21 £517	22 £534	23
24	25	26	27	28	29 £486	30
31 £458						

[About these prices](#)

Your Search History

LON to SEA

1 Oct - 4 Oct

[modify](#)



- Start at kayak.co.uk:  
`doc("kayak.co.uk")`

KAYAK

SEARCH ONE AND DONE

Flights

Hotels

Cars

Holidays

Explore

More

☒ Return

☐ One-way

☐ Multi-city

From

☐ Add nearby airports

custom

My dates are flexible

Depart

01/10/2011

Anytime

⌵

Sat, 1 Oct 2011

1 traveller

⌵

Economy

⌵

☐ Non-stops Only

☐ Find Hotels (in new window)

Compare hundreds of travel sites at once.

Search

Flexible Dates

Best Fares: LON to SEA

Oct 2011

⌵

Mon	Tue	Wed	Thu	Fri	Sat	Sun
					1 £522	2 £508
3 £508	4 £528	5	6 £538	7	8	9
10 £488	11	12 £538	13 £538	14 £618	15 £528	16
17	18	19	20	21 £517	22 £534	23
24	25	26	27	28	29 £486	30
31 £458						

About these prices

Your Search History

⌵

LON to SEA

1 Oct - 4 Oct

modify



- Start at kayak.co.uk:  
doc("kayak.co.uk")
- To select an airport, type a few letters and select from completion list  
//field().destination/{"Sea" /}  
//div#smartbox//li[1]/{click /}

KAYAK SEARCH ONE AND DONE

[Flights](#)
[Hotels](#)
[Cars](#)
[Holidays](#)
[Explore](#)
[More](#)

☒ Return
☐ One-way
☐ Multi-city

**From**  
  
☐ Add nearby airports [custom](#)

**To**  
  
☐ Add nearby airports [custom](#)

[My dates are flexible](#)

**Depart**  
01/10/2011 Anytime   
Sat, 1 Oct 2011

**Return**  
04/10/2011 Anytime   
Tue, 4 Oct 2011

1 traveller 
Economy 
☐ Non-stops Only

☐ Find Hotels (in new window)

Compare hundreds of travel sites at once.

Search

**Flexible Dates**

**Best Fares: LON to SEA** Oct 2011

Mon	Tue	Wed	Thu	Fri	Sat	Sun
					1 £522	2 £508
3 £508	4 £528	5	6 £538	7	8	9
10 £488	11	12 £538	13 £538	14 £618	15 £528	16
17	18	19	20	21 £517	22 £534	23
24	25	26	27	28	29 £486	30
31 £458						

[About these prices](#)

**Your Search History**

LON to SEA 1 Oct - 4 Oct [modify](#)



- Start at kayak.co.uk:  
`doc("kayak.co.uk")`
- To select an airport, type a few letters and select from completion list  
`//field().destination/{ "Sea" / }`  
`//div#smartbox//li[1]/{click / }`
- Submit the form

KAYAK

SEARCH ONE AND DONE

[Flights](#)
[Hotels](#)
[Cars](#)
[Holidays](#)
[Explore](#)
[More](#)

☒ Return
 ☐ One-way
 ☐ Multi-city

From

London, ENG, United Kingdom - All air

☐ Add nearby airports
 [custom](#)

My dates are flexible

Depart

01/10/2011

Anytime

Sat, 1 Oct 2011

1 traveller

Economy

☐ Find Hotels (in new window)

Choose Sites to Compare -vs- KAYAK (new windows)

☐ eDreams.co.uk
 ☐ TravelPack
 ☐ ebookers.com

To

Sea

☒ Seattle, WA, United States - Seattle/Tacoma Intl (SEA)
 ☐ Seattle, WA, United States - King County/Boeing (BFI)
 ☐ Seattle, WA, United States - Lake Union SPB (LKE)
 ☐ Seal Bay, AK, United States - Seal Bay (SYB)
 ☐ Stillwater, OK, United States - Searcy Fld (SWO)

Show airports on map:

☒ Seaside, OR, United States
 ☒ Seaside Heights, NJ, United States
 ☒ Seaside, CA, United States
 ☒ Searcy, AR, United States
 ☒ SeaTac, WA, United States

Recently searched:

☒ Seattle, WA, United States - Seattle/Tacoma Intl (SEA)

Flexible Dates

Best Fares: LON to SEA Oct 2011

Mon	Tue	Wed	Thu	Fri	Sat	Sun
					1 £522	2 £508
		5	6 £538	7	8	9
12 £538	13 £538	14 £618	15 £528	16		
19	20	21 £517	22 £534	23		
26	27	28	29 £486	30		

[About these prices](#)

History

LON to SEA

1 Oct - 4 Oct

modify





London, ENG, United Kingdom ▶ Seattle, WA, United States

01/10/2011 ▶ 04/10/2011

[Search](#) [more options](#)

### [Change your search](#)

[Get a price alert](#)

[Show fare charts](#)

#### Stops

- ☒ non-stop £547
- ☒ 1 stop £503
- ☒ 2+ stops £558

#### Times

- ☒ Show Red Eye/Overnight
- ☒ Take-off ☐ Landing
- Take-off (Depart Flight)  
Sat 06:00 - 20:00



Take-off (Return Flight)  
Tue 06:00 - Wed 00:00



2110 of 2110 return trips shown

[Show +/- 3 days](#) [Show matrix](#)

Price * ▲	Airline	Take-off	Landing	Stops (🕒)
-----------	---------	----------	---------	-----------

### [Airline Tickets To Seattle](#)

Book your Seattle flight & get our lowest fares at [united.com](#)  
[www.united.com](#)

**£503**

Economy

[Select](#)



US Airways

LGW 09:40 ▶ SEA 20:55 1 19h 15m  
SEA 11:15 ▶ LHR 10:00 1 14h 45m

Top Deals:

[Majortravel.co.uk](#) **£503**

[Usairways.com](#) **£509**

[Lastminute.com](#) **£508**

[Expedia.co.uk](#) **£508**

found on 4 Websites

[details ▶](#)

[pin ▶](#)

[share ▶](#)



**£503**

Economy

[Select](#)



US Airways

LGW 09:40 ▶ SEA 20:55 1 19h 15m  
SEA 08:40 ▶ LHR 10:00 1 17h 20m

Top Deals:

[Majortravel.co.uk](#) **£503**

[Expedia.co.uk](#) **£508**

[Lastminute.com](#) **£508**

[Flights To Seattle](#)

Fares Just Dropped

Flights Starting at

[www.lowfares.com](#)



Refine the results by unchecking the “2+ stops”:

`//*#stops2/{uncheck }`

**KAYAK**

London, ENG, United Kingdom ▶ Seattle, WA, United States

01/10/2011 ▶ 04/10/2011

Search [more options](#)

[Change your search](#)  
[Get a price alert](#)  
[Show fare charts](#)

**Stops**  
☒ non-stop £547  
☒ 1 stop £503  
☒ 2+ stops £558

**Times**  
☒ Show Red Eye/Overnight  
☒ Take-off ☐ Landing  
Take-off (Depart Flight)  
Sat 06:00 - 20:00  
Take-off (Return Flight)  
Tue 06:00 - Wed 00:00

2110 of 2110 return trips shown [Show +/- 3 days](#) [Show matrix](#)

Price * ▲	Airline	Take-off	Landing	Stops (🕒)
<b>Airline Tickets To Seattle</b> <small>ad</small> Book your Seattle flight & get our lowest fares at united.com <a href="http://www.united.com">www.united.com</a>				
<b>£503</b> Economy <b>Select</b>	US Airways	LGW 09:40 ▶ SEA 20:55 SEA 11:15 ▶ LHR 10:00	1 1	19h 15m 14h 45m
Top Deals: <a href="#">Majortravel.co.uk £503</a> <a href="#">Lastminute.com £508</a> <a href="#">Usairways.com £509</a> <a href="#">Expedia.co.uk £508</a>				
found on 4 Websites <a href="#">details ▶</a> <a href="#">pin ▶</a> <a href="#">share ▶</a> <small>!</small>				
<b>£503</b> Economy <b>Select</b>	US Airways	LGW 09:40 ▶ SEA 20:55 SEA 08:40 ▶ LHR 10:00	1 1	19h 15m 17h 20m
Top Deals: <a href="#">Majortravel.co.uk £503</a> <a href="#">Lastminute.com £508</a> <a href="#">Expedia.co.uk £508</a>				

[Flights To Seattle](#)  
Fares Just Dropped  
Flights Starting at  
[www.lowfares.com](http://www.lowfares.com)



- Refine the results by unchecking the “2+ stops”:

//\*#stops2/{**uncheck** }

- On all result pages

/ (//a[.='Next']/**{click /}**)\*

**KAYAK**

London, ENG, United Kingdom ▶ Seattle, WA, United States

01/10/2011 ▶ 04/10/2011

**Search** [more options](#)

[Change your search](#)  
[Get a price alert](#)  
[Show fare charts](#)

**Stops**  
☒ non-stop £547  
☒ 1 stop £503  
☒ 2+ stops £558

**Times**  
☒ Show Red Eye/Overnight  
☒ Take-off ☐ Landing  
Take-off (Depart Flight)  
Sat 06:00 - 20:00  
Take-off (Return Flight)  
Tue 06:00 - Wed 00:00

2110 of 2110 return trips shown [Show +/- 3 days](#) [Show matrix](#)

Price * ▲	Airline	Take-off	Landing	Stops (🕒)
<b>Airline Tickets To Seattle</b> <small>ad</small> Book your Seattle flight & get our lowest fares at united.com <a href="http://www.united.com">www.united.com</a>				
<b>£503</b> Economy <b>Select</b>	US Airways	LGW 09:40 ▶ SEA 20:55 SEA 11:15 ▶ LHR 10:00	1 1	19h 15m 14h 45m
Top Deals: <a href="#">Majortravel.co.uk £503</a> <a href="#">Lastminute.com £508</a> <a href="#">Usairways.com £509</a> <a href="#">Expedia.co.uk £508</a>				
found on 4 Websites <b>details ▶</b> <b>pin ▶</b> <b>share ▶</b> <small>!</small>				
<b>£503</b> Economy <b>Select</b>	US Airways	LGW 09:40 ▶ SEA 20:55 SEA 08:40 ▶ LHR 10:00	1 1	19h 15m 17h 20m
Top Deals: <a href="#">Majortravel.co.uk £503</a> <a href="#">Lastminute.com £508</a> <a href="#">Expedia.co.uk £508</a>				

[Flights To Seattle](#)  
Fares Just Dropped  
Flights Starting at  
[www.lowfares.com](http://www.lowfares.com)



- Refine the results by unchecking the “2+ stops”:

```
//*#stops2/{uncheck }
```

- On all result pages

```
/ ( //a[.='Next']/{click /} ) *
```

- and for each flight

```
//body.resultrow:<flight>
```

**KAYAK**

London, ENG, United Kingdom → Seattle, WA, United States

01/10/2011 → 04/10/2011 [Search](#) [more options](#)

[Change your search](#)  
[Get a price alert](#)  
[Show fare charts](#)

**Stops**

- ☒ non-stop £547
- ☒ 1 stop £503
- ☒ 2+ stops £558

**Times**

- ☒ Show Red Eye/Overnight
- ☒ Take-off ☐ Landing
- Take-off (Depart Flight): Sat 06:00 - 20:00
- Take-off (Return Flight): Tue 06:00 - Wed 00:00

2110 of 2110 return trips shown [Show +/- 3 days](#) [Show matrix](#)

Price	Airline	Take-off	Landing	Stops
<b>Airline Tickets To Seattle</b> Book your Seattle flight & get our lowest fares at united.com <a href="http://www.united.com">www.united.com</a>				
<b>£503</b> Economy <a href="#">Select</a>	US Airways	LGW 09:40 → SEA 20:55 SEA 11:15 → LHR 10:00	1 1	19h 15m 14h 45m
<b>Top Deals:</b> <a href="#">Majortravel.co.uk £503</a> <a href="#">Lastminute.com £508</a> <a href="#">Usairways.com £509</a> <a href="#">Expedia.co.uk £508</a>				
found on 4 Websites <a href="#">details</a> <a href="#">pin</a> <a href="#">share</a>				
<b>£503</b> Economy <a href="#">Select</a>	US Airways	LGW 09:40 → SEA 20:55 SEA 08:40 → LHR 10:00	1 1	19h 15m 17h 20m
<b>Top Deals:</b> <a href="#">Majortravel.co.uk £503</a> <a href="#">Lastminute.com £508</a> <a href="#">Expedia.co.uk £508</a>				

[Flights To Seattle](#)  
 Fares Just Dropped  
 Flights Starting at [www.lowfares.com](http://www.lowfares.com)





London, ENG, United Kingdom ▶ Seattle, WA, United States

01/10/2011 ▶ 04/10/2011

[Search](#) [more options](#)

[Change your search](#)

[Get a price alert](#)

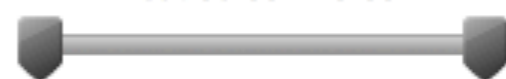
[Show fare charts](#)

**Stops**

- ☒ non-stop £547
- ☒ 1 stop £503
- ☒ 2+ stops £558

**Times**

- ☒ Show Red Eye/Overnight
- ☒ Take-off ☐ Landing
- Take-off (Depart Flight)
- Sat 06:00 - 20:00



Take-off (Return Flight)

Tue 06:00 - Wed 00:00



2110 of 2110 return trips shown

[Show +/- 3 days](#) [Show matrix](#)

Price * ▲	Airline	Take-off	Landing	Stops (🕒)
-----------	---------	----------	---------	-----------

[Airline Tickets To Seattle](#)

Book your Seattle flight & get our lowest fares at united.com  
[www.united.com](#)

**£503**  
Economy

[Select](#)

US Airways

**Top Deals:** [Majortravel.co.uk £503](#) [Lastminute.com £508](#)  
[Usairways.com £509](#) [Expedia.co.uk £508](#)

found on 4 Websites [details ▶](#) [pin ▶](#) [share ▶](#)

**LGW** 09:40 ▶ **SEA** 20:55 1 19h 15m

**SEA** 11:15 ▶ **LHR** 10:00 1 14h 45m

**£503**  
Economy

[Select](#)

US Airways

**Top Deals:** [Majortravel.co.uk £503](#) [Lastminute.com £508](#)  
[Expedia.co.uk £508](#)

**LGW** 09:40 ▶ **SEA** 20:55 1 19h 15m

**SEA** 08:40 ▶ **LHR** 10:00 1 17h 20m

[Flights To Seattle](#)  
Fares Just Dropped  
Flights Starting at  
[www.lowfares.com](#)





London, ENG, United Kingdom ▶ Seattle, WA, United States

01/10/2011 ▶ 04/10/2011

[Search](#) [more options](#)

[Change your search](#)

[Get a price alert](#)

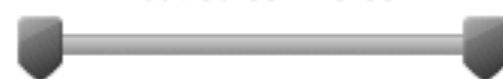
[Show fare charts](#)

**Stops**

- ☒ non-stop £547
- ☒ 1 stop £503
- ☒ 2+ stops £558

**Times**

- ☒ Show Red Eye/Overnight
- ☒ Take-off ☐ Landing
- Take-off (Depart Flight)
- Sat 06:00 - 20:00



Take-off (Return Flight)

Tue 06:00 - Wed 00:00



2110 of 2110 return trips shown

[Show +/- 3 days](#) [Show matrix](#)

Price * ▲	Airline	Take-off	Landing	Stops (⌚)
-----------	---------	----------	---------	-----------

[Airline Tickets To Seattle](#)

Book your Seattle flight & get our lowest fares at united.com  
[www.united.com](#)

**£503**  
Economy

[Select](#)

US Airways

LGW 09:40 ▶ SEA 20:55 1 19h 15m

SEA 11:15 ▶ LHR 10:00 1 14h 45m

Top Deals:

[Majortravel.co.uk £503](#) [Lastminute.com £508](#)

[Usairways.com £509](#) [Expedia.co.uk £508](#)

found on 4 Websites

[details ▶](#) [pin ▶](#) [share ▶](#)

**£503**  
Economy

[Select](#)

US Airways

LGW 09:40 ▶ SEA 20:55 1 19h 15m

SEA 08:40 ▶ LHR 10:00 1 17h 20m

Top Deals:

[Majortravel.co.uk £503](#) [Lastminute.com £508](#)

[Expedia.co.uk £508](#)

[Flights To Seattle](#)  
Fares Just Dropped  
Flights Starting at  
[www.lowfares.com](#)



## Extract the attributes

London, ENG, United Kingdom

Seattle, WA, United States

01/10/2011

04/10/2011

Search

[more options](#)

[Change your search](#)

[Get a price alert](#)

[Show fare charts](#)

**Stops**

☒ non-stop £547

☒ 1 stop £503

☒ 2+ stops £558

**Times**

☒ Show Red Eye/Overnight

☒ Take-off ☐ Landing

Take-off (Depart Flight)

Sat 06:00 - 20:00

Take-off (Return Flight)

Tue 06:00 - Wed 00:00

2110 of 2110 return trips shown

[Show +/- 3 days](#) [Show matrix](#)

Price * ▲	Airline	Take-off	Landing	Stops (🕒)
<div> <a href="#">Airline Tickets To Seattle</a> <span>ad</span> </div> <div> Book your Seattle flight &amp; get our lowest fares at united.com  <a href="http://www.united.com">www.united.com</a> </div>				
<div> <div>£503</div> <div>Economy</div> <div>Select</div> </div>	<div>  US Airways </div>	<div> <div>LGW 09:40 ▶ SEA 20:55</div> <div>SEA 11:15 ▶ LHR 10:00</div> </div>	<div> <div>1 19h 15m</div> <div>1 14h 45m</div> </div>	
<div> Top Deals: <div> <a href="#">Majortravel.co.uk £503</a> <a href="#">Lastminute.com £508</a> </div> <div> <a href="#">Usairways.com £509</a> <a href="#">Expedia.co.uk £508</a> </div> </div>				
<div> found on 4 Websites <div> <div>details ▶</div> <div>pin ▶</div> <div>share ▶</div> </div> </div>				
<div> <div>£503</div> <div>Economy</div> <div>Select</div> </div>	<div>  US Airways </div>	<div> <div>LGW 09:40 ▶ SEA 20:55</div> <div>SEA 08:40 ▶ LHR 10:00</div> </div>	<div> <div>1 19h 15m</div> <div>1 17h 20m</div> </div>	
<div> Top Deals: <div> <a href="#">Majortravel.co.uk £503</a> <a href="#">Lastminute.com £508</a> </div> <div> <a href="#">Expedia.co.uk £508</a> </div> </div>				

[Flights To Seattle](#)

Fares Just Dropped

Flights Starting at

[www.lowfares.com](http://www.lowfares.com)



- Extract the attributes
- Mouseover the ! to extract flight quality warnings  

```
//span.qualityWarningIcon/{mouseover /}
```

KAYAK

London, ENG, United Kingdom

Seattle, WA, United States

01/10/2011

04/10/2011

Search

more options

Change your search

Get a price alert

Show fare charts

Stops

☒ non-stop £547
 ☒ 1 stop £503
 ☒ 2+ stops £558

Times

☒ Show Red Eye/Overnight
 ☒ Take-off ☐ Landing

Take-off (Depart Flight)

Sat 06:00 - 20:00

Take-off (Return Flight)

Tue 06:00 - Wed 00:00

2110 of 2110 return trips shown

[Show +/- 3 days](#)
[Show matrix](#)

Price * ▲	Airline	Take-off	Landing	Stops (🕒)
<div> <div>Airline Tickets To Seattle</div> <div>Book your Seattle flight &amp; get our lowest fares at united.com</div> <div>www.united.com</div> </div>				
<div>£503</div> <div>Economy</div> <div>Select</div>	<div>US Airways</div>	<div>LGW 09:40 ▶ SEA 20:55</div> <div>SEA 11:15 ▶ LHR 10:00</div>	<div>1 19h 15m</div> <div>1 14h 45m</div>	<div>Top Deals:</div> <div> <a href="#">Majortravel.co.uk £503</a> <a href="#">Lastminute.com £508</a> </div> <div> <a href="#">Usairways.com £509</a> <a href="#">Expedia.co.uk £508</a> </div> <div> <div>found on 4 Websites</div> <div> <div>details ▶</div> <div>pin ▶</div> <div>share ▶</div> </div> </div>
<div>£503</div> <div>Economy</div> <div>Select</div>	<div>US Airways</div>	<div>LGW 09:40 ▶ SEA 20:55</div> <div>SEA 08:40 ▶ LHR 10:00</div>	<div>1 19h 15m</div> <div>1 17h 20m</div>	<div>Top Deals:</div> <div> <a href="#">Majortravel.co.uk £503</a> <a href="#">Lastminute.com £508</a> </div> <div> <a href="#">Expedia.co.uk £508</a> </div>

Flight Quality

🕒 A stopover is more than 4 h

📊 Less than 75% on time

👁 Red Eye (overnight flight)

[See all flight details](#)

Flights Starting at

www.lowfares.com

58



- Extract the attributes
- Mouseover the ! to extract flight quality warnings  

```
//span.qualityWarningIcon/{mouseover /}
```
- Click on the details to extract layovers

The screenshot shows a flight search results page for a route from London, ENG, United Kingdom to Seattle, WA, United States. The page displays a multi-airport itinerary with two flights: Flight 1 (Sat, 1 Oct) from LGW to SEA via CLT, and Flight 2 (Tue, 4 Oct) from SEA to LHR via CLT. The itinerary is highlighted with a yellow background. A red circle highlights the '2+ stops' option in the 'Stops' filter. A red circle highlights the 'Details' link for the first flight. A red circle highlights the '!' icon in the 'Flight Quality' section. The 'Flight Quality' section shows a warning: 'A stopover is more than 4 h'. The 'Details' section shows the flight details for the first flight, including the airline (US Airways - Flights 733), departure time (Sat 9:40), arrival time (Sat 13:35), and connection details (CLT Charlotte, NC, United States, 4h 35m). The 'Details' section also shows the flight details for the second flight, including the airline (US Airways - Flights 575), departure time (Sat 18:10), arrival time (Sat 20:55), and connection details (SEA Seattle, WA, United States, 9h 05m). The 'Details' section also shows the flight details for the third flight, including the airline (US Airways - Flights 575), departure time (Sat 18:10), arrival time (Sat 20:55), and connection details (SEA Seattle, WA, United States, 9h 05m). The 'Details' section also shows the flight details for the fourth flight, including the airline (US Airways - Flights 575), departure time (Sat 18:10), arrival time (Sat 20:55), and connection details (SEA Seattle, WA, United States, 9h 05m).

**Flight 1 Sat, 1 Oct** LGW to SEA – 1 stop 19h 15m

This is a multi-airport itinerary. Please review all the airports below.

**US Airways – Flights 733** 14h 15m

**Take-off Sat 9:40** LGW London, ENG, United Kingdom  
**Landing Sat 13:35** CLT Charlotte, NC, United States

Coach | Airbus A330-300 (Wide-body Jet) | 14h 15m | 3998 miles

**Connection** CLT Charlotte, NC, United States 4h 35m

**US Airways – Flights 575** 9h 05m

**Take-off Sat 18:10** CLT Charlotte, NC, United States  
**Landing Sat 20:55** SEA Seattle, WA, United States

Coach | Airbus A320-100/200 (Narrow-body Jet) | 9h 05m | 50% on time | 2273 miles

**Flight 2 Tue, 4 Oct** SEA to LHR – 1 stop 14h 45m

**Flight Quality**

- A stopover is more than 4 h
- Less than 75% on time
- Red Eye (overnight flight)

[See all flight details](#)

Flights Starting at [www.lowfares.com](#)

Top Deals: [Majortravel.co.uk £503](#) [Expedia.co.uk £508](#) [Lastminute.com £508](#)



# ① Actions: Browser Interaction

- Actions correspond to **DOM events**, e.g.,

Document	<code>doc("rightmove.co.uk")</code>
Click	<code>{click}</code>
Fill	<code>{"Sea"}</code>
Mouseover	<code>{mouseover}</code>

- Executed once on **each context node**
- Return **context nodes** (contextual actions) or **root nodes for new DOM** (absolute actions)



## ② Extraction: Compact Tree Construction

- Extraction marker **select nodes for extraction**
  - record markers: **:<flight>**
  - attribute markers: **:<price=string(.)>**
- Extracted data **has tree shape**
  - nesting of extraction markers** in XPath expression defines **nesting of records** and attribute-record associations in the output



## ② Extraction: Compact Tree Construction

- Extraction marker **select nodes for extraction**
  - record markers: `:<flight>`
  - attribute markers: `:<price=string(.)>`
- Extracted data **has tree shape**
  - nesting of extraction markers** in XPath expression defines **nesting of records** and attribute-record associations in the output

```
doc("news.google.com")//div[@class~="story"]:<story>  
  [./h2:<title=string(.)>]  
  [./span[style::color="#767676"]:<source=string(.)>]
```

```
<story><title>Tax cuts ...</title>  
  <source>Washington Post</source>  
  <source>Wall Street Journal</source> ... </story>
```



## ③ Iteration: Kleene Star

- Most web sites use **pagination for results**
  - traversing paginated results require **iteration**
  - extraction from any **unbounded component** of a link graph
- Kleene Star from Regular XPath [Marx TODS '05]
  - extended to XPath, i.e., with action in the iterated expression
  - `/ (//a[.='Next']/{click /}) *`
  - XPath's Page-at-a-time algorithm
    - buffers in practice only a constant number of pages
    - even for very large components



## ④ Style: Querying Visual Attributes

- Access to all CSS properties via **style axis**

Visibility	<code>style::display</code> or <code>style::visibility</code>
Font size	<code>style::font-size</code>
Geometry	<code>style::top</code> , <code>style::left</code> , ...
Color	<code>style::color</code> or <code>style::background-color</code>

- Joins** on style properties possible
  - but: no rich spatial relations as in XPath



# Summary of Complexity

**Combined: PTIME-hard**

**Data: NLOGSPACE**



# Summary of Complexity

Combined: PTIME-hard PTIME-hard

Data: NLOGSPACE LOGSPACE



# Summary of Complexity

Combined: PTIME-hard PTIME-hard

Data: NLOGSPACE LOGSPACE

	Time	Space
OXPath w/o Actions & Kleene	$O(n^6 \cdot q^2)$	$O(n^5 \cdot q^2)$



OXPath

OXPath

Supervised, Site-specific Data Extraction

Summary of Complexity

Combined: PTIME-hard PTIME-hard

Data: NLOGSPACE LOGSPACE

	Time	Space
OXPath w/o Actions & Kleene	$O(n^6 \cdot q^2)$ $O(n^4 \cdot q^2)$	$O(n^5 \cdot q^2)$ $O(n^3 \cdot q^2)$
OXPath w/o Kleene	$O((p \cdot n)^6 \cdot q^3)$	$O(n^5 \cdot q^3)$
OXPath w/o unbounded Kleene	$O((p \cdot n)^6 \cdot q^3)$	$O(n^5 \cdot q_{\Sigma}^3)$

63



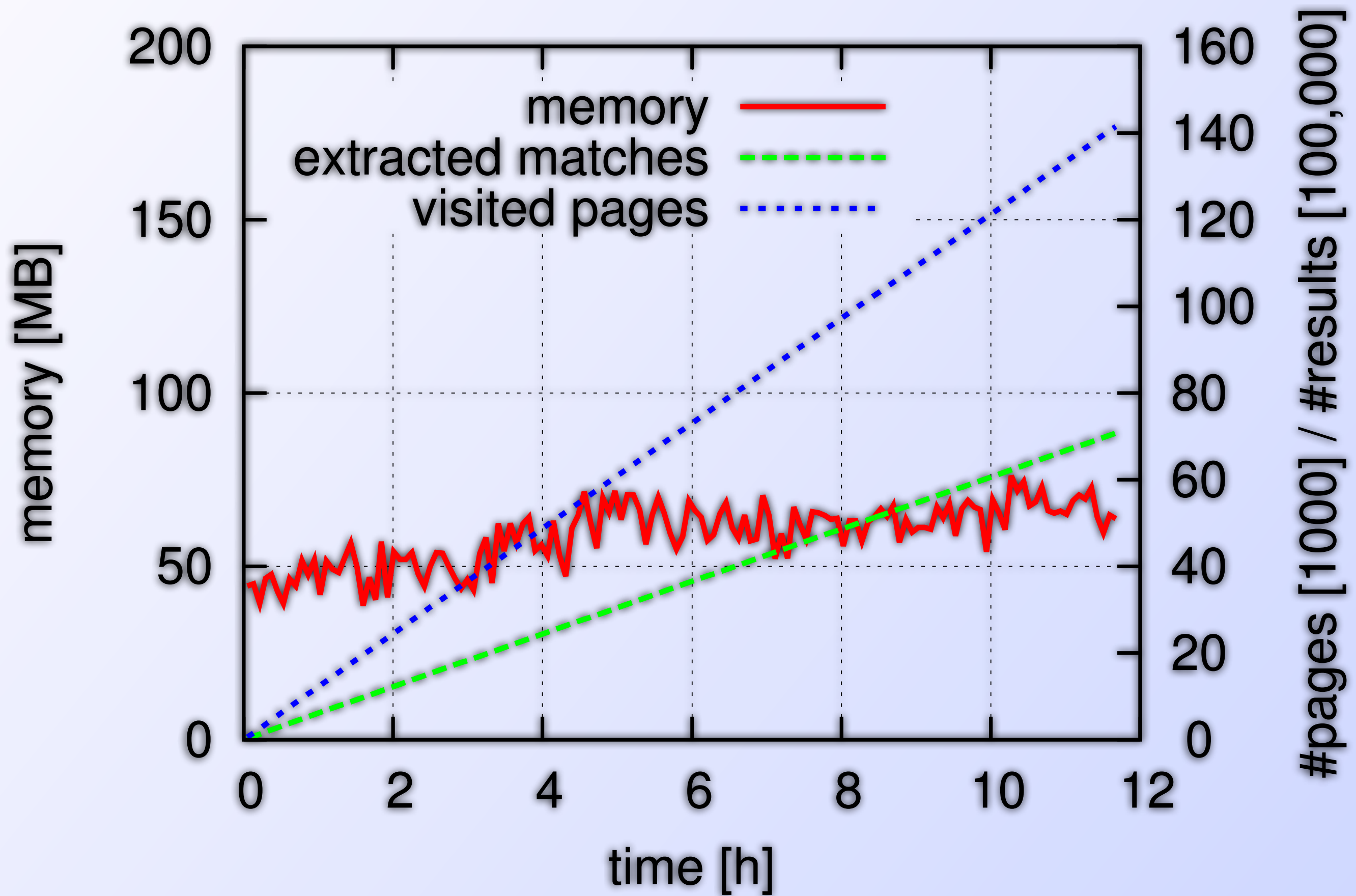
# Summary of Complexity

Combined: PTIME-hard PTIME-hard

Data: NLOGSPACE LOGSPACE

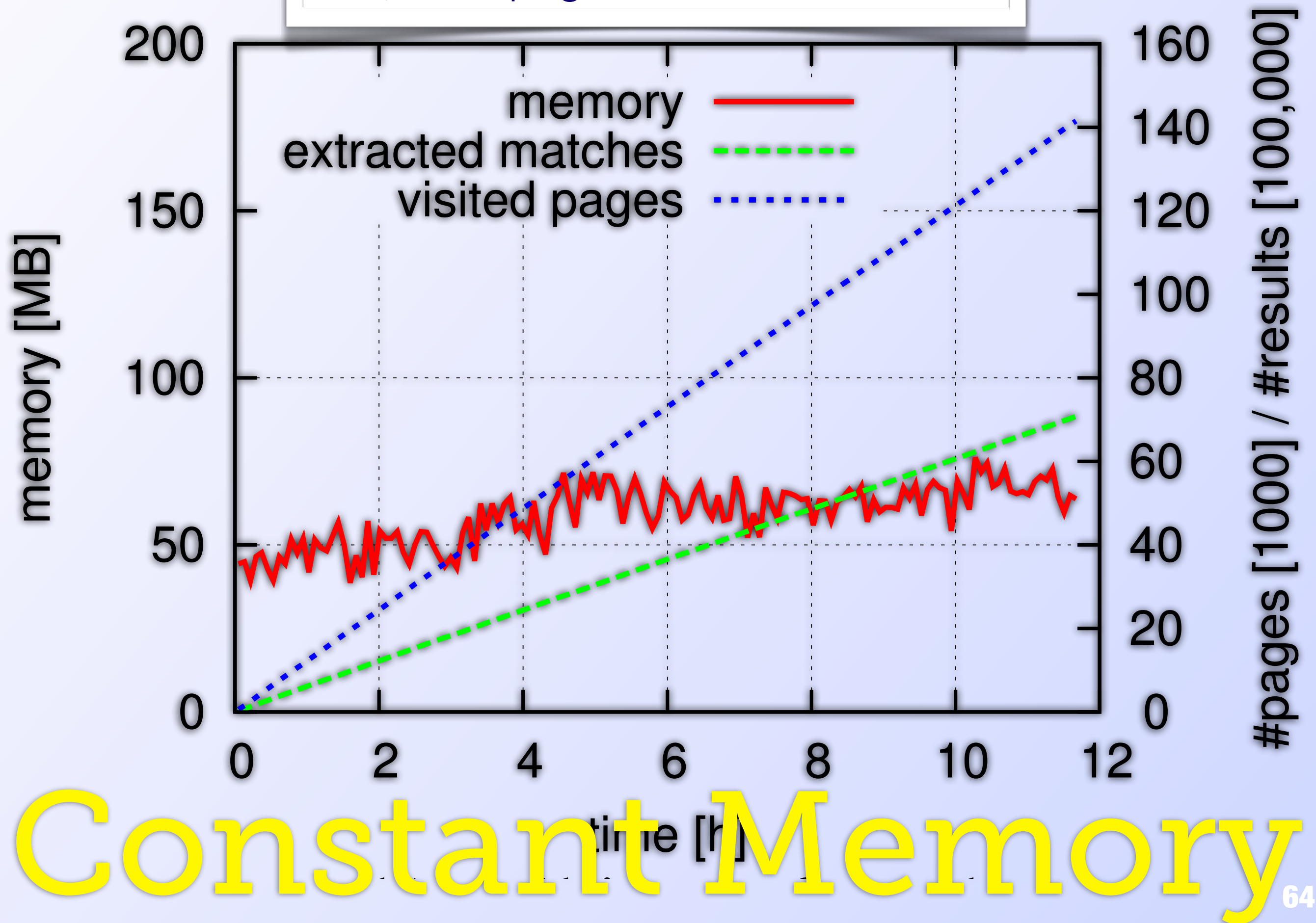
	Time	Space
OXPath <b>w/o Actions &amp; Kleene</b>	$O(n^6 \cdot q^2)$ $O(n^4 \cdot q^2)$	$O(n^5 \cdot q^2)$ $O(n^3 \cdot q^2)$
OXPath <b>w/o Kleene</b>	$O((p \cdot n)^6 \cdot q^3)$	$O(n^5 \cdot q^3)$
OXPath <b>w/o unbounded Kleene</b>	$O((p \cdot n)^6 \cdot q^3)$	$O(n^5 \cdot q_{\Sigma}^3)$
OXPath <b>(full)</b>	$O((p \cdot n)^6 \cdot q^3)$	$O(n^5 \cdot (q+d)^3)$



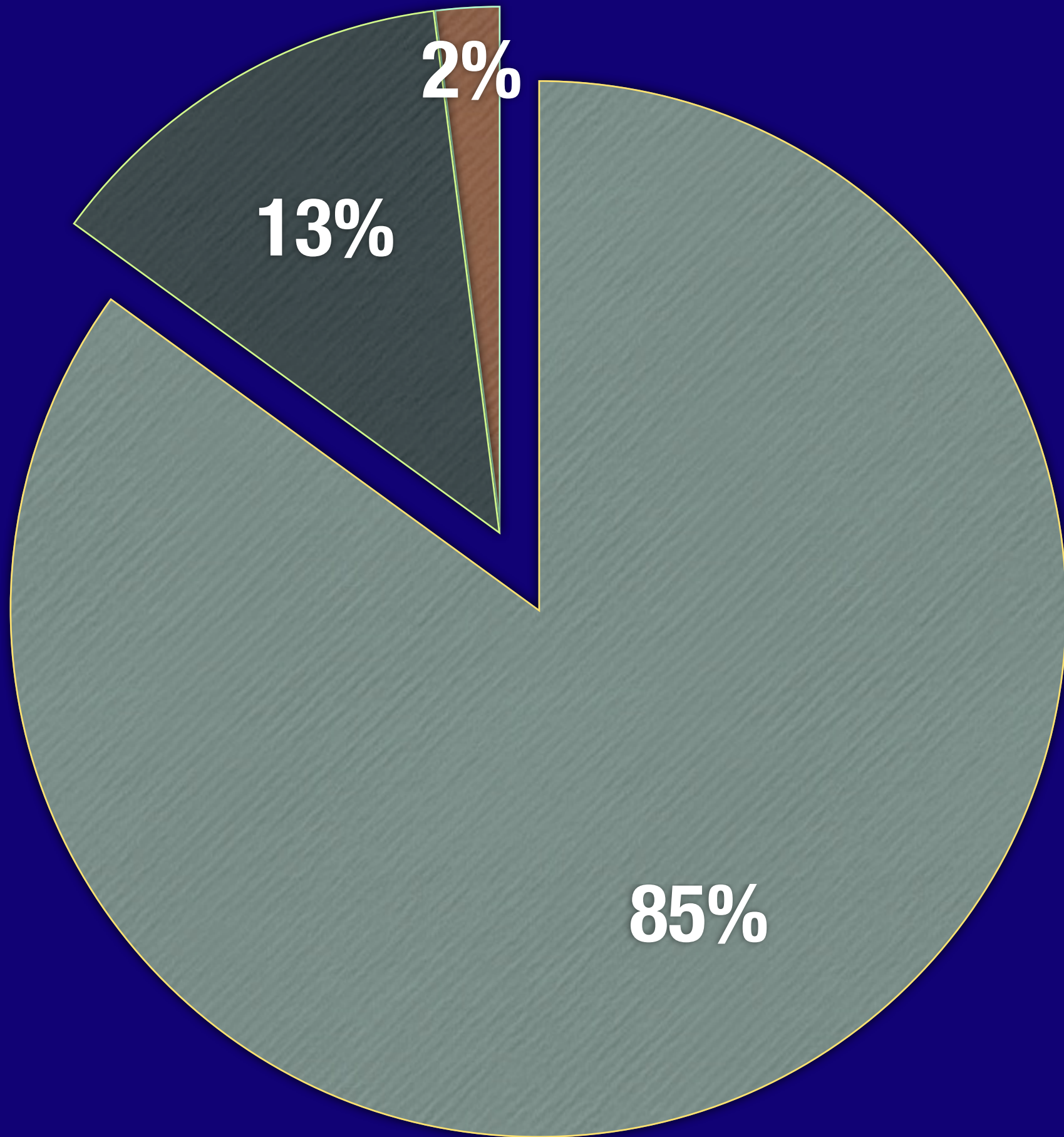




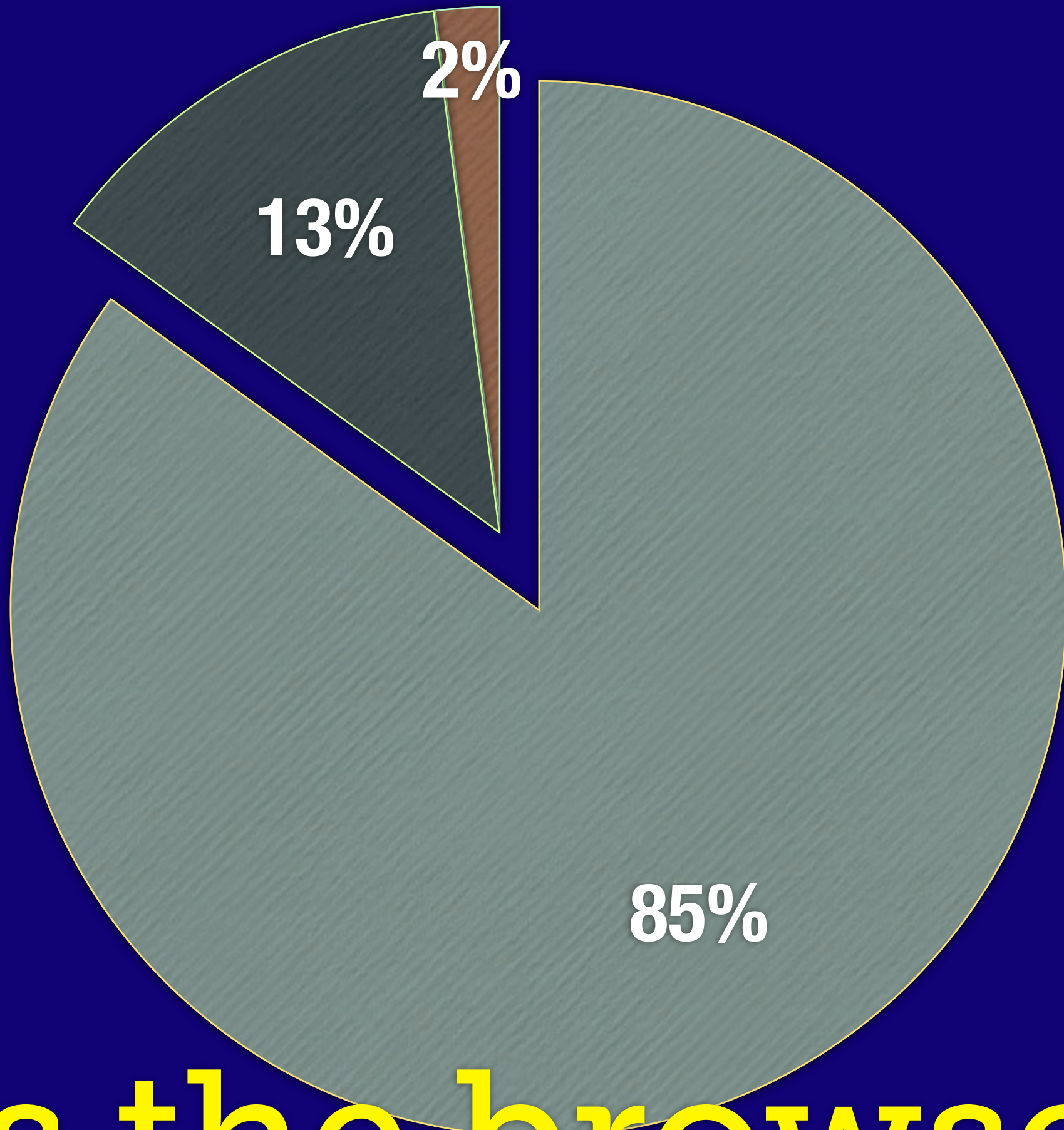
**100,000+** pages, **millions** of results





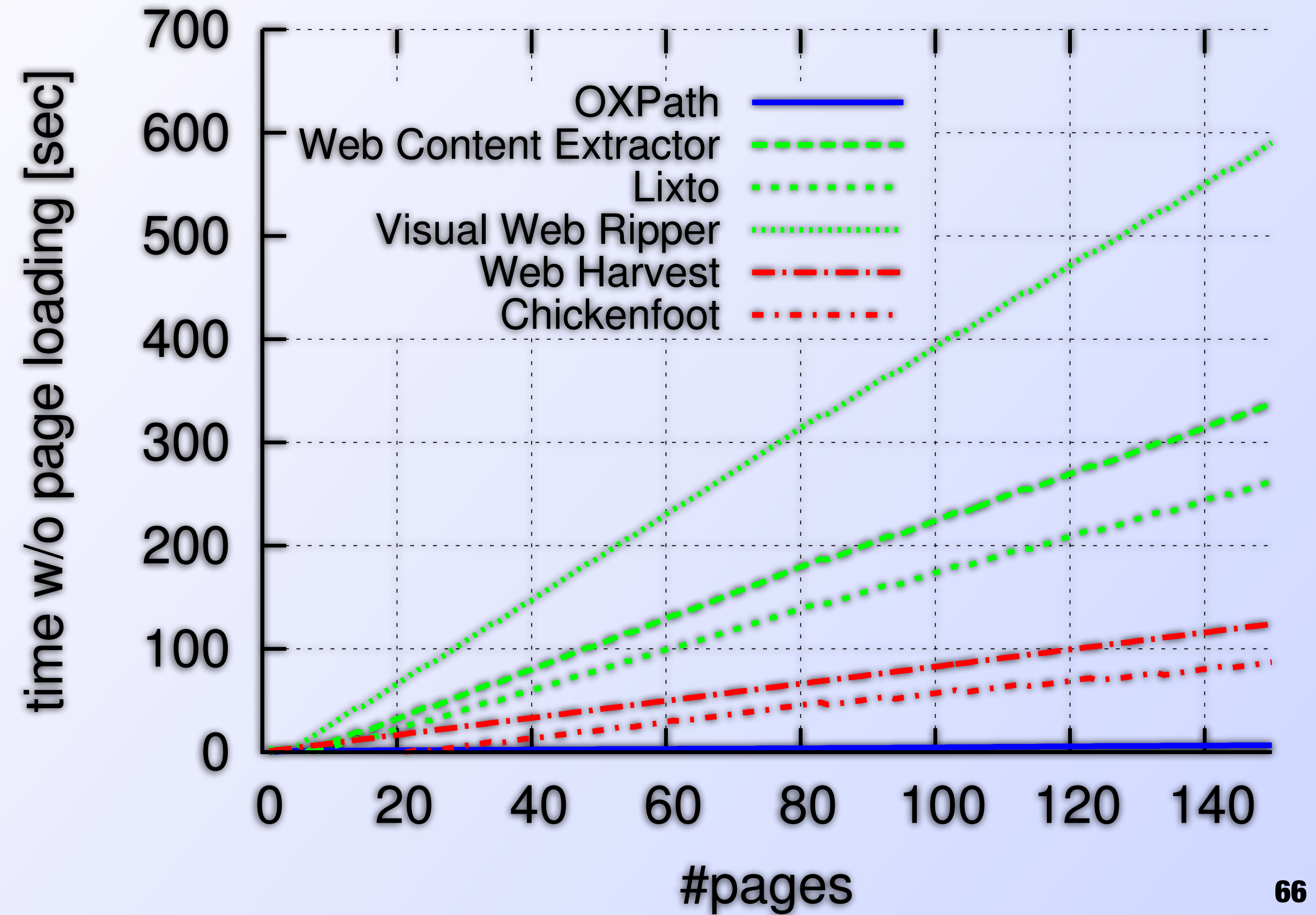




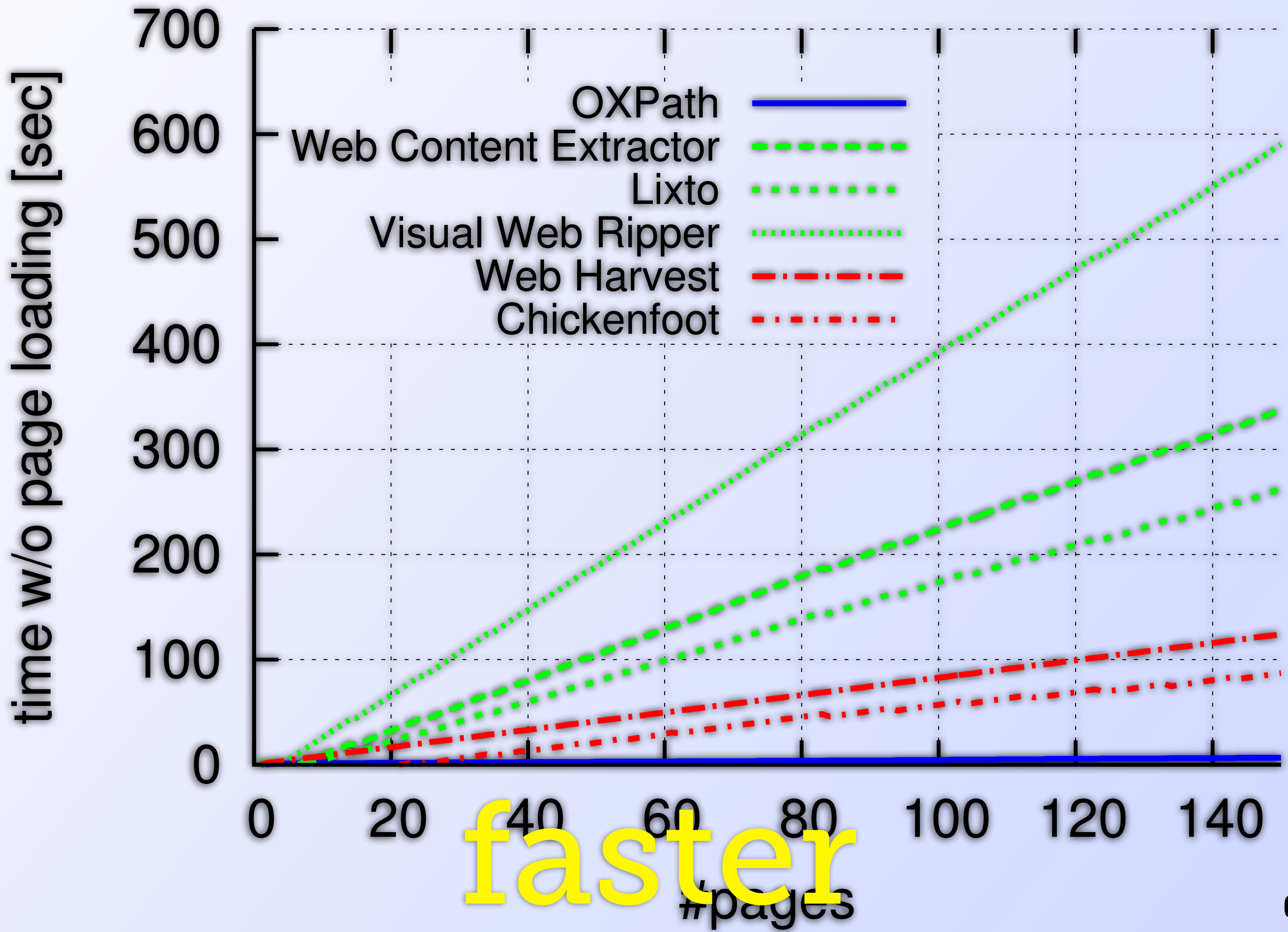


it's the browser

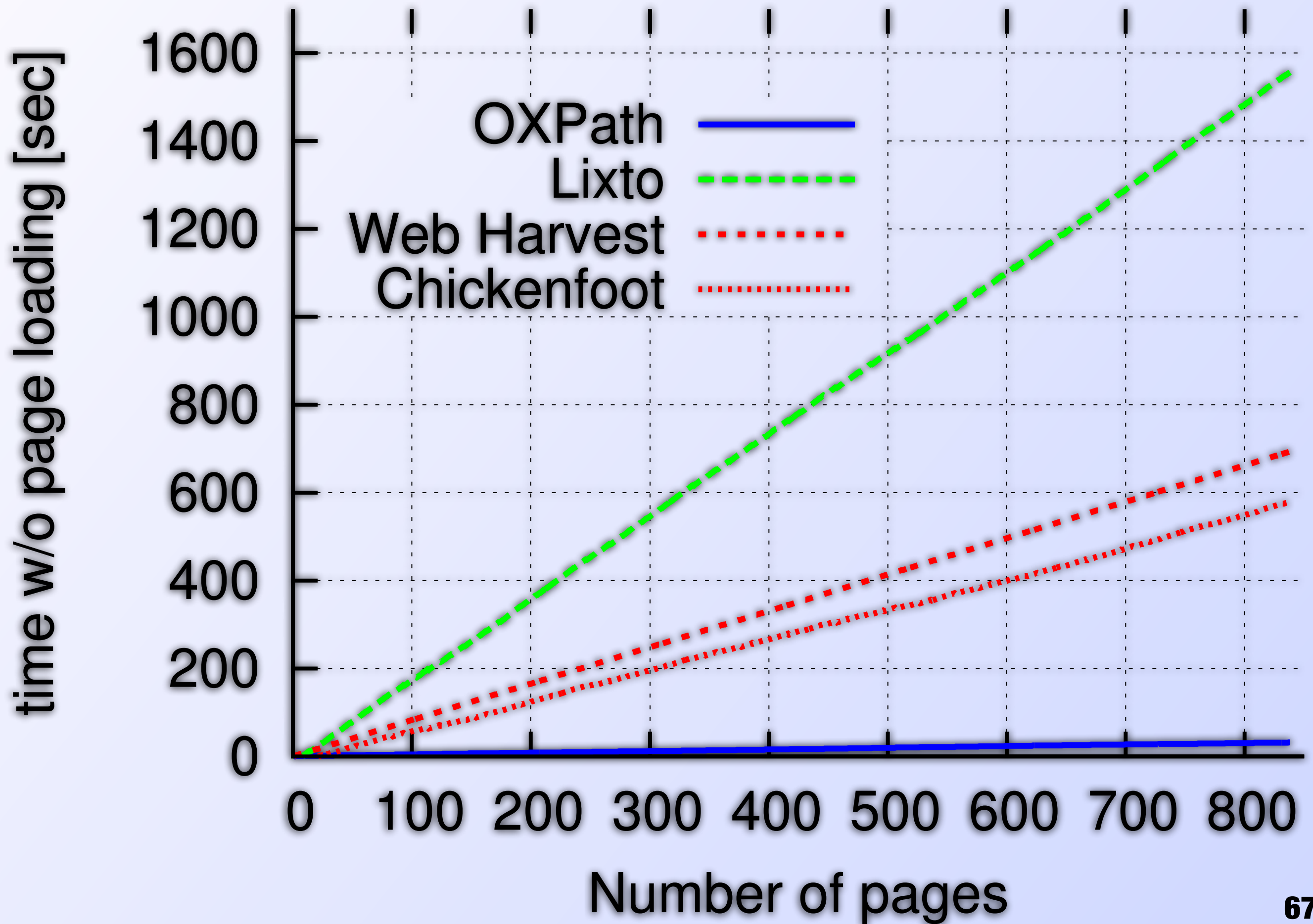




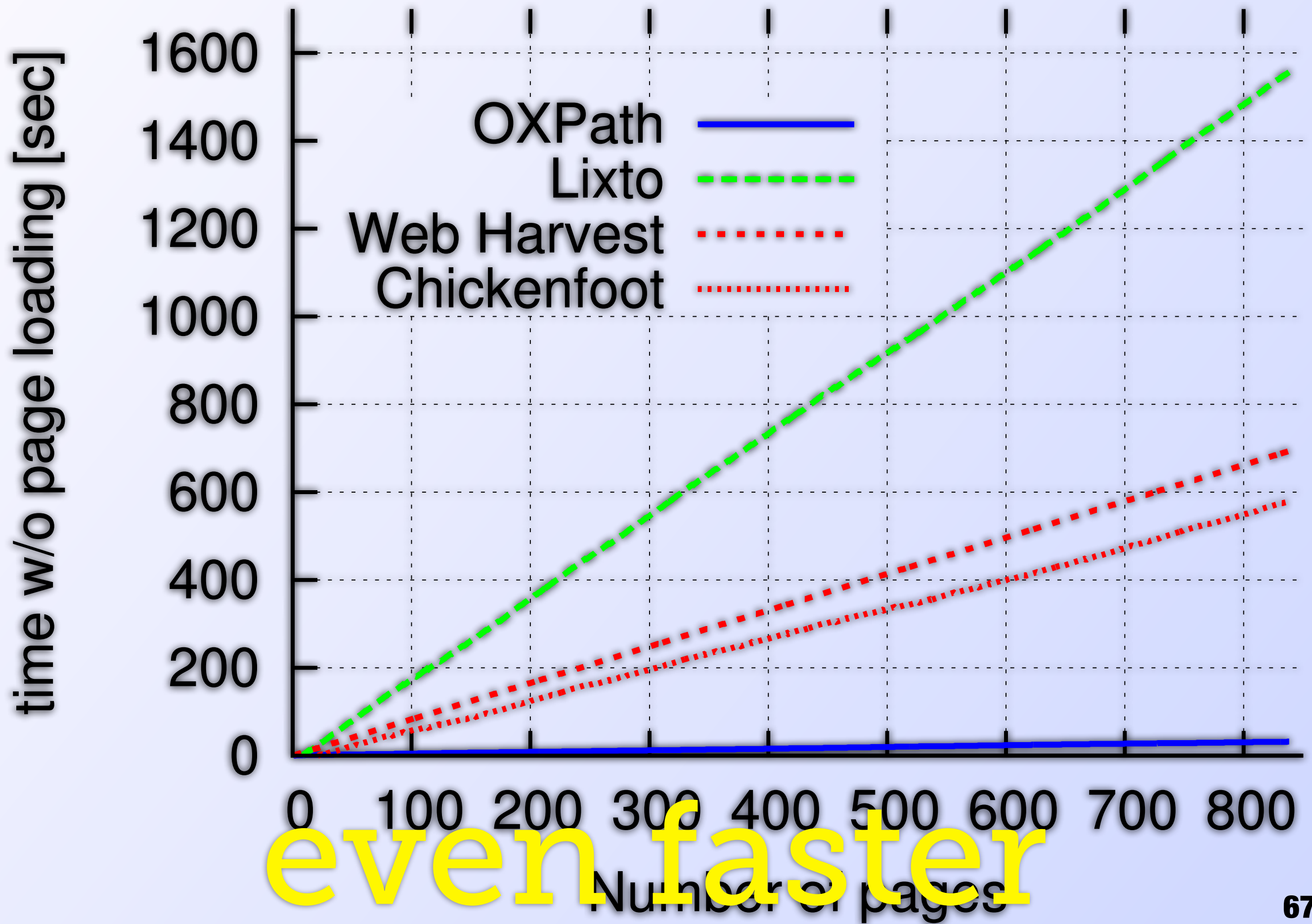














# Visual OXPath: Semi-supervised Generation

The screenshot displays the Visual OXPath tool interface, which is used for semi-supervised generation of XPath expressions. The interface is divided into several panels:

- Actions Panel (Left):** Shows a sequence of actions performed on the web page, such as "Go to http://scholar.google.com", "Enter 'OXPath' into //input[@name='q']", "Click on //input[@name='btnG']", and "Extract publication at //div[@class='gs\_r']". It also lists extracted data points like "title at .//div[@class='gs\_r']", "authors at .//span[@class='gs\_a']", and "citations at .//span[@class='gs\_c']".
- Browser (Recording) Panel (Center):** Displays the Google Scholar search results for the query "OXPath". The search results include a list of articles, with the first result highlighted: "[CITATION] Finding an OXPath to Cherries Hidden in the Scripted Web" by T Furche, G Gottlob, G Grasso, C Schallhart. The article abstract is visible, discussing the challenges of data extraction from web applications and the introduction of OXPath as a superset of XPath 1.0.
- DOM Panel (Right):** Shows the Document Object Model (DOM) tree of the current page. The selected element is the search results container, with the XPath expression `//div[@class='gs_r']` displayed.
- Results Panel (Bottom Right):** Displays the search results for the query "OXPath". The results are organized into a tree structure, showing the hierarchy of the search results. The first result is highlighted, showing the title, authors, and citations.

The interface also includes a search bar at the top, a "Search" button, and a "Did you mean: XPath" suggestion. The search results are displayed in a list format, with each result showing the title, authors, and a brief abstract. The first result is highlighted, and its details are shown in the Results panel.



# Visual OXPath: Semi-supervised Generation

Interaction recording

The screenshot displays the Visual OXPath interface, which is used for semi-supervised generation of XPath expressions. The interface consists of several panels:

- Actions Panel (Left):** A list of recorded interactions for the Google Scholar search. The first interaction is highlighted with a red circle and a red box. The actions are:
  - Go to `http://scholar.google.com`
  - Enter "OXPath" into `//input[@name='q']`
  - Click on `//input[@name='btnG']`
  - Extract publication at `//div[@class='gs_r']`
    - title at `./div[@class='gs_r']`
    - authors at `./span[@class='gs_a']`
    - citations at `./span[@class='gs_c']`
- Browser (Recording) Panel (Center):** A web browser window showing the Google Scholar search results for "OXPath". The search bar contains "OXPath", and the search button is labeled "Search". The results show a list of publications, with the first one highlighted in red. The first result is titled "Finding an OXPath to Cherries Hidden in the Scripted Web" by T Furche, G Gottlob, G Grasso, and C Schallhart. The second result is titled "Taking the oxpath down the deep web" by A Sellers, T Furche, G Gottlob, and G Grasso. The third result is titled "Oxpath: little language, little memory, great value" by AJ Sellers, T Furche, G Gottlob, and G Grasso. The fourth result is titled "Oxpath: A language for scalable, memory-efficient data extraction from web applications" by T Furche, G Gottlob, G Grasso, and C Schallhart.
- DOM Panel (Right):** A panel showing the DOM tree of the current page. The root node is `HTML`, with children `HEAD` and `BODY`. The `BODY` node is expanded, showing the `div[@class='gs_r']` node, which is the source of the first result.
- Results Panel (Bottom Right):** A panel showing the results of the search. It lists 10 results, with the first one highlighted in red. The first result is titled "Finding an OXPath to Cherries Hidden in the Scripted Web" by T Furche, G Gottlob, G Grasso, and C Schallhart. The second result is titled "Taking the oxpath down the deep web" by A Sellers, T Furche, G Gottlob, and G Grasso. The third result is titled "Oxpath: little language, little memory, great value" by AJ Sellers, T Furche, G Gottlob, and G Grasso. The fourth result is titled "Oxpath: A language for scalable, memory-efficient data extraction from web applications" by T Furche, G Gottlob, G Grasso, and C Schallhart.



# Visual OXPath: Semi-supervised Generation

Generation of OXPath expressions  
ranked by **robustness & specificity**

Action	Extract Root
Description	Extract publication at <code>//div[@class='gs_r']</code>
Marker	publication
OXPath	<code>//div[@class='gs_r']</code>
OXPath selects 10 nodes	

Score	Generator	OXPath	Count
403	position	<code>//div[@class='gs_r']</code>	10
358	position	<code>(//div[@class='gs_r'])</code>	10
250	position	<code>//div</code>	34
166	position	<code>(//div)</code>	34
100	canonical	<code>/html[1]/body[1]/div</code>	15

[CITATION] Finding an **OXPath** to Cherries Hidden in the Scripted Web  
T Furche, G Gottlob, G Grasso, C Schallhart... - ... rep., diadem-project. info/OXPath

Taking the **oxpath** down the deep web

A Sellers, T Furche, G Gottlob... - Proceedings of the ..., 2011 - portal.acm.org

ABSTRACT Although deep web analysis has been studied extensively, there is no succinct formalism to describe user interactions with AJAX-enabled web applications. Toward this end, we introduce **OXPath** as a superset of XPath 1.0. Beyond XPath, **OXPath** is able (1) to fill web forms ...

Cited by 1 - Related articles - Check Availability - All 11 versions

**Oxpath**: little language, little memory, great value

AJ Sellers, T Furche, G Gottlob... - Proceedings of the ..., 2011 - portal.acm.org

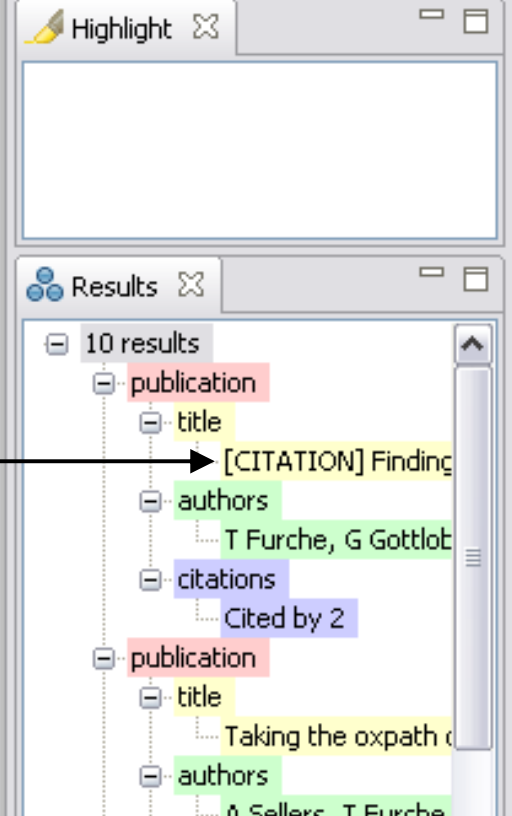
ABSTRACT Data about everything is readily available on the web—but often only accessible through elaborate user interactions. For automated decision support, extracting that data is essential, but infeasible with existing heavy-weight data extraction systems. In this ...

Cited by 1 - Related articles - Check Availability - All 11 versions

[PDF] **Oxpath**: A language for scalable, memory-efficient data extraction from web applications

T Furche, G Gottlob, G Grasso, C Schallhart... - Proceedings of the VLDB ..., 2011 - fox7.eu

ABSTRACT The evolution of the web has outpaced itself: The growing wealth of information



Actions

- Go to `http://scholar.google.com`
- Enter "OXPath" into `//input[@name='q']`
- Click on `//input[@name='btnG']`
- Extract publication at `//div[@class='gs_r']`
  - title at `./div[@class='gs_r']`
  - authors at `./span[@class='gs_a']`
  - citations at `./span[@class='gs_c']`

OXPath

```
doc("http://scholar.google.com")//input[@name='q']/{OXPath}/input[@name='btnG']/{click}/div.gs_r: <publication>[./div.gs_rt/h3: <title=string(.)>][./span.gs_a: <authors=string(.)>][./span.gs_fl/a[1]: <citations=string(.)>]
```



# SXPath

**Extending XPath towards  
Spatial Querying  
on Web Document**



# SXPath

## Extending XPath towards Spatial Querying on Web Documents

E. Oro, M. Ruffolo, S. Staab. **SXPath—Extending XPath towards Spatial Querying on Web Documents.** *PVLDB* 4 (2): 129-140 (2010)



# Motivation: Spatial Arrangement

- **Goal:** Robust wrappers usable on a wide range of web sites
- **Observation:**
  - Web pages are made for humans and provide **visual clues**
    - through the **spatial arrangement** of text and other media
  - XPath (and others) are established selection mechanisms
    - but **lack any support** for spatial arrangement on Web pages
- **Solution:**
  - **Extend** XPath with facilities for querying spatial arrangement



# Presentation-Oriented Documents

- Web page from lastfm (<http://www.lastfm.it/>)
- Music band profile:** a photo with descriptive information **at its east**

The screenshot shows the Last.fm website interface. At the top, there's a red navigation bar with the Last.fm logo and links for Musica, Radio, Eventi, Classifiche, and Community. Below this is a blue banner with the text "SOLO fino al 30 settembre raddoppi i vantaggi!". The main content area is divided into two columns. The left column has a search bar and a list of music genres. The right column features a "Tra i tuoi consigli" section with a registration prompt and a Groupon advertisement for airplane tickets. The central part of the page displays "Musica più ascoltata su Last.fm" with a list of top artists, including Radiohead and Adele, each with a photo, name, listener count, and similar artists.

last.fm Musica Radio Eventi Classifiche Community Registrati Accedi

Come work with us! Last.fm is hiring » Italiano | Guida Ricerca di musica

SOLO fino al 30 settembre raddoppi i vantaggi!

Trova musica su Last.fm

Cerca in catalogo:  Inserisci artista o tag Cerca

Tutti

- acoustic
- ambient
- blues
- classical
- country
- electronic
- emo
- folk
- hardcore
- hip hop
- indie
- jazz

Musica più ascoltata su Last.fm

Più ascoltata | Del momento | Più ascoltata in: Italia

IN TOUR

Radiohead

303.041.555 ascolti (3.483.089 ascoltatori)

Simile a: Thom Yorke, Jonny Greenwood, Sigur Rós, Portishead, Blur

alternative

IN TOUR

Adele

43.453.878 ascolti (1.275.590 ascoltatori)

Simile a: Jessie J, Amy Winehouse, Christina Perri, Beyoncé, Glee Cast

soul

★ Tra i tuoi consigli

Ottieni consigli musicali basati sui tuoi gusti.

Registrati, raccontaci che musica ti piace e potrai subito scoprire tanta nuova musica.

Registrati ora

GROUPON

Biglietti Aerei\*

FINO A -70%

Qui >



# Presentation-Oriented Documents

- Web page from lastfm (<http://www.lastfm.it/>)
- Music band profile:** a photo with descriptive information **at its east**

band profile

The screenshot shows the Last.fm website interface. At the top, there's a navigation bar with links for Musica, Radio, Eventi, Classifiche, and Community. Below this is a search bar and a language selector set to Italiano. A large blue banner advertises a promotion: "SOLO fino al 30 settembre raddoppi i vantaggi!". The main content area is titled "Trova musica su Last.fm" and includes a search bar. On the left, there's a sidebar with a list of music genres: Tutti, acoustic, ambient, blues, classical, country, electronic, emo, folk, hardcore, hip hop, indie, and jazz. The main content area features a section titled "Musica più ascoltata su Last.fm" with sub-sections for "Più ascoltata", "Del momento", and "Più ascoltata in: Italia". Two band profiles are highlighted with red boxes. The first profile is for Radiohead, showing a photo of the band, their name, listener count (303.041.555 ascolti), and a list of similar artists (Thom Yorke, Jonny Greenwood, Sigur Rós, Portishead, Blur). The second profile is for Adele, showing a photo of her, her name, listener count (43.453.878 ascolti), and a list of similar artists (Jessie J, Amy Winehouse, Christina Perri, Beyoncé, Glee Cast). A red arrow points from the text "band profile" to the Radiohead profile. On the right side of the page, there's a section titled "Tra i tuoi consigli" with a registration prompt and a Groupon advertisement for "Biglietti Aerei\*" with a 70% discount.

last.fm Musica Radio Eventi Classifiche Community Registrati Accedi

Come work with us! Last.fm is hiring » Italiano | Guida Ricerca di musica

SOLO fino al 30 settembre raddoppi i vantaggi!

Trova musica su Last.fm

Cerca in catalogo: Inserisci artista o tag Cerca

Tutti  
acoustic  
ambient  
blues  
classical  
country  
electronic  
emo  
folk  
hardcore  
hip hop  
indie  
jazz

Musica più ascoltata su Last.fm

Più ascoltata | Del momento | Più ascoltata in: Italia

IN TOUR

Radiohead  
303.041.555 ascolti (3.483.089 ascoltatori)  
Simile a: Thom Yorke, Jonny Greenwood, Sigur Rós, Portishead, Blur  
alternative

IN TOUR

Adele  
43.453.878 ascolti (1.275.590 ascoltatori)  
Simile a: Jessie J, Amy Winehouse, Christina Perri, Beyoncé, Glee Cast  
soul

★ Tra i tuoi consigli

Ottieni consigli musicali basati sui tuoi gusti.  
Registrati, raccontaci che musica ti piace e potrai subito scoprire tanta nuova musica.  
Registrati ora

GROUPON

Biglietti Aerei\*

FINO A -70%

Qui >



# Presentation-Oriented Documents

- Web page from lastfm (<http://www.lastfm.it/>)
- Music band profile:** a photo with descriptive information **at its east**

band profile

The screenshot shows the Last.fm website interface. At the top, there's a red navigation bar with the 'last.fm' logo and links for 'Musica', 'Radio', 'Eventi', 'Classifiche', and 'Community'. Below this is a search bar with the text 'Cerca in catalogo: Inserisci artista o tag' and a 'Cerca' button. A large blue banner across the middle reads 'SOLO fino al 30 settembre raddoppi i vantaggi!'. On the left, there's a sidebar with a list of music genres: 'Tutti', 'acoustic', 'ambient', 'blues', 'classical', 'country', 'electronic', 'emo', 'folk', 'hardcore', 'hip hop', 'indie', and 'jazz'. The main content area is titled 'Trova musica su Last.fm' and features a section 'Musica più ascoltata su Last.fm'. This section lists two bands: 'Radiohead' and 'Adele'. Each band entry includes a photo, the band name, the number of listens (e.g., '303.041.555 ascolti' for Radiohead), a list of similar artists, and a genre tag (e.g., 'alternative' for Radiohead, 'soul' for Adele). On the right side, there's a section 'Tra i tuoi consigli' with a message about getting musical advice based on user preferences and a 'Registrati ora' button. At the bottom right, there's a 'GROUPON' advertisement for 'Biglietti Aerei\*' with a '-70%' discount and a 'Qui >' button.



# Presentation-Oriented Documents

- Web page from lastfm (<http://www.lastfm.it/>)
- Music band profile:** a photo with descriptive information **at its east**

band profile

photo

The screenshot shows the Last.fm website interface. At the top, there's a red navigation bar with the Last.fm logo and links for Musica, Radio, Eventi, Classifiche, and Community. Below this is a blue banner with the text "SOLO fino al 30 settembre raddoppi i vantaggi!". The main content area is titled "Trova musica su Last.fm" and features a search bar. On the left, there's a sidebar with a list of music genres: Tutti, acoustic, ambient, blues, classical, country, electronic, emo, folk, hardcore, hip hop, indie, and jazz. The main content area displays "Musica più ascoltata su Last.fm" with a sub-header "Più ascoltata | Del momento | Più ascoltata in: Italia". The first band profile shown is for Radiohead, featuring a photo of the band members (highlighted by a red box and an arrow from the "photo" label) and their statistics: 303.041.555 ascolti (3.483.089 ascoltatori). Below the photo, it lists similar artists: Thom Yorke, Jonny Greenwood, Sigur Rós, Portishead, and Blur, along with a "alternative" tag. The second band profile shown is for Adele, featuring a photo of her (highlighted by a red box and an arrow from the "photo" label) and her statistics: 43.453.878 ascolti (1.275.590 ascoltatori). Below her photo, it lists similar artists: Jessie J, Amy Winehouse, Christina Perri, Beyoncé, and Glee Cast, along with a "soul" tag. On the right side of the page, there's a section titled "★ Tra i tuoi consigli" with the text "Ottieni consigli musicali basati sui tuoi gusti. Registrati, raccontaci che musica ti piace e potrai subito scoprire tanta nuova musica." and a "Registrati ora" button. At the bottom right, there's a Groupon advertisement for "Biglietti Aerei\*" with a "FINO A -70%" discount and a "Qui >" button.



# Presentation-Oriented Documents

- Web page from lastfm (<http://www.lastfm.it/>)
- Music band profile:** a photo with descriptive information **at its east**

band profile

photo

The screenshot shows the Last.fm website interface. At the top, there's a red navigation bar with the Last.fm logo and links for Musica, Radio, Eventi, Classifiche, and Community. Below this is a blue banner with the text "SOLO fino al 30 settembre raddoppi i vantaggi!". The main content area is divided into two columns. The left column features a search bar and a list of music genres (acoustic, ambient, blues, classical, country, electronic, emo, folk, hardcore, hip hop, indie, jazz). The right column displays "Musica più ascoltata su Last.fm" with a list of top artists. The first artist listed is Radiohead, with a photo of the band members and their statistics (303.041.555 ascolti, 3.483.089 ascoltatori). Below Radiohead is Adele, also with a photo and statistics (43.453.878 ascolti, 1.275.590 ascoltatori). On the far right, there's a sidebar with a "Tra i tuoi consigli" section and a "GROUPON Biglietti Aerei\*" advertisement featuring a red "-70%" discount tag and a "Qui >" button.

last.fm Musica Radio Eventi Classifiche Community Registrati Accedi

Come work with us! Last.fm is hiring » Italiano | Guida Ricerca di musica

SOLO fino al 30 settembre raddoppi i vantaggi!

Trova musica su Last.fm

Cerca in catalogo:  Cerca

Tutti

- acoustic
- ambient
- blues
- classical
- country
- electronic
- emo
- folk
- hardcore
- hip hop
- indie
- jazz

Musica più ascoltata su Last.fm

Più ascoltata | Del momento | Più ascoltata in: Italia

IN TOUR

 **Radiohead**  
303.041.555 ascolti (3.483.089 ascoltatori)  
Simile a: Thom Yorke, Jonny Greenwood, Sigur Rós, Portishead, Blur  
alternative

IN TOUR

 **Adele**  
43.453.878 ascolti (1.275.590 ascoltatori)  
Simile a: Jessie J, Amy Winehouse, Christina Perri, Beyoncé, Glee Cast  
soul

★ Tra i tuoi consigli

Ottieni consigli musicali basati sui tuoi gusti.

Registrati, raccontaci che musica ti piace e potrai subito scoprire tanta nuova musica.

Registrati ora

GROUPON

Biglietti Aerei\*

FINO A -70%

Qui >



# Presentation-Oriented Documents

- Web page from lastfm (<http://www.lastfm.it/>)
- Music band profile:** a photo with descriptive information **at its east**

band profile

photo

name

The screenshot shows the Last.fm website interface. At the top, there's a red navigation bar with the 'last.fm' logo and links for 'Musica', 'Radio', 'Eventi', 'Classifiche', and 'Community'. Below this is a blue banner with the text 'SOLO fino al 30 settembre raddoppi i vantaggi!'. The main content area is titled 'Trova musica su Last.fm' and includes a search bar. On the left, there's a sidebar with a list of music genres: 'Tutti', 'acoustic', 'ambient', 'blues', 'classical', 'country', 'electronic', 'emo', 'folk', 'hardcore', 'hip hop', 'indie', and 'jazz'. The main content area features a section titled 'Musica più ascoltata su Last.fm' with sub-sections for 'Più ascoltata', 'Del momento', and 'Più ascoltata in: Italia'. Two band profiles are displayed: 'Radiohead' and 'Adele'. The 'Radiohead' profile includes a photo of the band, their name in a red box, and a list of similar artists: 'Thom Yorke, Jonny Greenwood, Sigur Rós, Portishead, Blur'. The 'Adele' profile includes a photo of the artist, her name, and a list of similar artists: 'Jessie J, Amy Winehouse, Christina Perri, Beyoncé, Glee Cast'. On the right side of the page, there's a section titled 'Tra i tuoi consigli' with a button 'Registrati ora'. At the bottom right, there's a 'GROUPON' advertisement for 'Biglietti Aerei\*' with a '-70%' discount and a 'Qui >' button.



# Presentation-Oriented Documents

- Web page from lastfm (<http://www.lastfm.it/>)
- Music band profile:** a photo with descriptive information **at its east**

band profile

photo

name

The screenshot shows the Last.fm website interface. At the top, there's a red navigation bar with the Last.fm logo and links for Musica, Radio, Eventi, Classifiche, and Community. Below this is a blue banner with the text "SOLO fino al 30 settembre raddoppi i vantaggi!". The main content area is divided into two columns. The left column features a search bar and a list of music genres (acoustic, ambient, blues, classical, country, electronic, emo, folk, hardcore, hip hop, indie, jazz). The right column displays a "Musica più ascoltata su Last.fm" section, which includes a "Tra i tuoi consigli" box and a "Biglietti Aerei\*" advertisement. The "Musica più ascoltata" section lists two bands: Radiohead and Adele, each with a photo, a list of similar artists, and a genre tag (alternative for Radiohead, soul for Adele).

last.fm Musica Radio Eventi Classifiche Community Registrati Accedi

Come work with us! Last.fm is hiring » Italiano | Guida Ricerca di musica

SOLO fino al 30 settembre raddoppi i vantaggi!

Trova musica su Last.fm

Cerca in catalogo:  Cerca

Tutti

- acoustic
- ambient
- blues
- classical
- country
- electronic
- emo
- folk
- hardcore
- hip hop
- indie
- jazz

Musica più ascoltata su Last.fm

Più ascoltata | Del momento | Più ascoltata in: Italia

IN TOUR

Radiohead

303.041.555 ascolti (3.483.089 ascoltatori)

Simile a: Thom Yorke, Jonny Greenwood, Sigur Rós, Portishead, Blur

alternative

IN TOUR

Adele

43.453.878 ascolti (1.275.590 ascoltatori)

Simile a: Jessie J, Amy Winehouse, Christina Perri, Beyoncé, Glee Cast

soul

★ Tra i tuoi consigli

Ottieni consigli musicali basati sui tuoi gusti.

Registrati, raccontaci che musica ti piace e potrai subito scoprire tanta nuova musica.

Registrati ora

GROUPON

Biglietti Aerei\*

FINO A -70%

Qui >



# Presentation-Oriented Documents

Web page

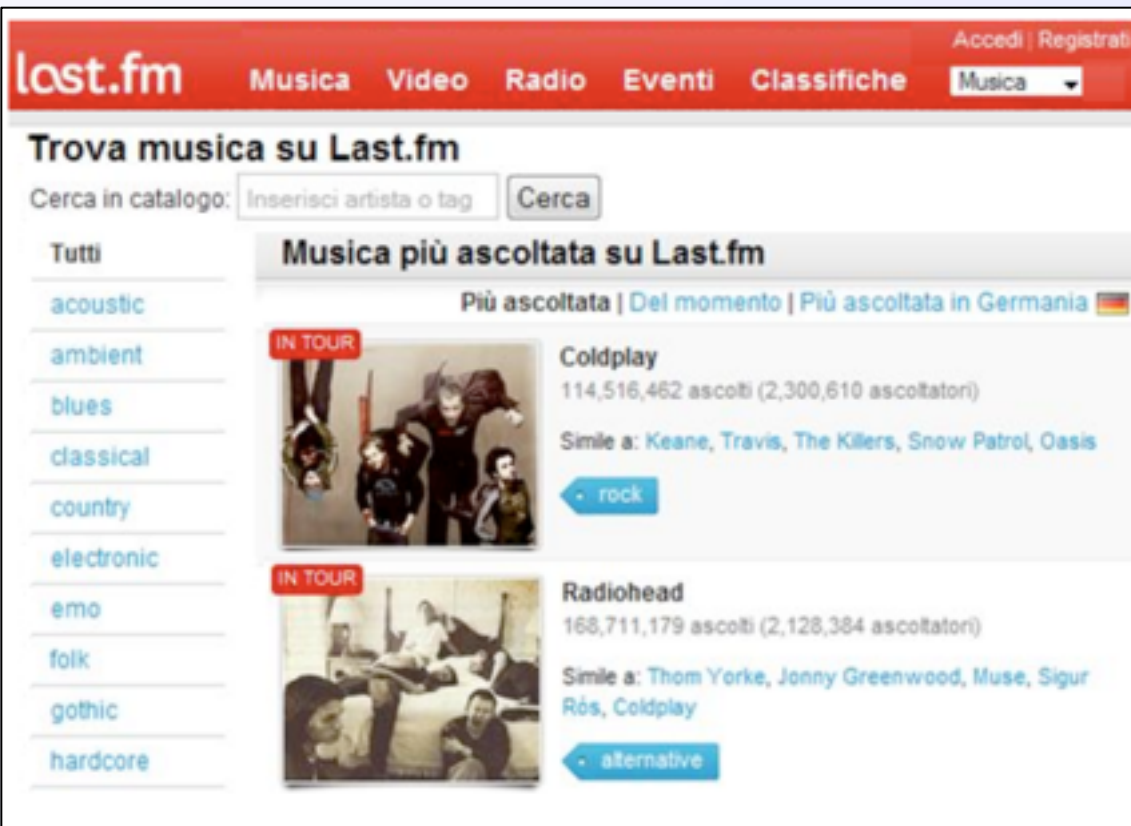
Document object model (DOM)

- Spatial arrangement rarely explicit
  - but hidden in complex nesting of layout elements in DOM
  - **difficult:** intricate tree structure conceptually difficult to query
  - **not robust:** small changes in layout structure break a wrapper



# Presentation-Oriented Documents

## Web page



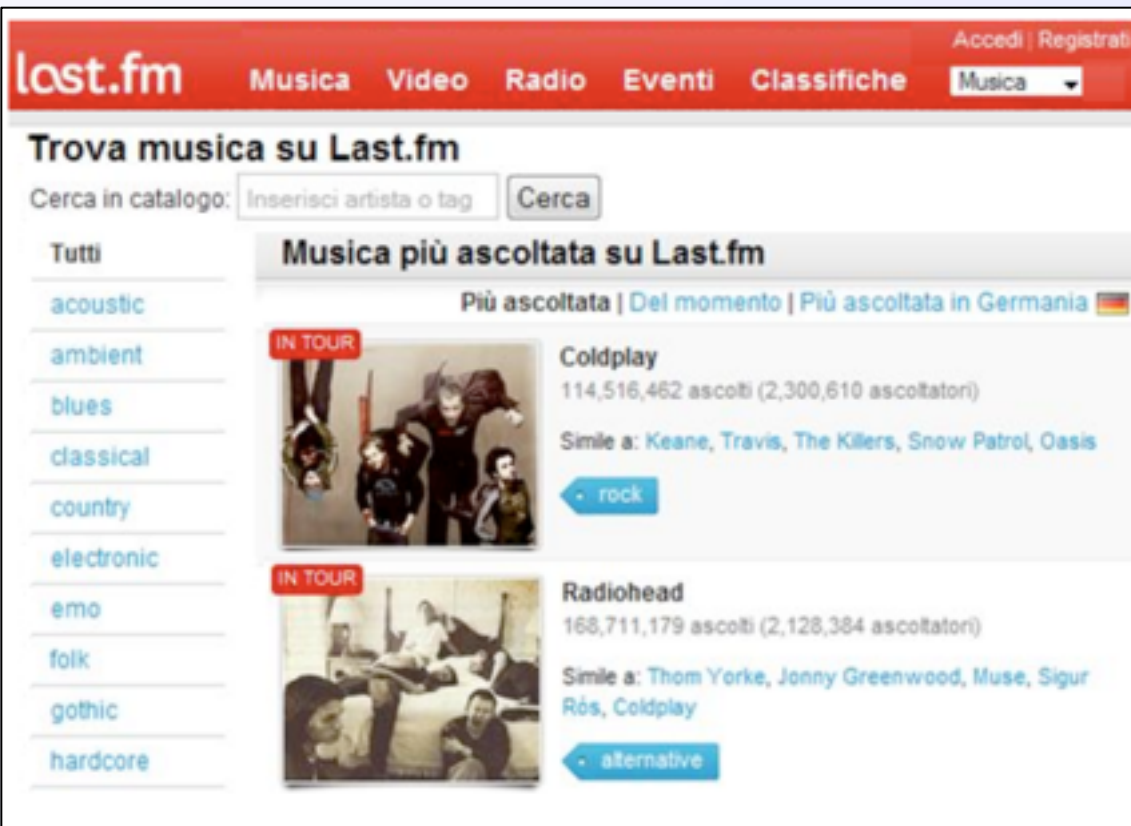
## Document object model (DOM)

- Spatial arrangement rarely explicit
  - but hidden in complex nesting of layout elements in DOM
  - **difficult:** intricate tree structure conceptually difficult to query
  - **not robust:** small changes in layout structure brake a wrapper

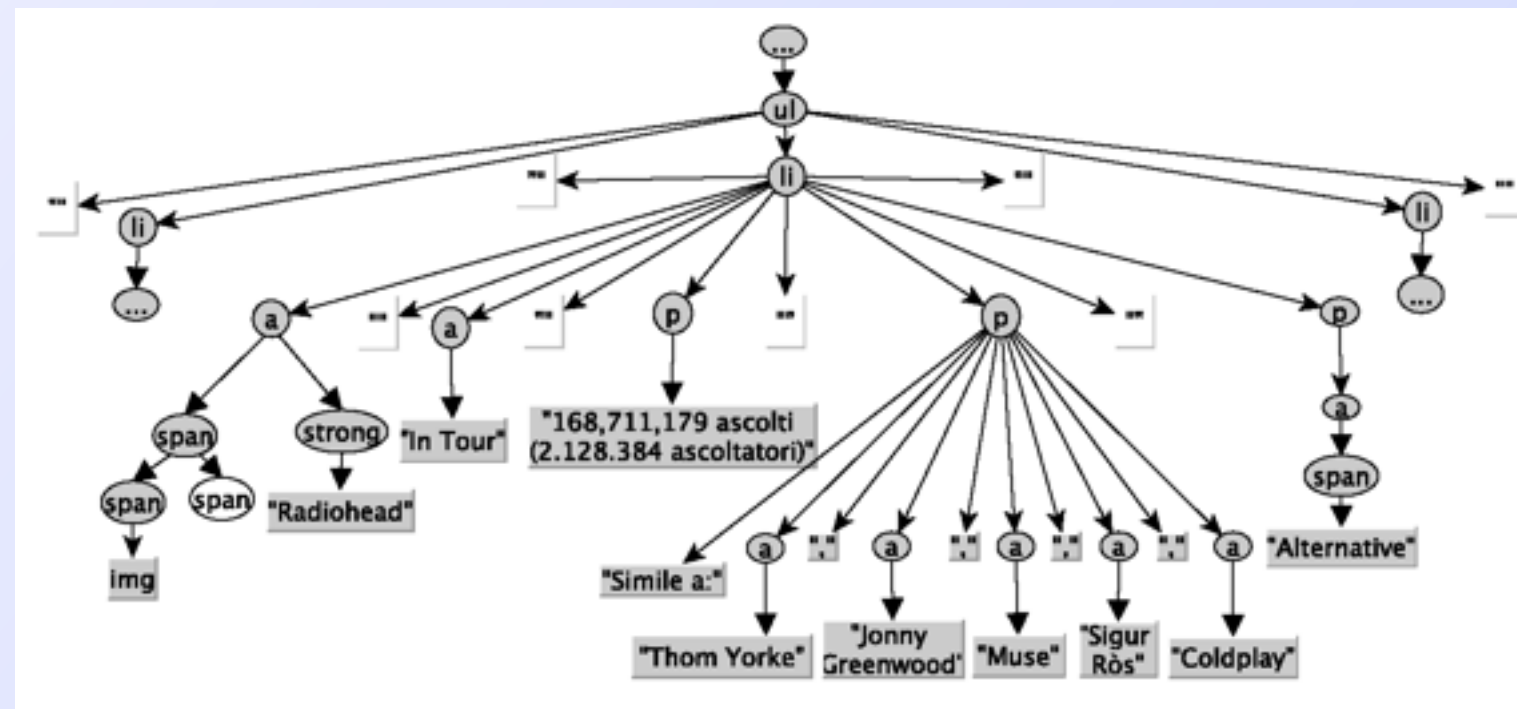


# Presentation-Oriented Documents

## Web page



## Document object model (DOM)

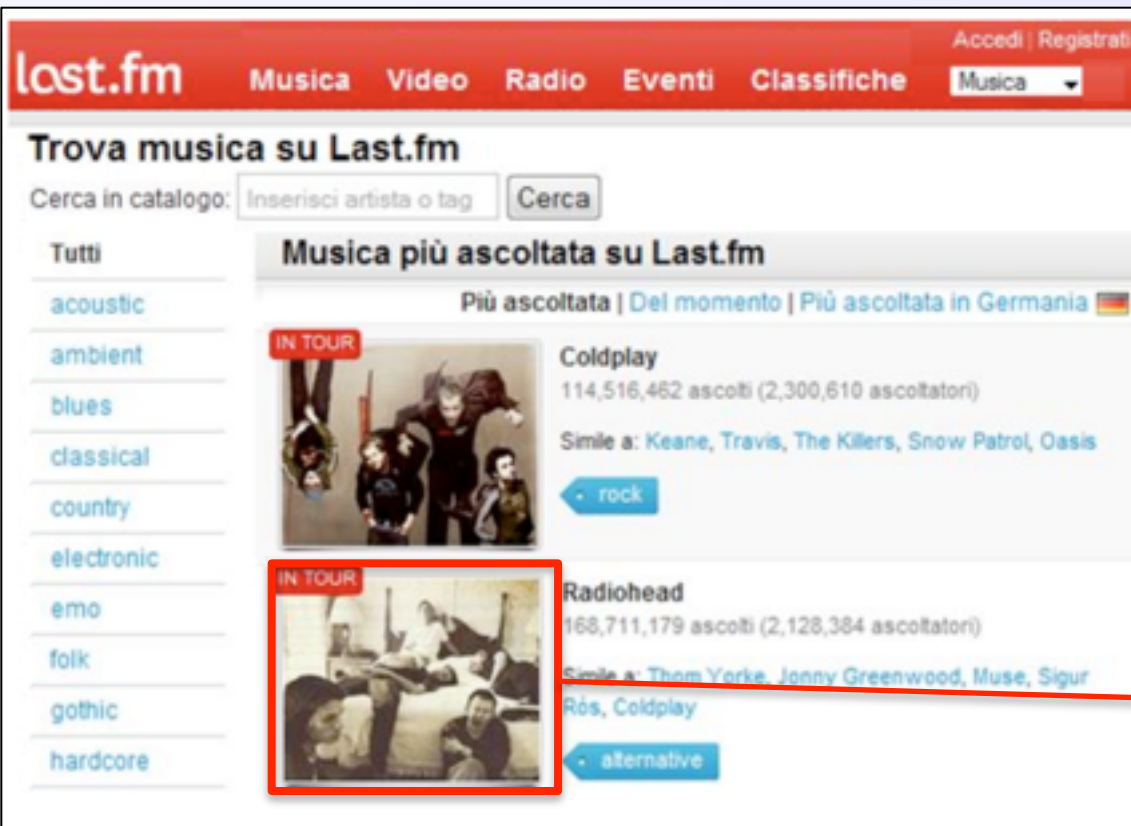


- Spatial arrangement rarely explicit
  - but hidden in complex nesting of layout elements in DOM
  - **difficult:** intricate tree structure conceptually difficult to query
  - **not robust:** small changes in layout structure brake a wrapper

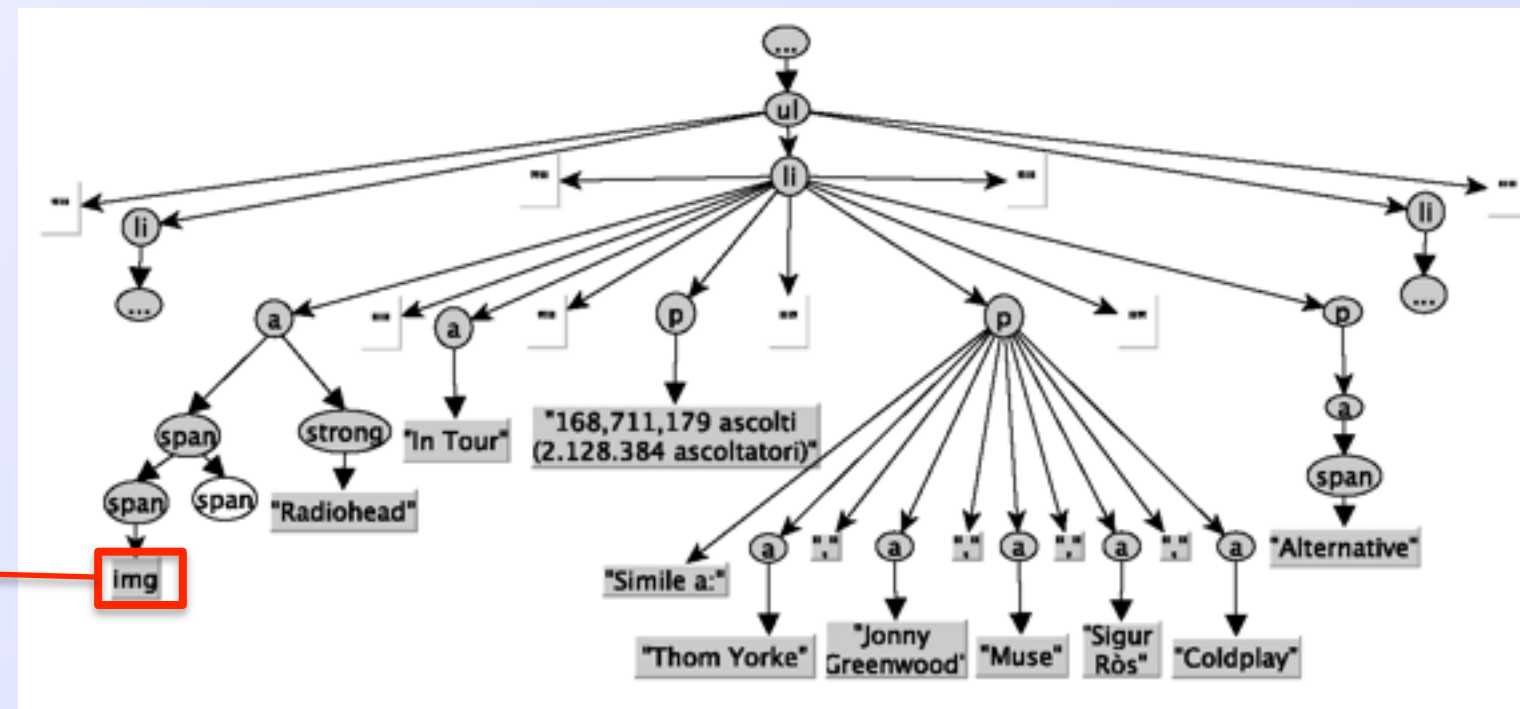


# Presentation-Oriented Documents

## Web page



## Document object model (DOM)

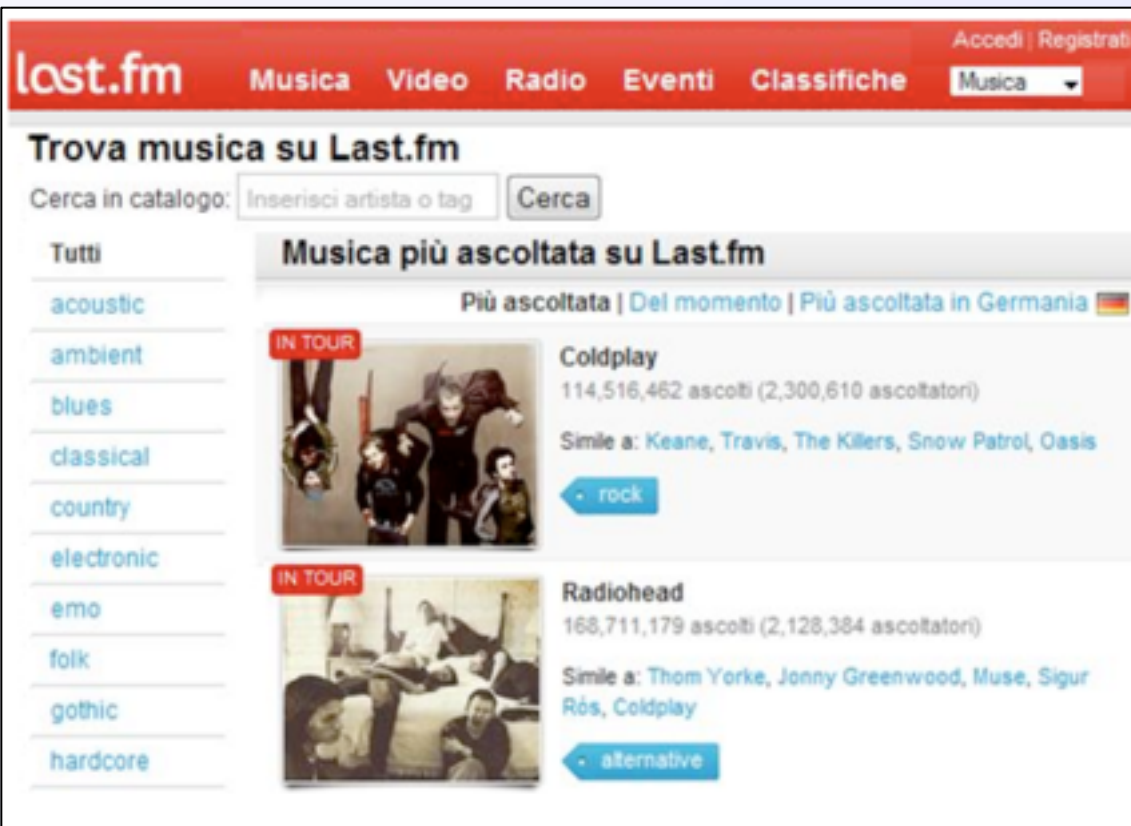


- Spatial arrangement rarely explicit
  - but hidden in complex nesting of layout elements in DOM
  - **difficult:** intricate tree structure conceptually difficult to query
  - **not robust:** small changes in layout structure brake a wrapper

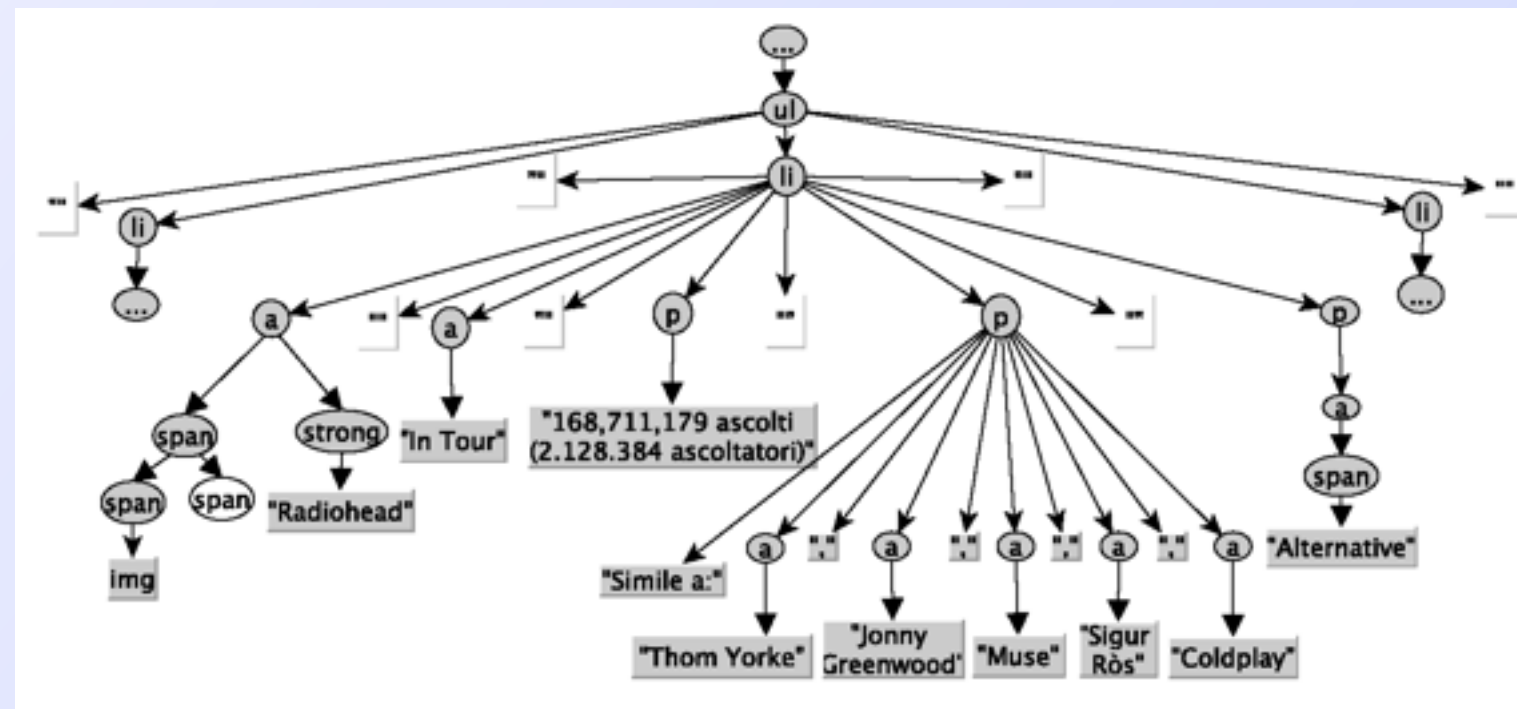


# Presentation-Oriented Documents

## Web page



## Document object model (DOM)

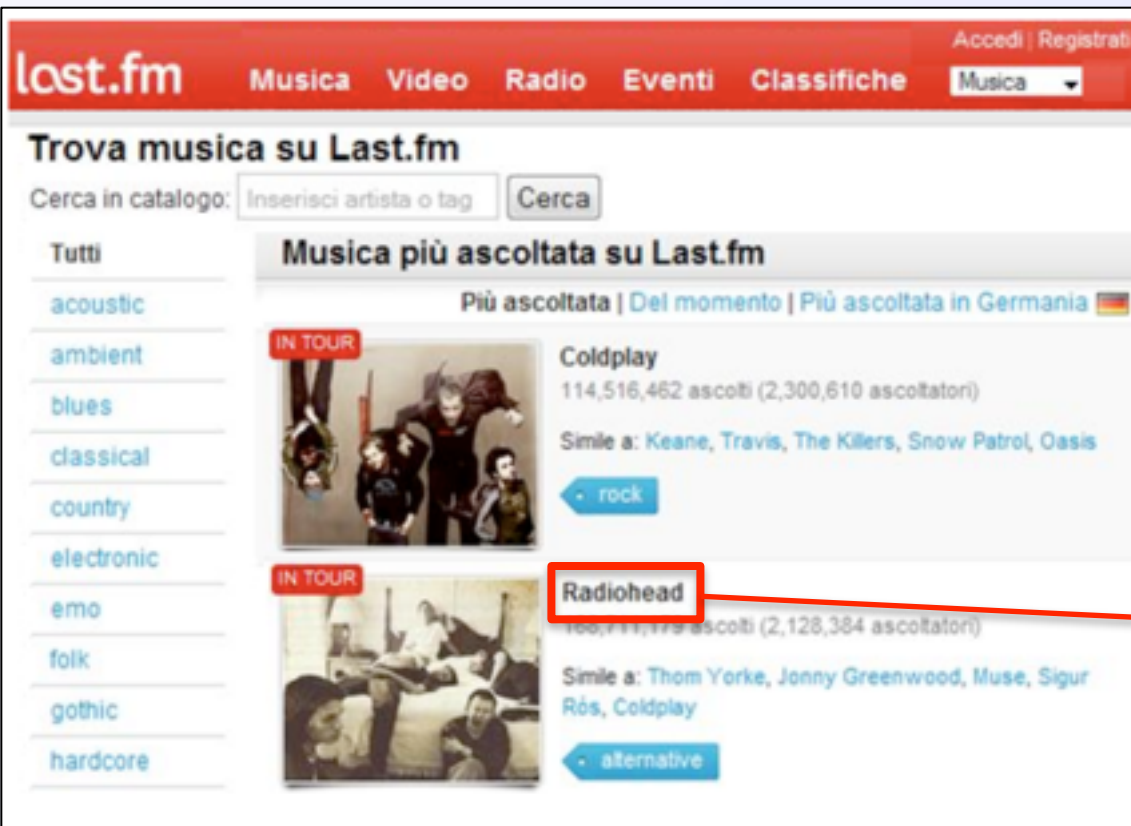


- Spatial arrangement rarely explicit
  - but hidden in complex nesting of layout elements in DOM
  - **difficult:** intricate tree structure conceptually difficult to query
  - **not robust:** small changes in layout structure brake a wrapper

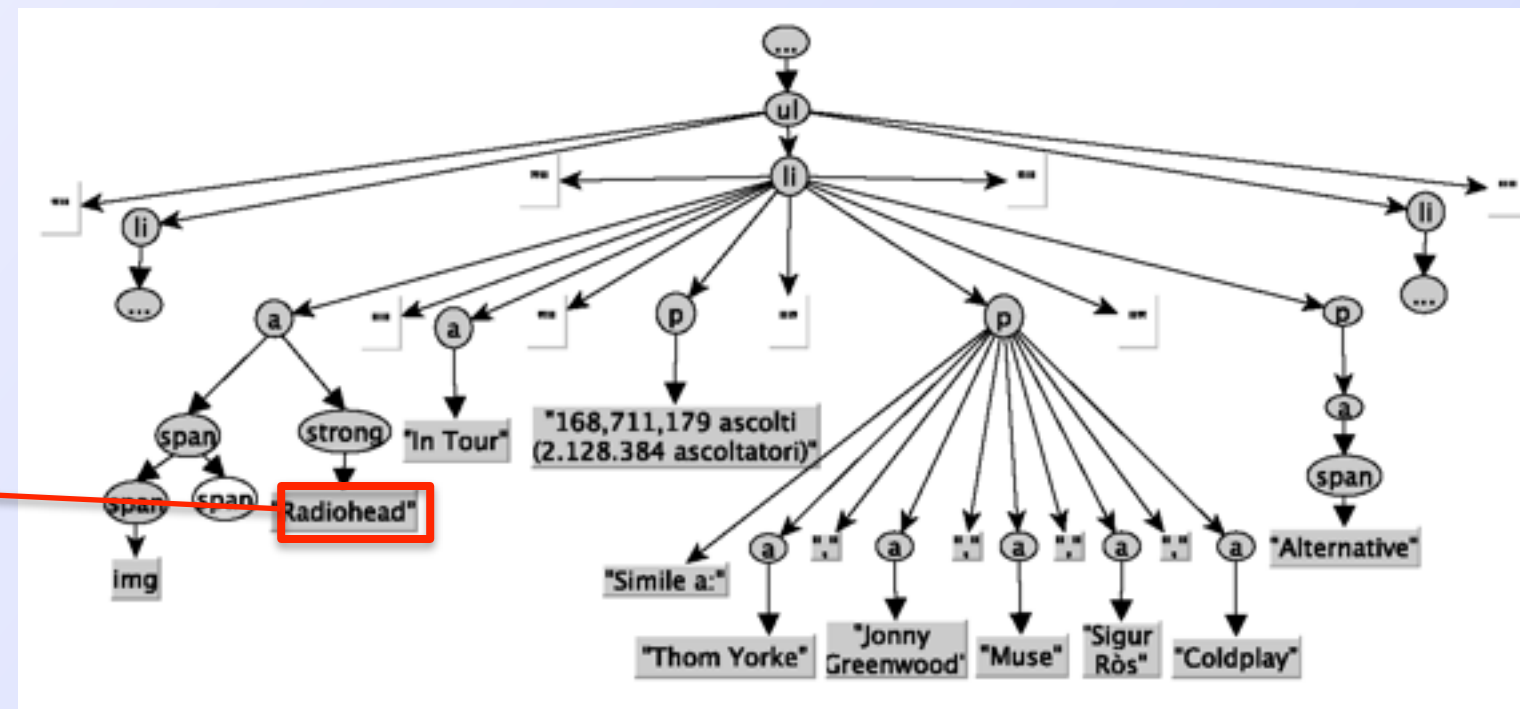


# Presentation-Oriented Documents

## Web page



## Document object model (DOM)

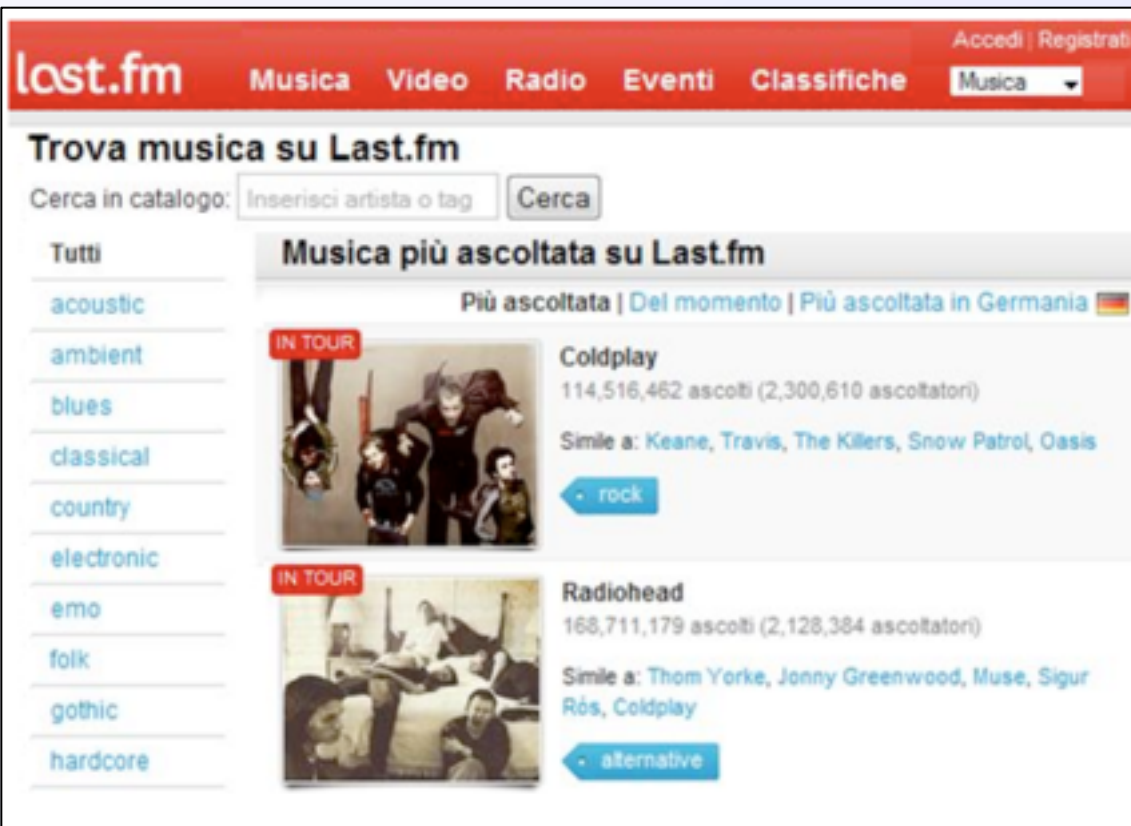


- Spatial arrangement rarely explicit
  - but hidden in complex nesting of layout elements in DOM
  - **difficult:** intricate tree structure conceptually difficult to query
  - **not robust:** small changes in layout structure brake a wrapper

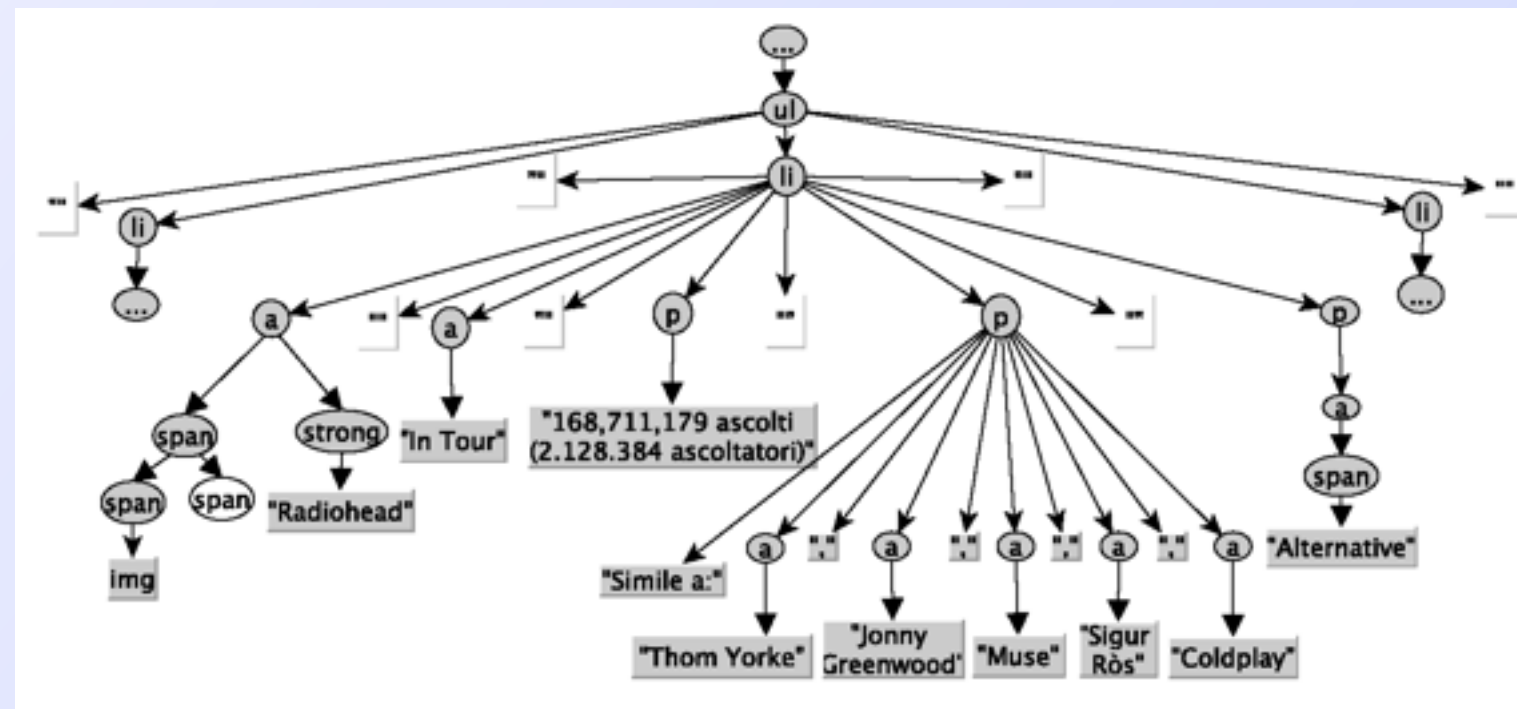


# Presentation-Oriented Documents

## Web page



## Document object model (DOM)

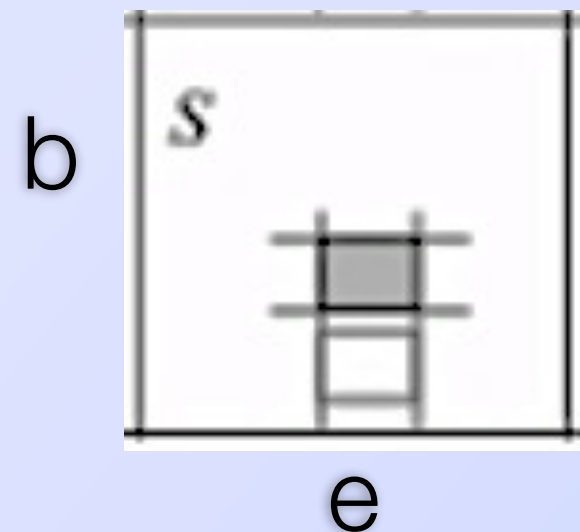


- Spatial arrangement rarely explicit
  - but hidden in complex nesting of layout elements in DOM
  - **difficult:** intricate tree structure conceptually difficult to query
  - **not robust:** small changes in layout structure brake a wrapper








# Spatial Relations among Nodes

- **Rectangular** algebra (RA):
  - Allen's temporal **interval** algebra extended to **2-dimensional** case [Balbiani et al.]
  - fine-grained, expressive model
  - many attractive properties
    - for query evaluation (invertibility)





# Spatial Relations among Nodes

-  **Rectangular algebra (RA):**
  -  Allen's temporal **interval** algebra extended to **2-dimensional** case [Balbiani et al.]
  -  fine-grained, expressive model
  -  many attractive properties
    -  for query evaluation (invertibility)

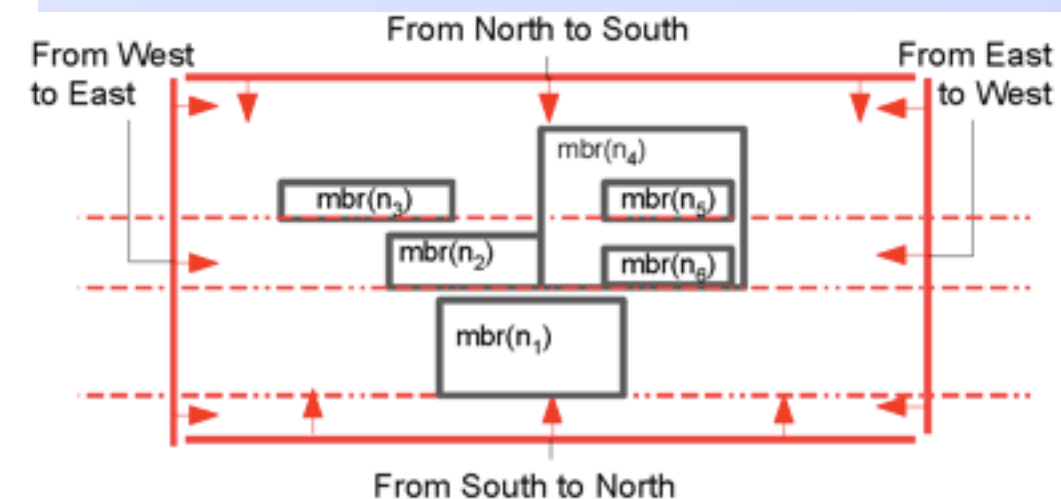
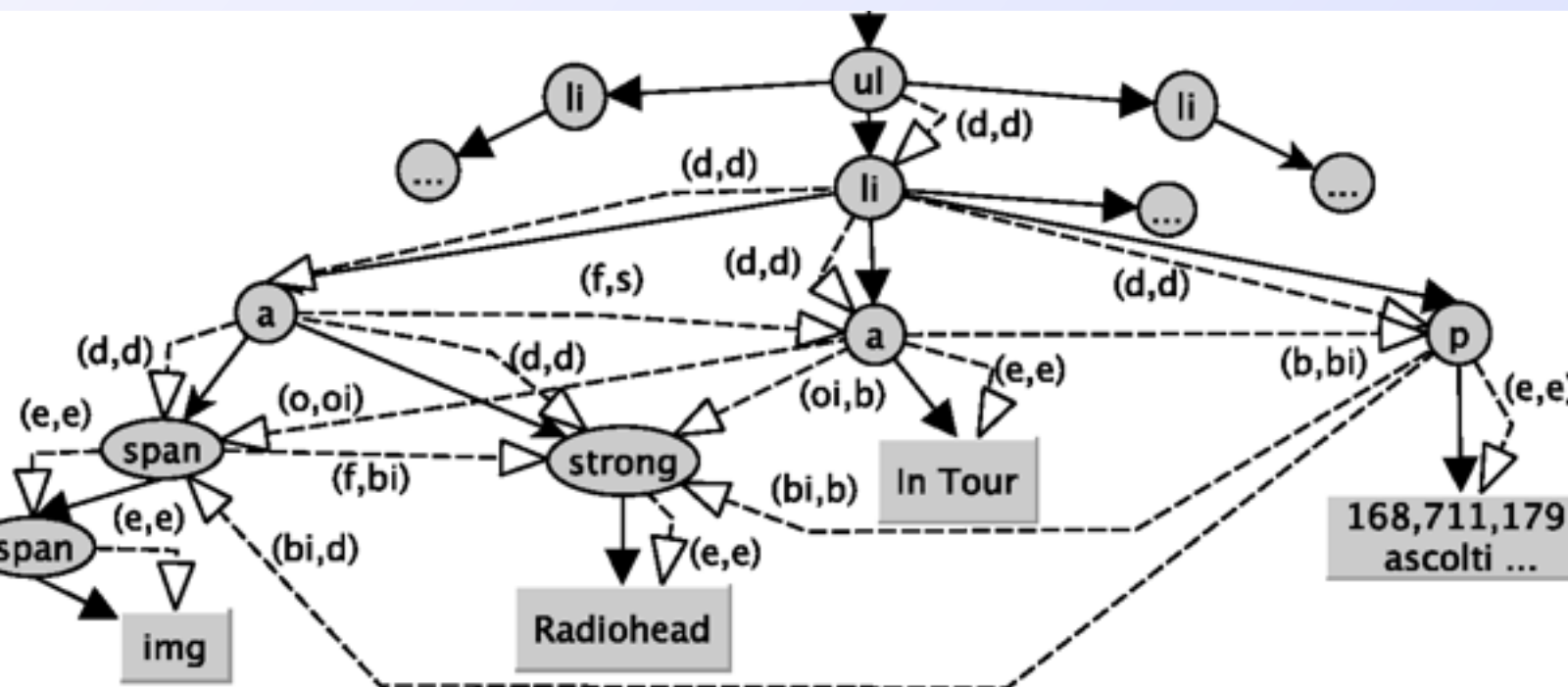
	b	m	o	fi	di	s	e	si	d	f	oi	mi	bi
bi													
mi													
oi													
si													
di													
f													
e													
fi													
d													
s													
o													
m													
b													



# Spatial Relations in Spatial DOM



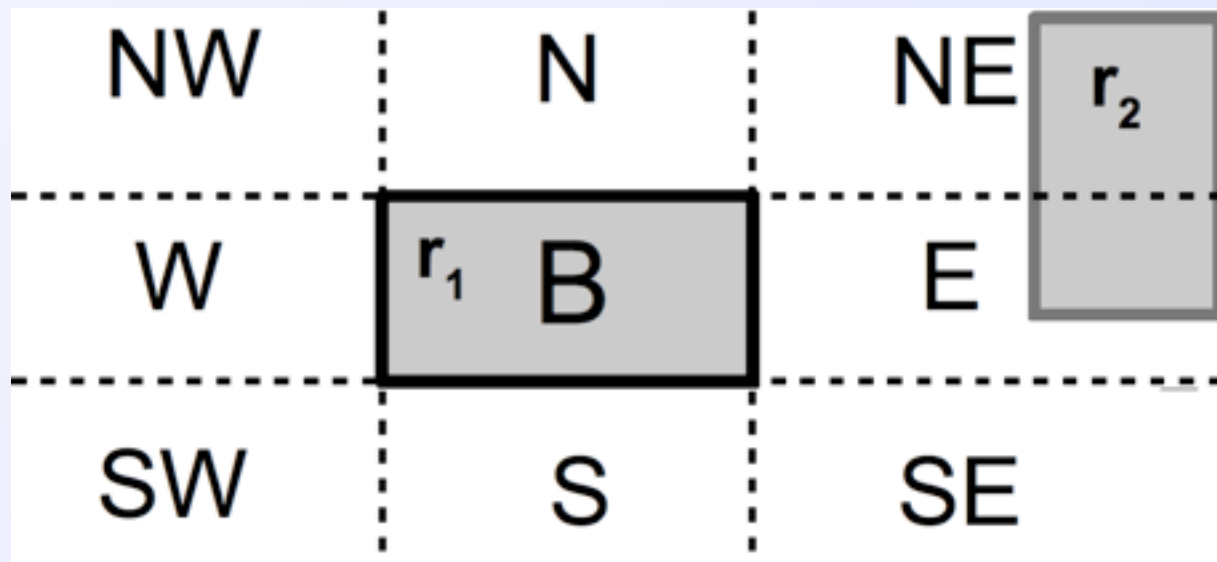
- SDOM extends DOM by
  - RA relations between nodes
  - spatial orders among nodes





# Spatial Relations among Nodes

## Rectangular Cardinal Relations (RCR)



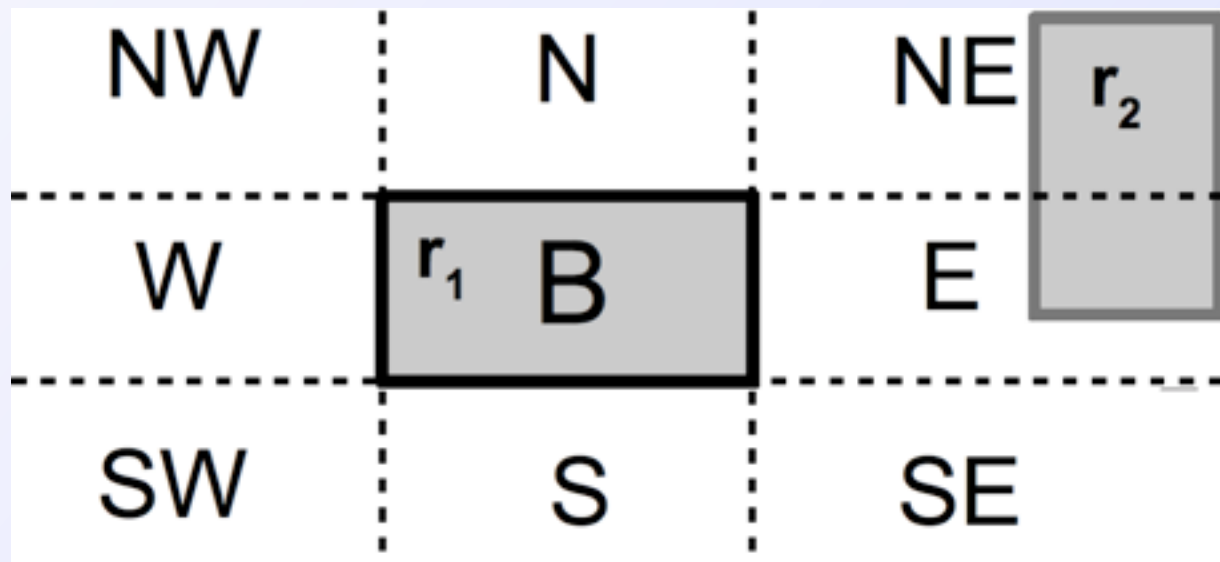
## Topological Relations

- inspired by the *Region Connection Calculus*
- contained (CD)
- container (ER)
- equivalent (EQ)



# Spatial Relations among Nodes

## Rectangular Cardinal Relations (RCR)



$r_1$  **E:NE**  $r_2$

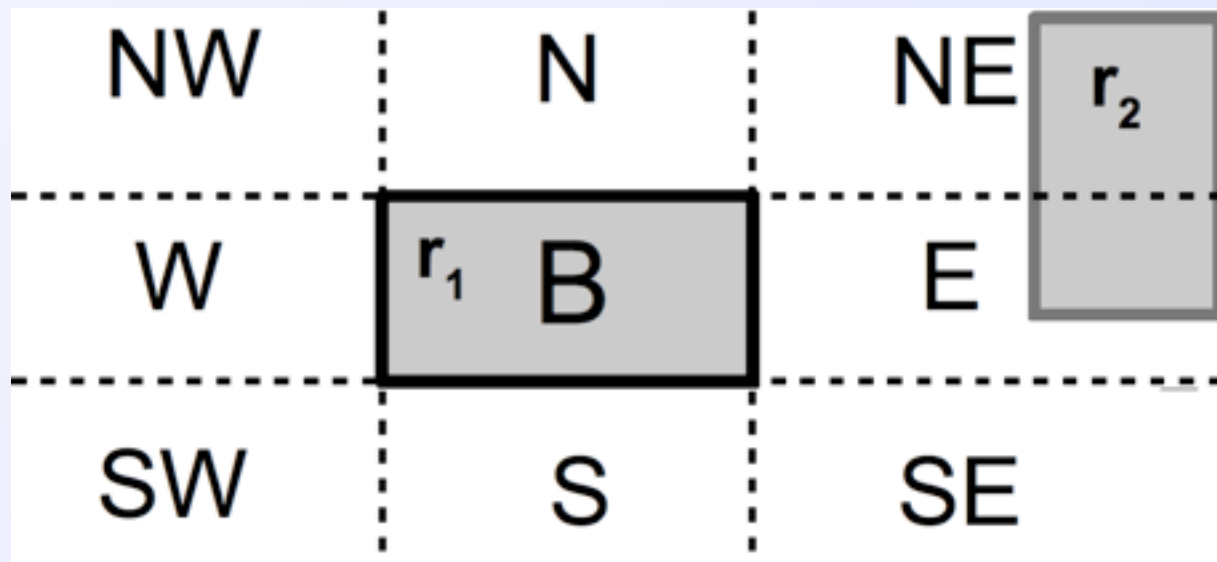
## Topological Relations

- inspired by the *Region Connection Calculus*
- contained (CD)
- container (ER)
- equivalent (EQ)



# Spatial Relations among Nodes

## Rectangular Cardinal Relations (RCR)



$r_1$  **E:NE**  $r_2$

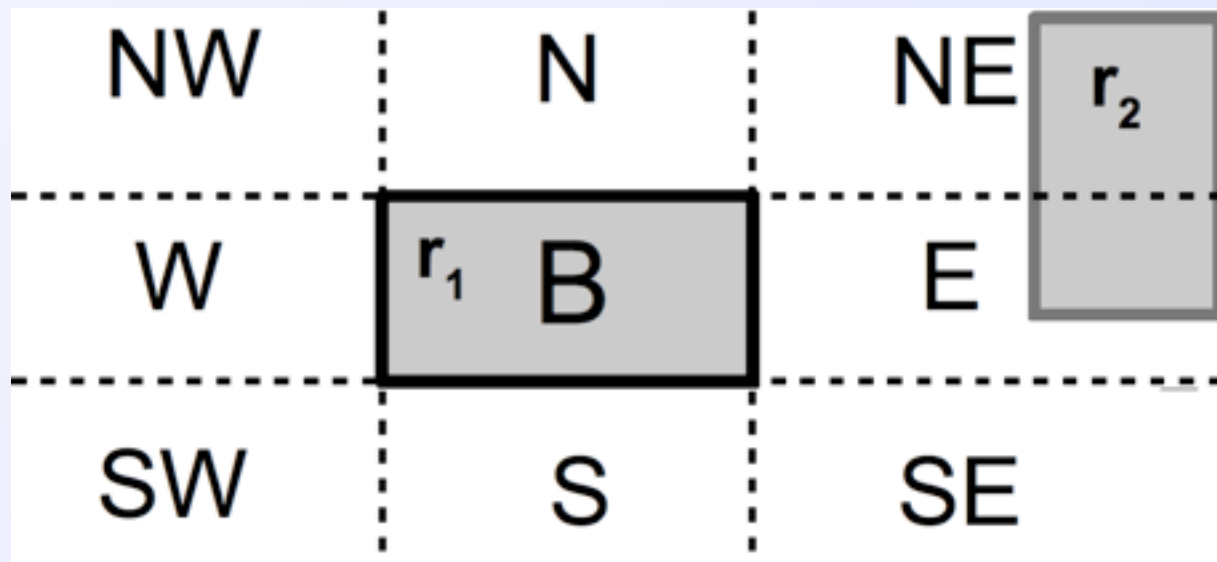
## Topological Relations

- inspired by the *Region Connection Calculus*
- contained (CD)
- container (ER)
- equivalent (EQ)



# Spatial Relations among Nodes

## Rectangular Cardinal Relations (RCR)



$r_1$  **E:NE**  $r_2$

Spatial models allows for expressing disjunctive relations among regions

## Topological Relations

- inspired by the *Region Connection Calculus*
- contained (CD)
- container (ER)
- equivalent (EQ)



# Extending XPath with Spatial Queries

- **Spatial XPath (XPath)** extends XPath 1.0
  - Intuitive path notation `/axis::nodetest[pred1]*`
  - A new set of **spatial axes** that allow for selecting nodes that have a specific spatial relation w.r.t. context nodes;
  - A new set of **spatial position functions** that allow for expressing predicates working on positions of nodes in the plane.
  - Abilities to query **attributes and stylesheet properties** as they are rendered from browsers (such as font size, font color, etc.).
- **Semantics:**
  - natural semantics for spatial queries
- **Complexity:**
  - maintains **polynomial combined complexity**

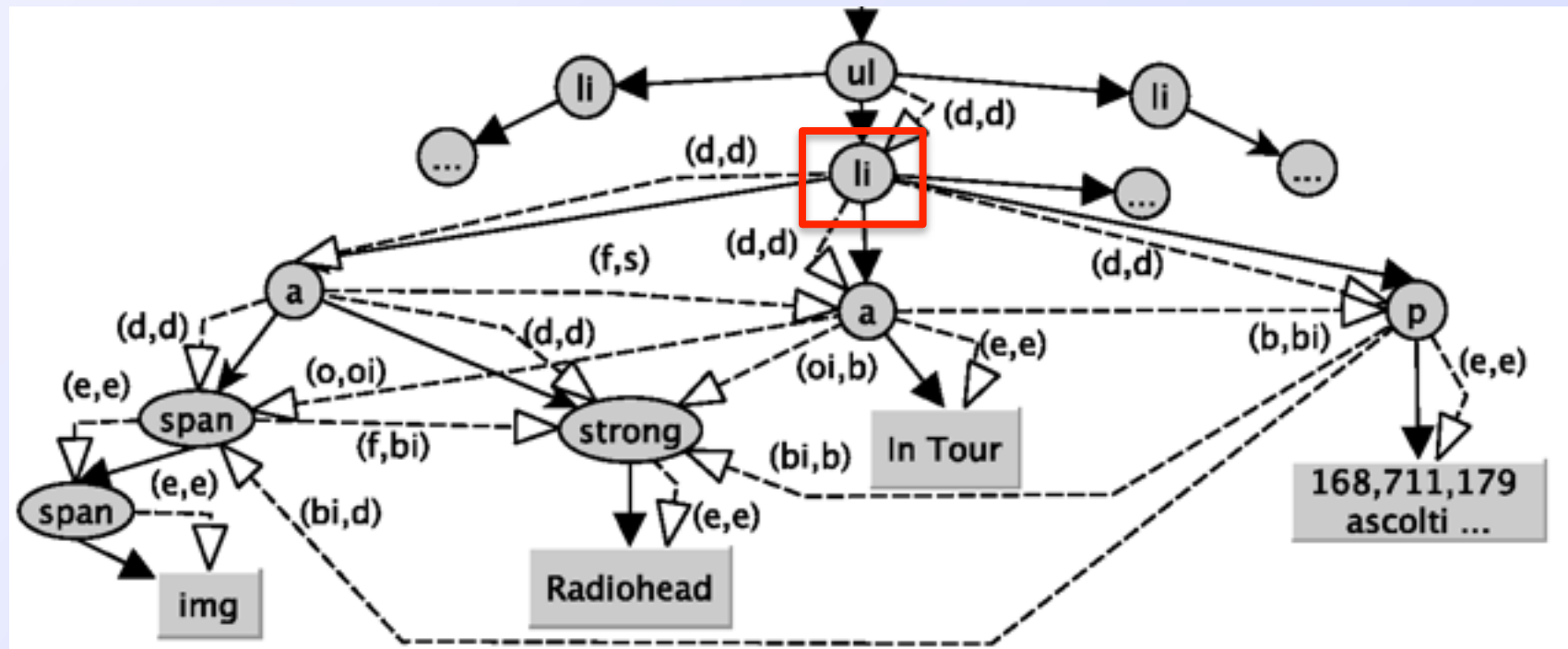


# Extending XPath with Spatial Queries

```
for $li in doc("last-fm.htm")
  /div[@id= 'content']//ul//li
return
  <music-band>
    <name>
      {$li/a/strong/text()}
    </name>
    ...
  </music-band>
```



# Extending XPath with Spatial Queries



```
for $li in doc("last-fm.htm")
  /div[@id= 'content']//ul//li
```

```
return
```

```
<music-band>
```

```
<name>
```

```
{ $li/a/strong/text() }
```

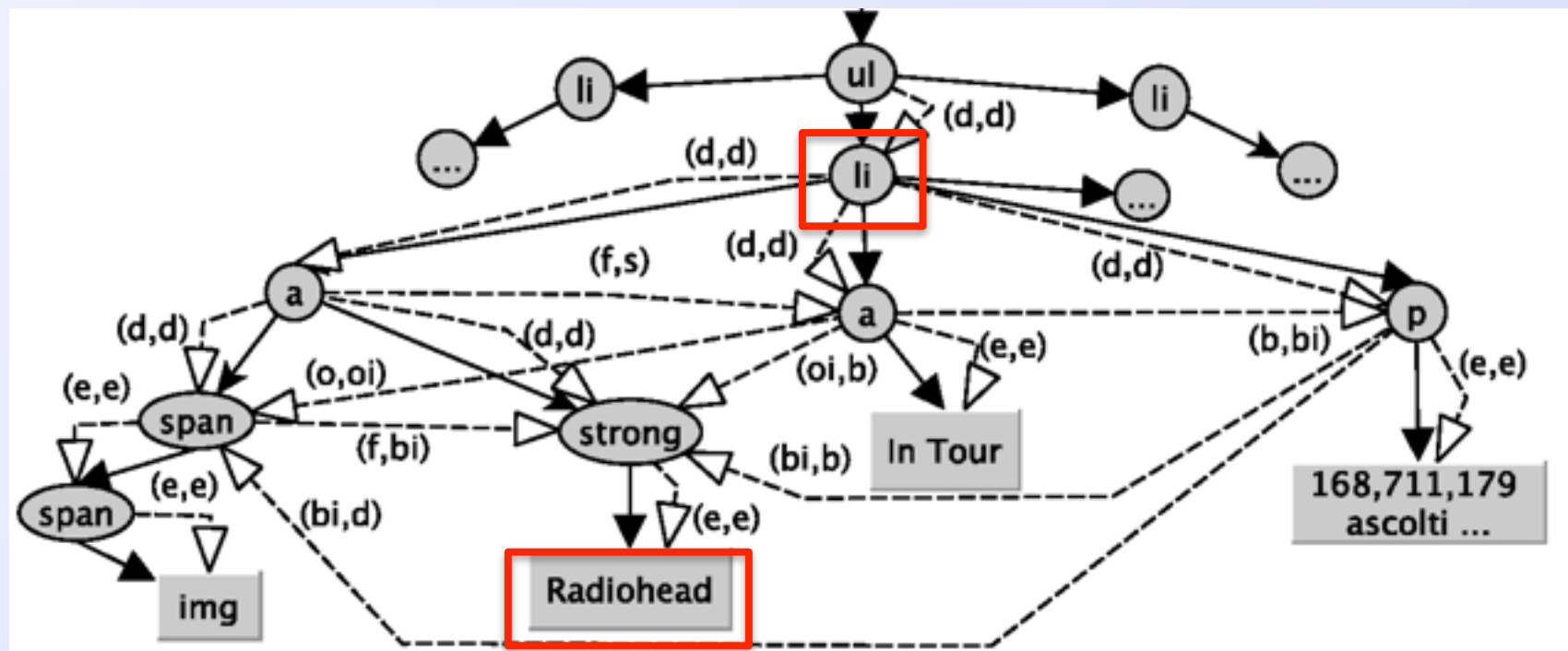
```
</name>
```

```
...
```

```
</music-band>
```



# Extending XPath with Spatial Queries



```
for $li in doc("last-fm.htm")
  /div[@id= 'content']//ul//li
```

```
return
```

```
<music-band>
```

```
<name>
```

```
{ $li/a/strong/text() }
```

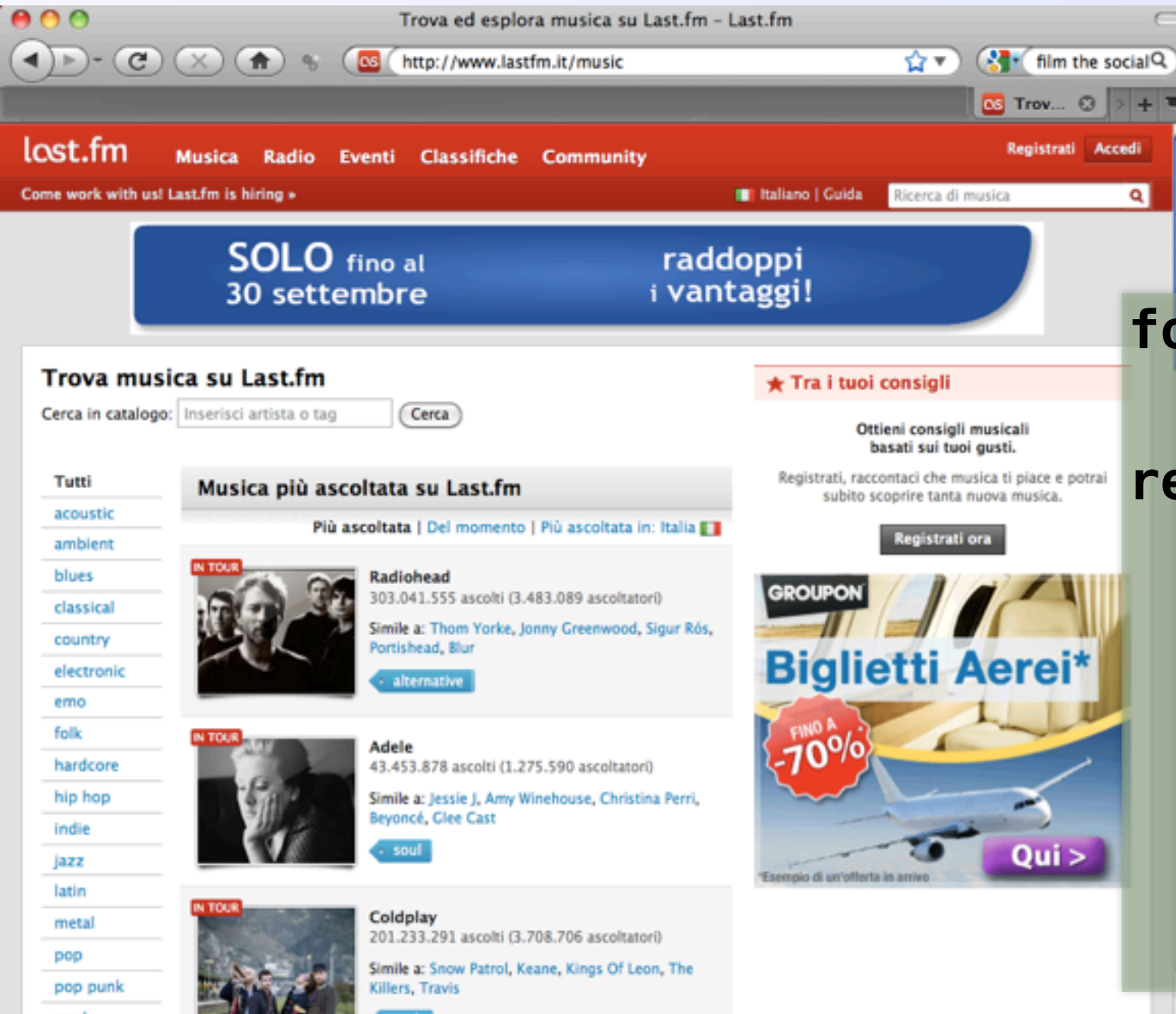
```
</name>
```

```
...
```

```
</music-band>
```



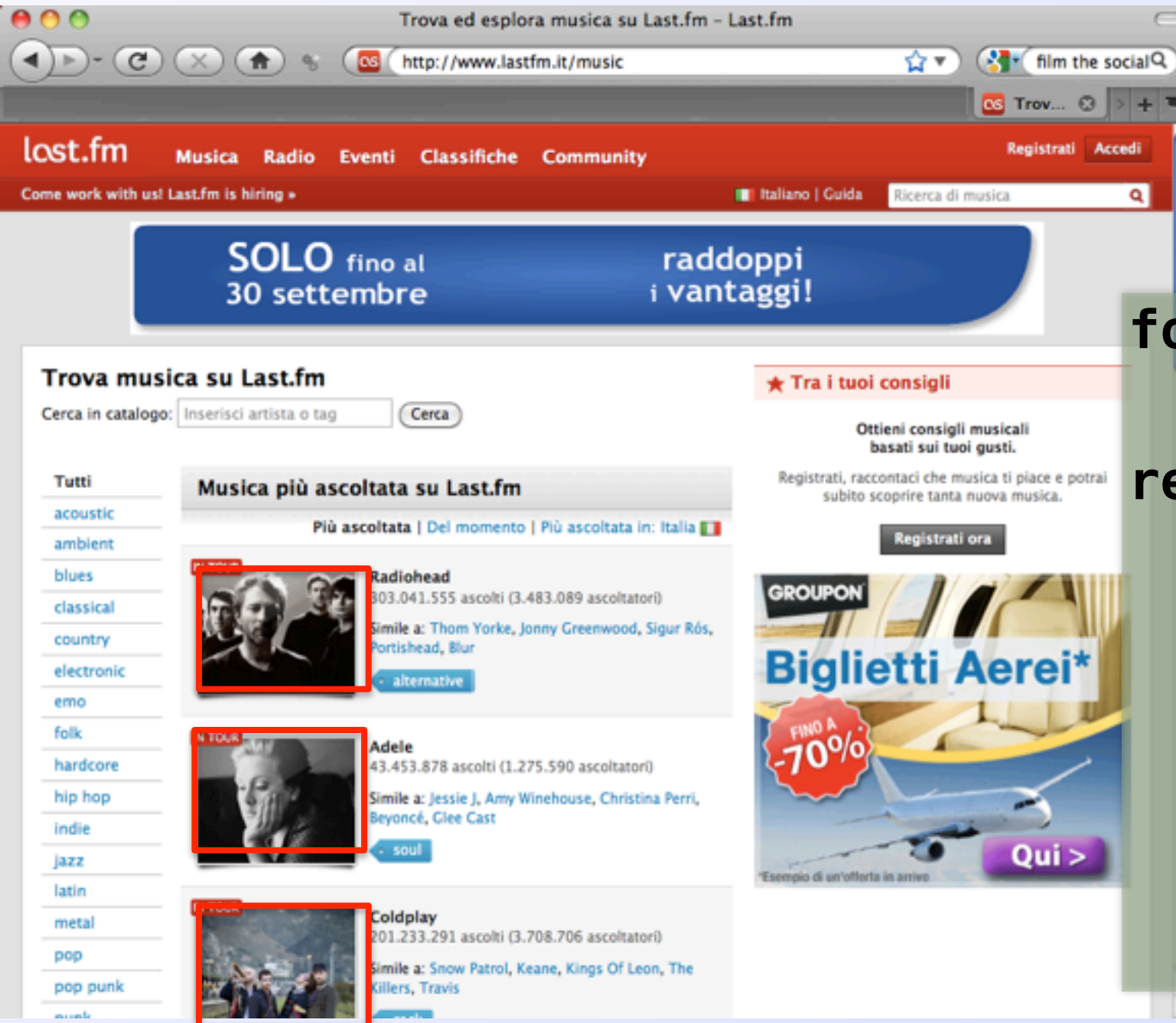
# Extending XPath with Spatial Queries



```
for $img in doc("last-fm.htm")
  /CD::img[N|S::img]
return
  <music-band>
    <name>
      {$img/E::text[N,1]}
    </name>
    ...
  </music-band>
```



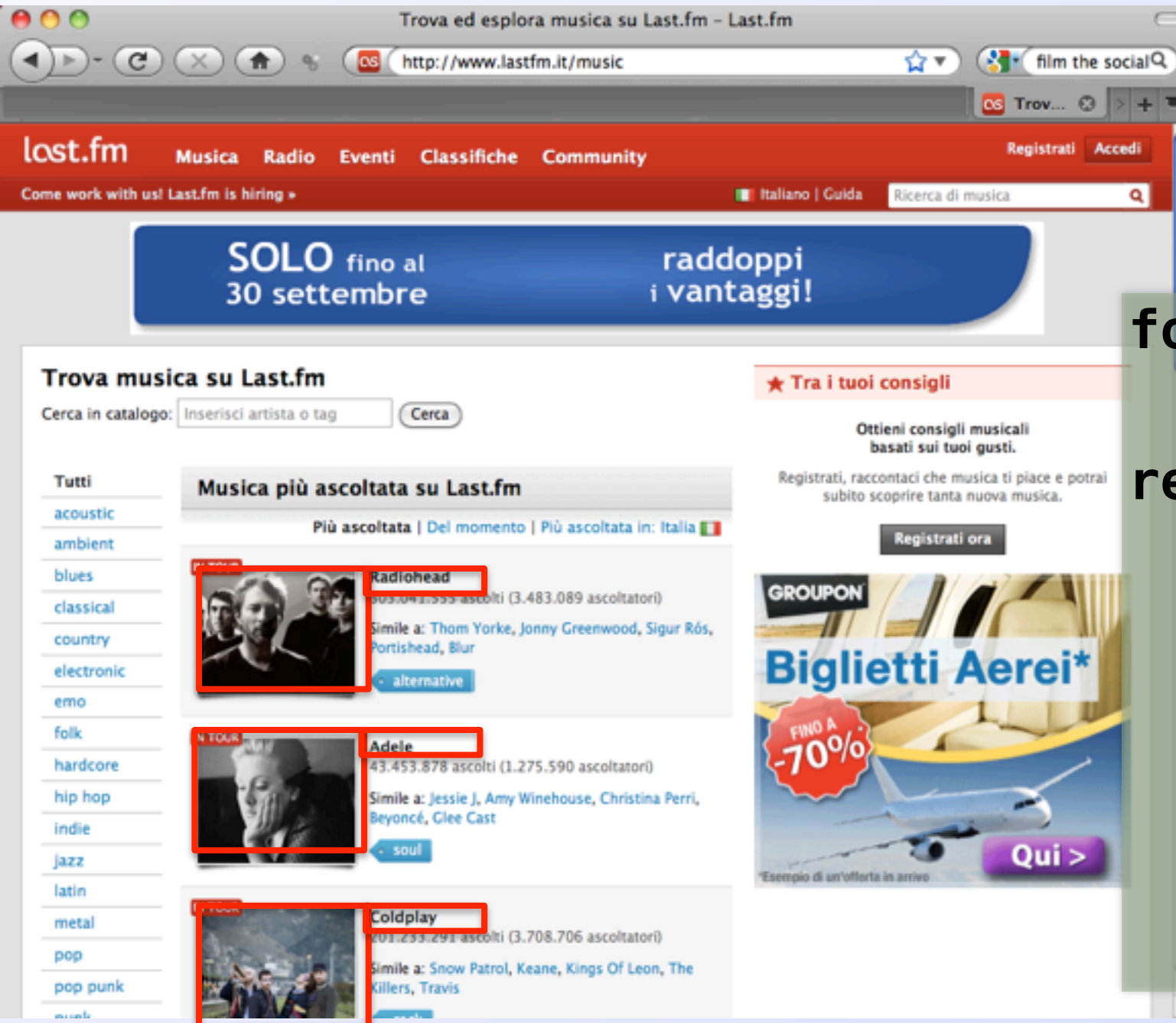
# Extending XPath with Spatial Queries



```
for $img in doc("last-fm.htm")
  /CD::img[N|S::img]
return
  <music-band>
    <name>
      {$img/E::text[N,1]}
    </name>
    ...
  </music-band>
```



# Extending XPath with Spatial Queries

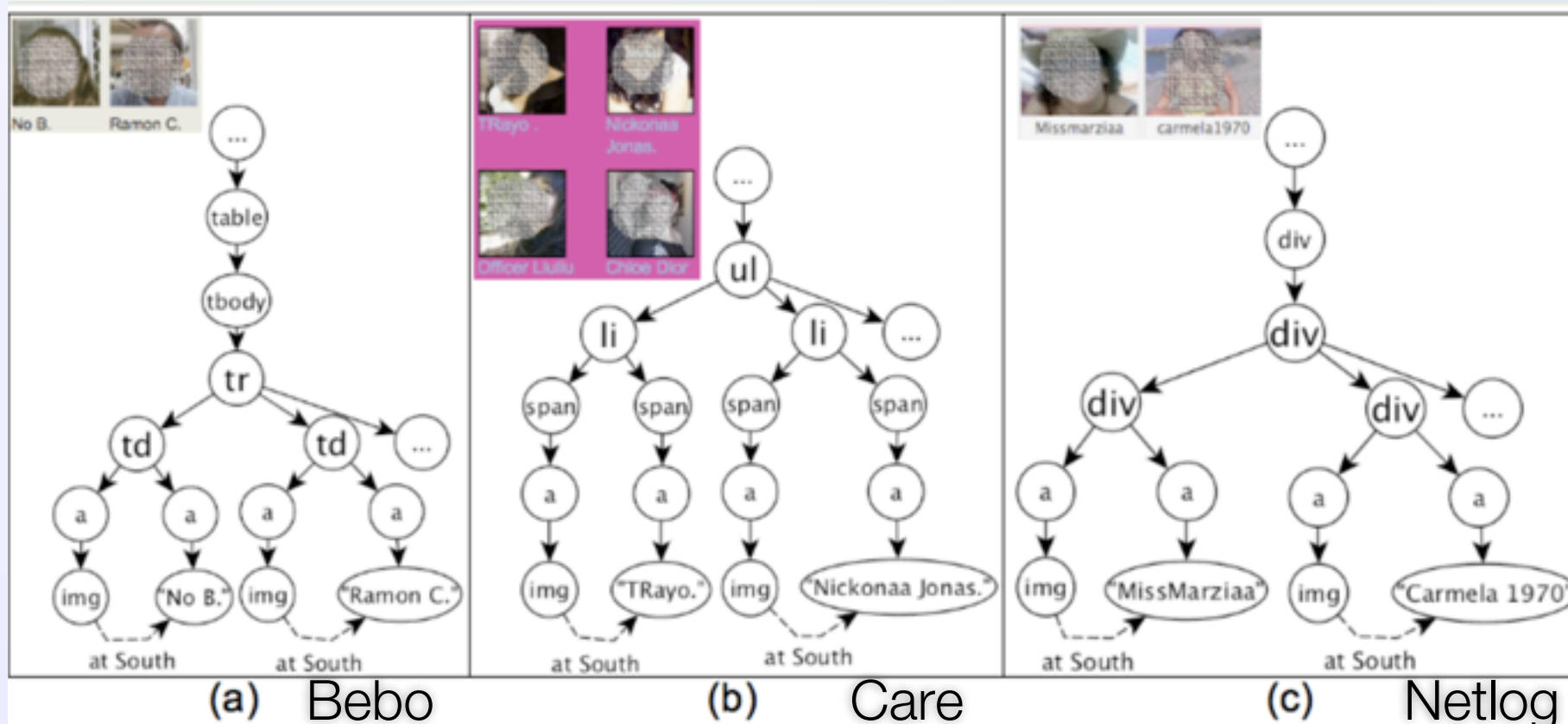


```
for $img in doc("last-fm.htm")
  /CD::img[N|S::img]
return
  <music-band>
    <name>
      {$img/E::text[N,1]}
    </name>
    ...
  </music-band>
```



# Extending XPath with Spatial Queries

- Goal: Extract friend lists from different social networks



```
for $img in doc("http://www.bebo.com/friendlist.html")
```

```
  //img[ N|S|E|W::img ]
```

```
return <friend>
```

```
  <photo> {$img} </photo>
```

```
  <name> {$img/S::text()[N,1]} </name>
```

```
</friend>
```



- GUI that supports **spatial querying**





# Summary of Complexity



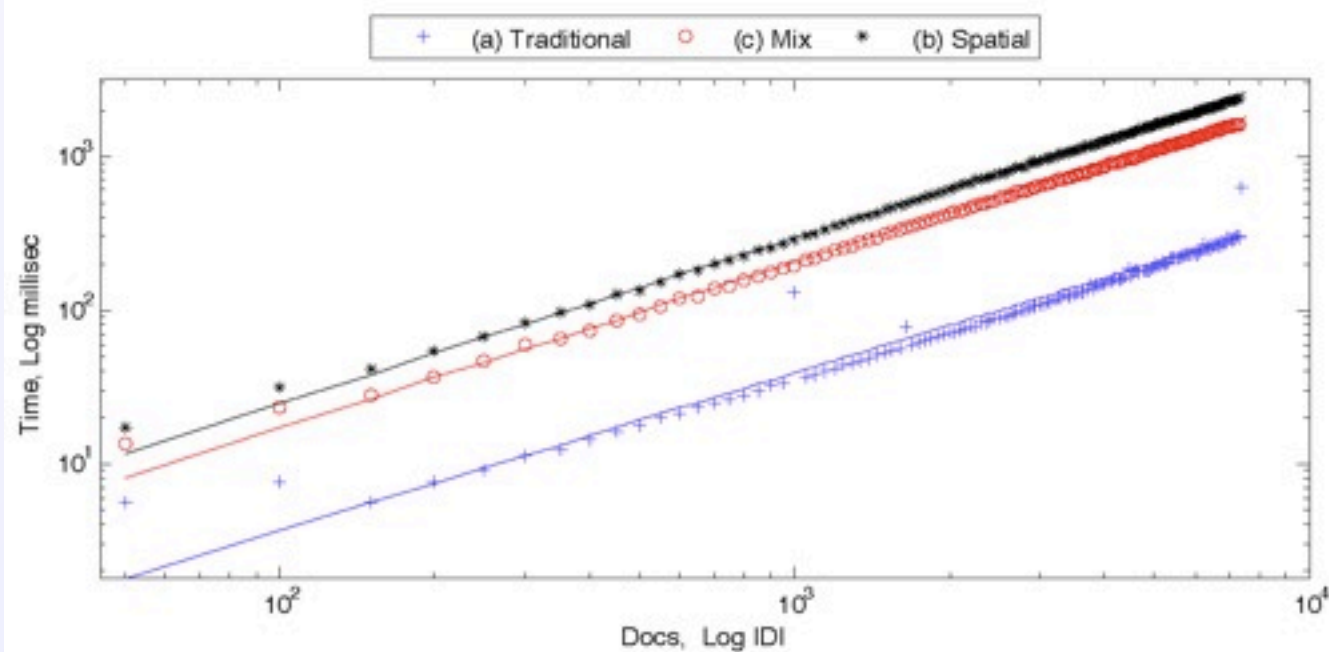
# Summary of Complexity

XPath 1.0		SXPath	
<b>Core</b>	space: $O(n \cdot q)$ time: $O(n \cdot q)$	<b>Spatial Core</b>	space: $O(n^2 \cdot q)$ time: $O(n^2 \cdot q)$
<b>EWf</b>	space: $O(n \cdot q^2)$ time: $O(n^2 \cdot q^2)$	<b>SWF</b>	space: $O(n^2 \cdot q^2)$ time: $O(n^3 \cdot q^2)$
<b>Full XPath</b>	<b>space: <math>O(n^2 \cdot q^2)</math></b> <b>time: <math>O(n^4 \cdot q^2)</math></b>	<b>Full SXPath</b>	<b>space: <math>O(n^2 \cdot q^2)</math></b> <b>time: <math>O(n^4 \cdot q^2)</math></b>

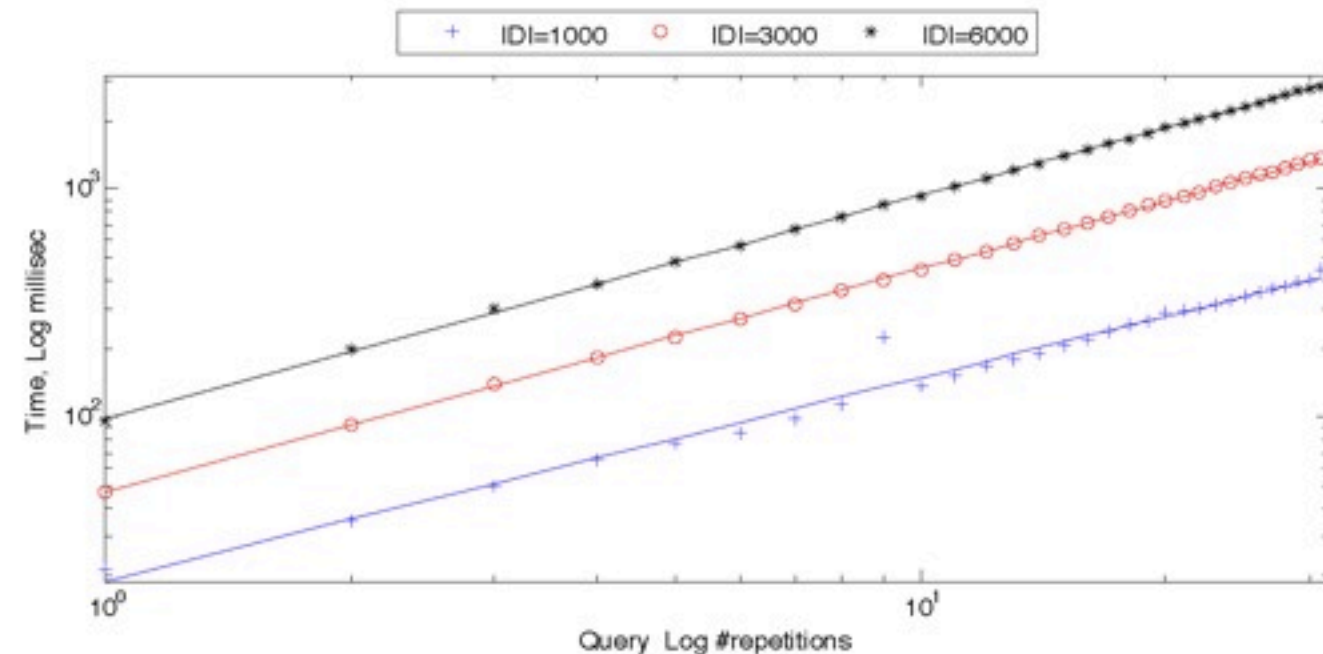


# Evaluation

## Data Efficiency of SXPath Query Evaluation



## Query Efficiency of SXPath Query Evaluation



The curves grows linear on log-log scale indicating the polynomial growth



# Summary

- XPath = XPath extended with spatial navigation
- SDOM = DOM + spatial relationships and orders
- XPath forms a stepping stone
  - minimal, but expressive extensions for spatial querying
  - basis of data extraction for presentation-oriented documents



# SCRAP



# SCRAP

B. Fazzinga, S. Flesca, A. Tagarelli: **Schema-based Web Wrapping**. *Knowl. Inf. Syst.* 26(1): 127-173 (2011)



# Hierarchical wrapping approaches

- Based on extraction rules
  - Identify the region(s) of the document to be extracted
  - Rules further extract subregions from the result of other rules
  - Rules often formulated in XPath, but may be SXPath, OXPath, ...
- **Observation:** These approaches do not exploit
  - **schema of desired information** in the extraction process
  - schema used by these approaches too simple
    - all the subelements have multiplicity 0..\*
  - more complex schema can be defined separately
    - verified a posteriori



# SCRAP approach

- A wrapper is a pair
  - (Schema, Set of Extraction Rules)
- **SCRAP:** New extraction strategy
  - schema information used in **identifying** information to be extracted:
    - A region is extracted only if it is possible to construct a **valid** element



# Schema-based Wrapper definition

- **Schema:** DTD, Entity-Relationship
  - each pair of nested elements  $(p_e, e)$  in the DTD
  - associated to an extraction rule.
- **Rules:** Works with **different types** of extraction rules
  - An extraction rule  $r$  is a function associating
    - a sequence  $S$  of document regions
    - to a document region.
  - XPath-based rules are adopted in the implementation



# Extraction: an example

```
<!ELEMENT doc (store)>
<!ELEMENT store (book+)>
<!ELEMENT book (title, author+, (customer_rate | no_rate), price, year)>
<!ELEMENT title (#PCDATA)>
<!ELEMENT author (name)>
<!ELEMENT name (#PCDATA)>
<!ELEMENT customer_rate (rate)>
<!ELEMENT no_rate EMPTY>
<!ELEMENT rate (#PCDATA)>
<!ELEMENT price (#PCDATA)>
<!ELEMENT year (#PCDATA)>
```



# Extraction: an example

Regions extracted for  
book

The screenshot shows the Amazon.com homepage with a search bar and navigation links. The search results are for the category 'Books > Subjects > Computers & Internet > Programming > Algorithms > General'. The results are sorted by 'Featured Items'.

Rank	Book Title	Author	Format	Release Date	Avg. Customer Review	Price
1.	Introduction to Algorithms, Second Edition	Thomas H. Cormen, et al	Hardcover	September 1, 2001	★★★★★	Used & new from \$12.95
2.	Microsoft Outlook Programming, Jumpstart for Administrators, Developers, and Power Users	Sue Mosher	Paperback	September 6, 2002	★★★★★	Used & new from \$30.42
3.	Information Theory, Inference & Learning Algorithms	David J. C. MacKay	Hardcover	June 15, 2002	★★★★★	Used & new from \$35.34



# Extraction: an example

Regions extracted for  
book

The screenshot shows the Amazon.com website with a search for "Books" in the "Computers & Internet > Programming > Algorithms > General" category. Three books are listed, each with a blue box highlighting its details and a separate blue box highlighting the "Used & new from" price. Arrows point from the text "Regions extracted for book" to these regions.

Rank	Book Title	Author	Format	Release Date	Avg. Customer Review	List Price	Buy new	In-store Pickup	Used & new from
1.	Introduction to Algorithms, Second Edition	Thomas H. Cormen, et al	Hardcover	September 1, 2001	★★★★★	\$80.00	\$80.00	\$71.96	\$12.95
2.	Microsoft Outlook Programming, Jumpstart for Administrators, Developers, and Power Users	Sue Mosher	Paperback	September 6, 2002	★★★★★	\$44.99	\$30.59		\$30.42
3.	Information Theory, Inference & Learning Algorithms	David J. C. MacKay	Hardcover	June 15, 2000	★★★★★	\$50.00	\$39.50		\$35.34



# Extraction: an example

<!ELEMENT book (title, author+, (customer\_rate | no\_rate), price, year)>

Regions extracted for book

The screenshot shows the Amazon.com website with a search for "Books" in the "Computers & Internet" category. Three book listings are shown, each with a blue box highlighting the extraction region. Arrows point from the "Regions extracted for book" text to these regions.

Rank	Title	Author	Format	Release Date	Avg. Customer Review	Price
1.	Introduction to Algorithms, Second Edition	by Thomas H. Cormen, et al	Hardcover	September 1, 2001	★★★★★	Used & new from \$12.95
2.	Microsoft Outlook Programming, Jumpstart for Administrators, Developers, and Power Users	by Sue Mosher	Paperback	September 6, 2002	★★★★★	Used & new from \$30.42
3.	Information Theory, Inference & Learning Algorithms	by David J. C. MacKay (Author)	Hardcover	June 15, 2000	★★★★★	Used & new from \$35.34



# Extraction: an example

```
<!ELEMENT book (title, author+, (customer_rate | no_rate), price, year)>
```

Regions extracted for  
book

The screenshot shows the Amazon.com website with search results for 'Books'. The breadcrumb trail is 'Books > Subjects > Computers & Internet > Programming > Algorithms > General'. The search results are sorted by 'Featured Items'. Three book entries are highlighted with blue boxes, and their corresponding price information is highlighted with red boxes. Arrows point from the 'Regions extracted for book' text to these regions.

Rank	Title	Author	Format	Release Date	Avg. Customer Review	List Price	Buy new	In-store Pickup	Used & new from
1.	Introduction to Algorithms, Second Edition	by Thomas H. Cormen, et al	Hardcover	-- September 1, 2001	★★★★★	\$80.00	<del>\$80.00</del>	\$71.96	<del>\$12.95</del>
2.	Microsoft Outlook Programming, Jumpstart for Administrators, Developers, and Power Users	by Sue Mosher	Paperback	-- September 6, 2002	★★★★★	<del>\$44.99</del>	\$30.59		<del>\$30.42</del>
3.	Information Theory, Inference & Learning Algorithms	by David J. C. MacKay (Author)	Hardcover	-- June 15, 2000	★★★★★	<del>\$50.00</del>	\$39.50		<del>\$35.34</del>

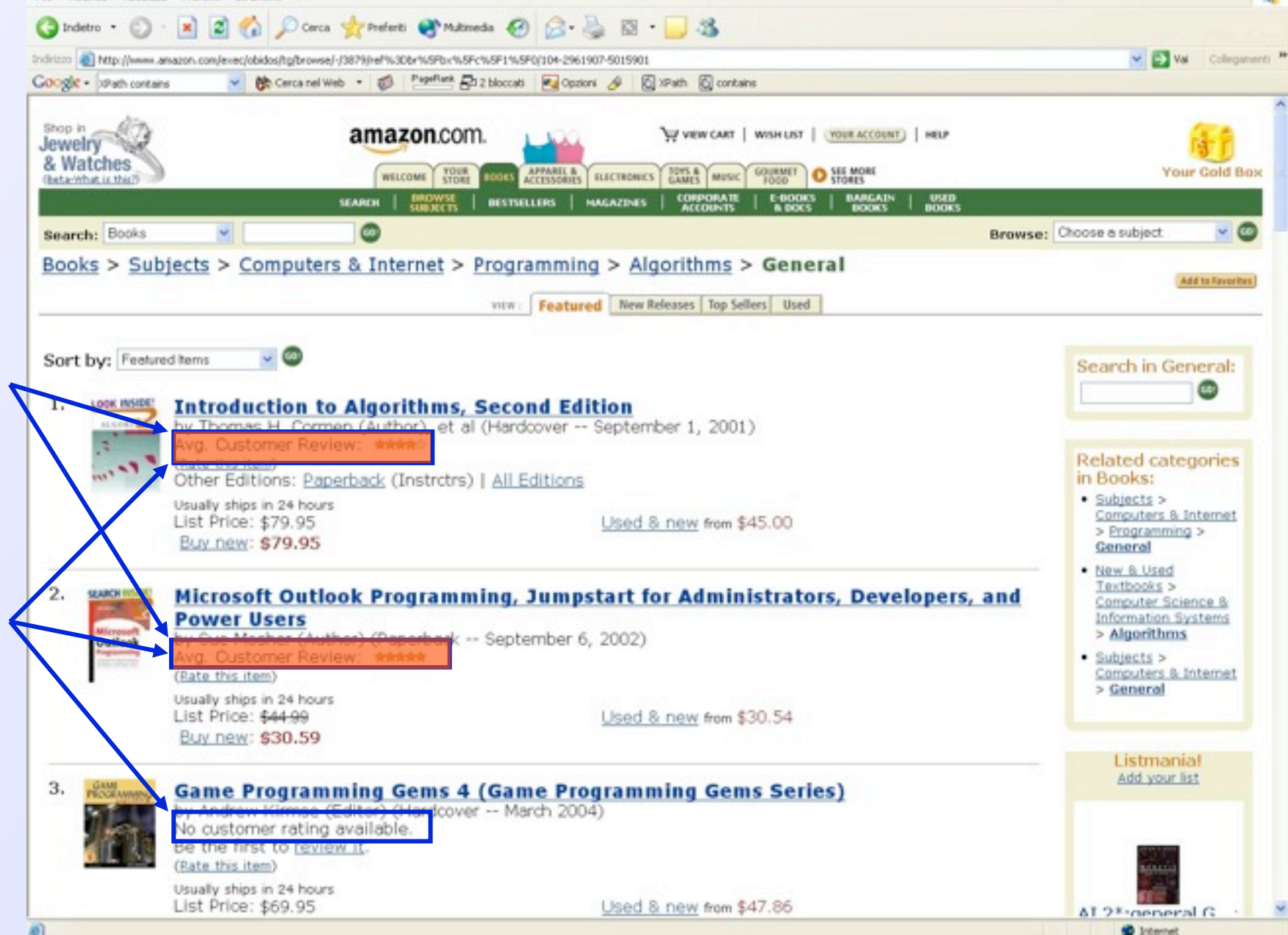


# Multiple extraction alternatives

```
<!ELEMENT book (title, author+, (customer_rate | no_rate), price, year)>
```

Regions extracted for  
customer\_rate

Regions extracted for  
no\_rate





# Preferred Element Content Model

- If there are multiple extraction model one is preferred
  - based on **syntactic/semantic conditions**

Preferred to

```
<book>
  <title> ... </title>
  <author> ... </author>
  ...
  <author> ... </author>
  <customer_rate> ... <customer_rate>
  <price> ... <price>
  <year> ... <year>
</book>
```

```
<book>
  <title> ... </title>
  <author> ... </author>
  ...
  <author> ... </author>
  <no_rate> ... <no_rate>
  <price> ... <price>
  <year> ... <year>
</book>
```



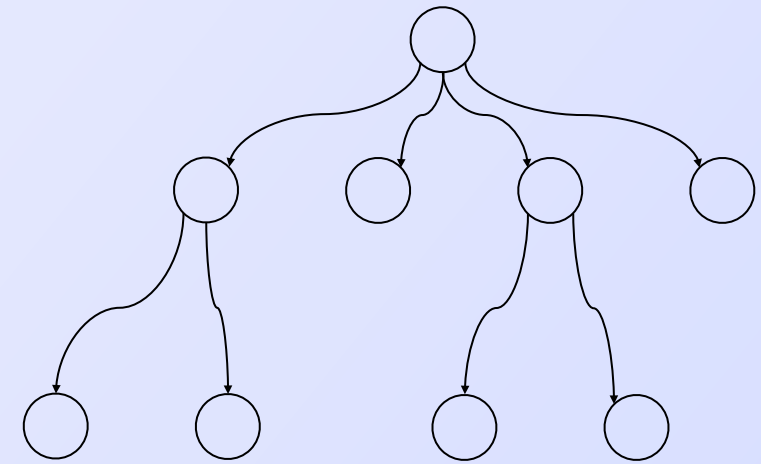
# Wrapper evaluation

- Run the extraction rules to obtain a tree of extraction candidates  $T$
- Work bottom-up on  $T$  (starting from its leaves)
  - If there is no valid content for  $e$  then remove  $e$
  - build the preferred content of  $e$  and remove the remaining sub-candidates from  $T$



# Wrapper evaluation

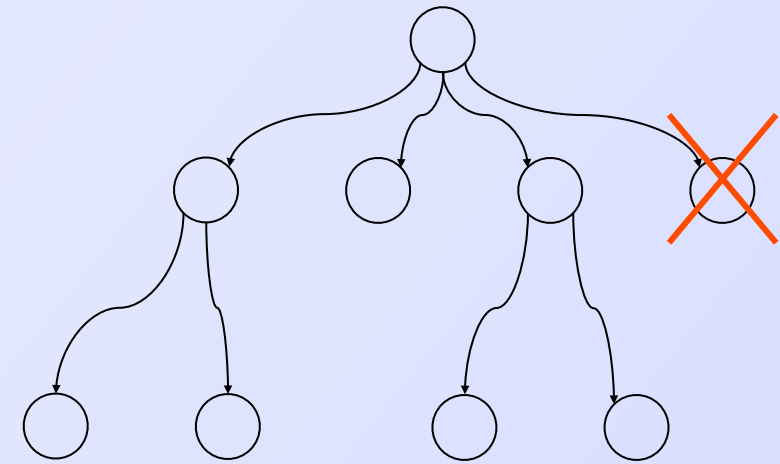
- Run the extraction rules to obtain a tree of extraction candidates  $T$
- Work bottom-up on  $T$  (starting from its leaves)
  - If there is no valid content for  $e$  then remove  $e$
  - build the preferred content of  $e$  and remove the remaining sub-candidates from  $T$





# Wrapper evaluation

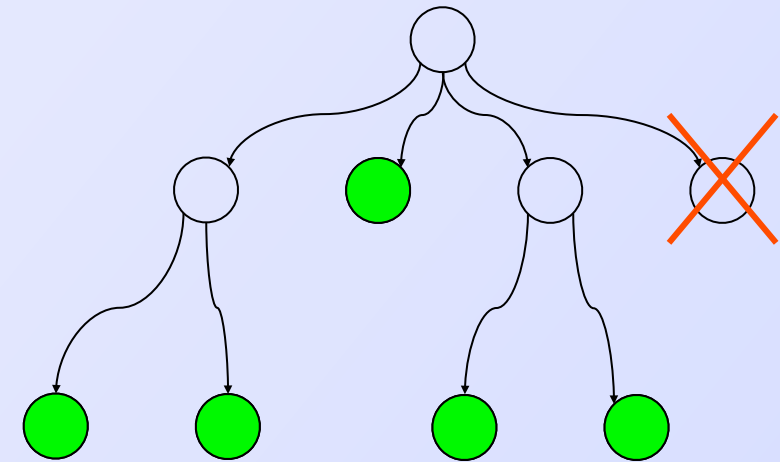
- Run the extraction rules to obtain a tree of extraction candidates  $T$
- Work bottom-up on  $T$  (starting from its leaves)
  - If there is no valid content for  $e$  then remove  $e$
  - build the preferred content of  $e$  and remove the remaining sub-candidates from  $T$





# Wrapper evaluation

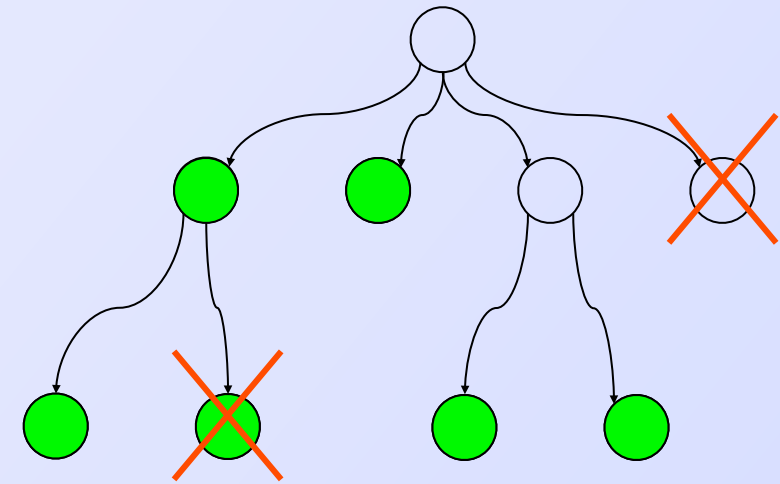
- Run the extraction rules to obtain a tree of extraction candidates  $T$
- Work bottom-up on  $T$  (starting from its leaves)
  - If there is no valid content for  $e$  then remove  $e$
  - build the preferred content of  $e$  and remove the remaining sub-candidates from  $T$





# Wrapper evaluation

- Run the extraction rules to obtain a tree of extraction candidates  $T$
- Work bottom-up on  $T$  (starting from its leaves)
  - If there is no valid content for  $e$  then remove  $e$
  - build the preferred content of  $e$  and remove the remaining sub-candidates from  $T$





# Advanced features

- SCRAP adopts a wrapper generalization strategy
  - Improves wrapper **robustness**
  - Works **without need of manual labeling** of example



# Experimental validation

## Average performance of SCRAP wrappers before and after generalization

Wrapper	Before generalization		After generalization	
	<i>P</i>	<i>R</i>	<i>P</i>	<i>R</i>
<i>Amazon</i> book list	0.83	0.83	0.99	1.0
<i>ANSA</i> home	1.0	0.88	1.0	1.0
<i>ANSA</i> top-news	0.75	0.74	0.99	1.0
<i>ANSA</i> local-news	1.0	0.99	1.0	1.0
<i>IMDb</i> movie list	1.0	0.97	1.0	1.0
<i>Kelkoo</i> tv list	1.0	0.97	1.0	1.0
<i>Kelkoo</i> compare-prices	1.0	0.96	1.0	1.0
<i>Trulia</i> home advices	0.75	0.74	1.0	0.97



# Summary

- **SCRAP's** improves:
  - wrapper robustness
  - makes schema-dependent wrapper refinement independent from
    - wrapper language or wrapper induction approach
- **SCRAP's** limitations:
  - doesn't work with **recursive** rules
    - It is not possible to define an item in terms of its siblings



# Lixto



# Lixto

R. Baumgartner, S. Flesca, and G. Gottlob. 2001. **Visual Web Information Extraction with Lixto.** In *Very Large Data Bases (VLDB '01)*.

G. Gottlob and C. Koch. 2004. **Monadic datalog and the expressive power of languages for Web information extraction.** *J. ACM* 51, 1 (2004).

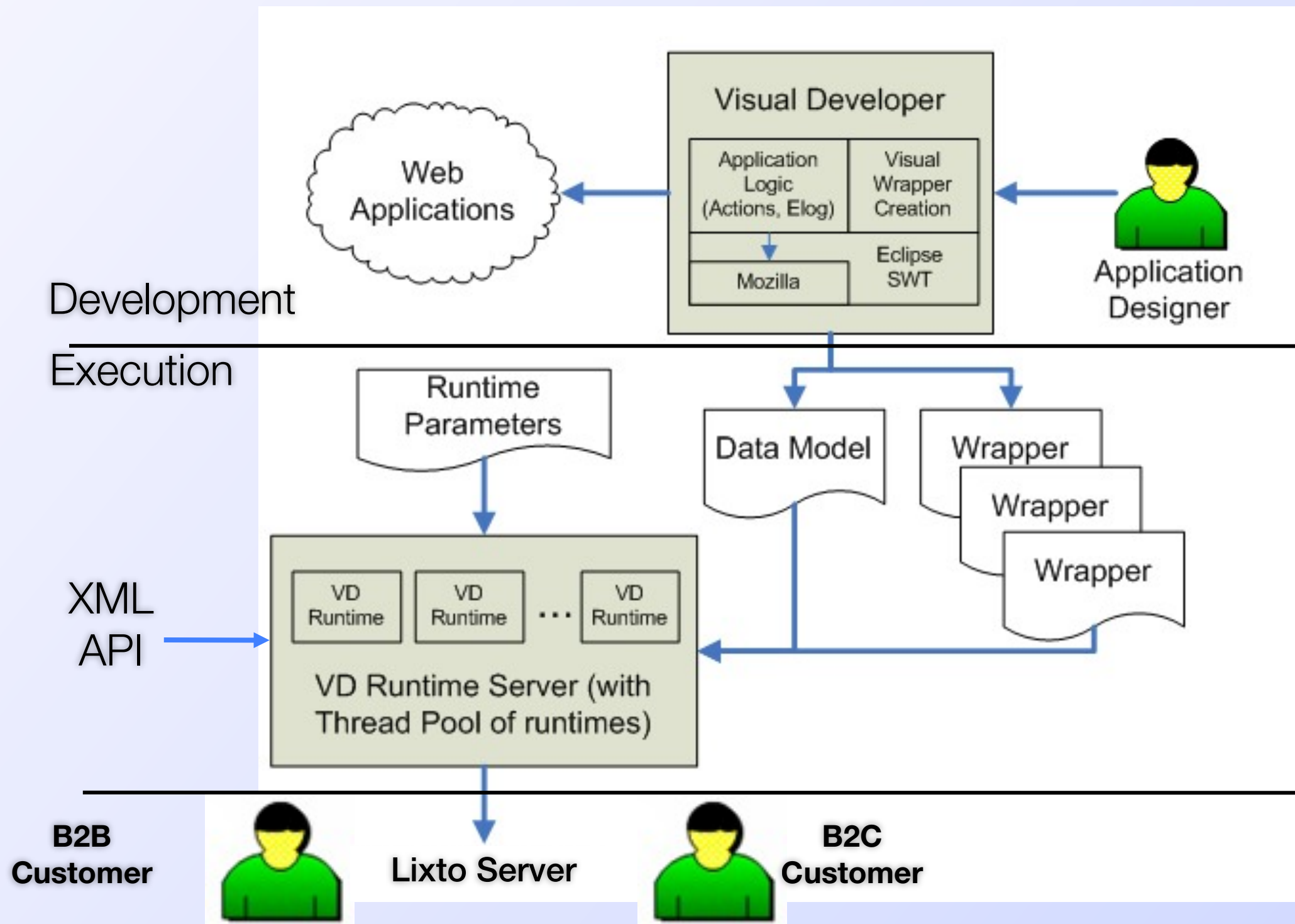


# Lixto: Visual Data Extraction

- **Lixto:** Full-fledged data extraction system
  - visual wrapper generation through **Lixto visual developer**
  - backed by Datalog-like language **Elog**
    - with optimal wrapper complexity and expressiveness
  - integrated into **Lixto suite**
    - including data integration, reporting, cloud execution, ...

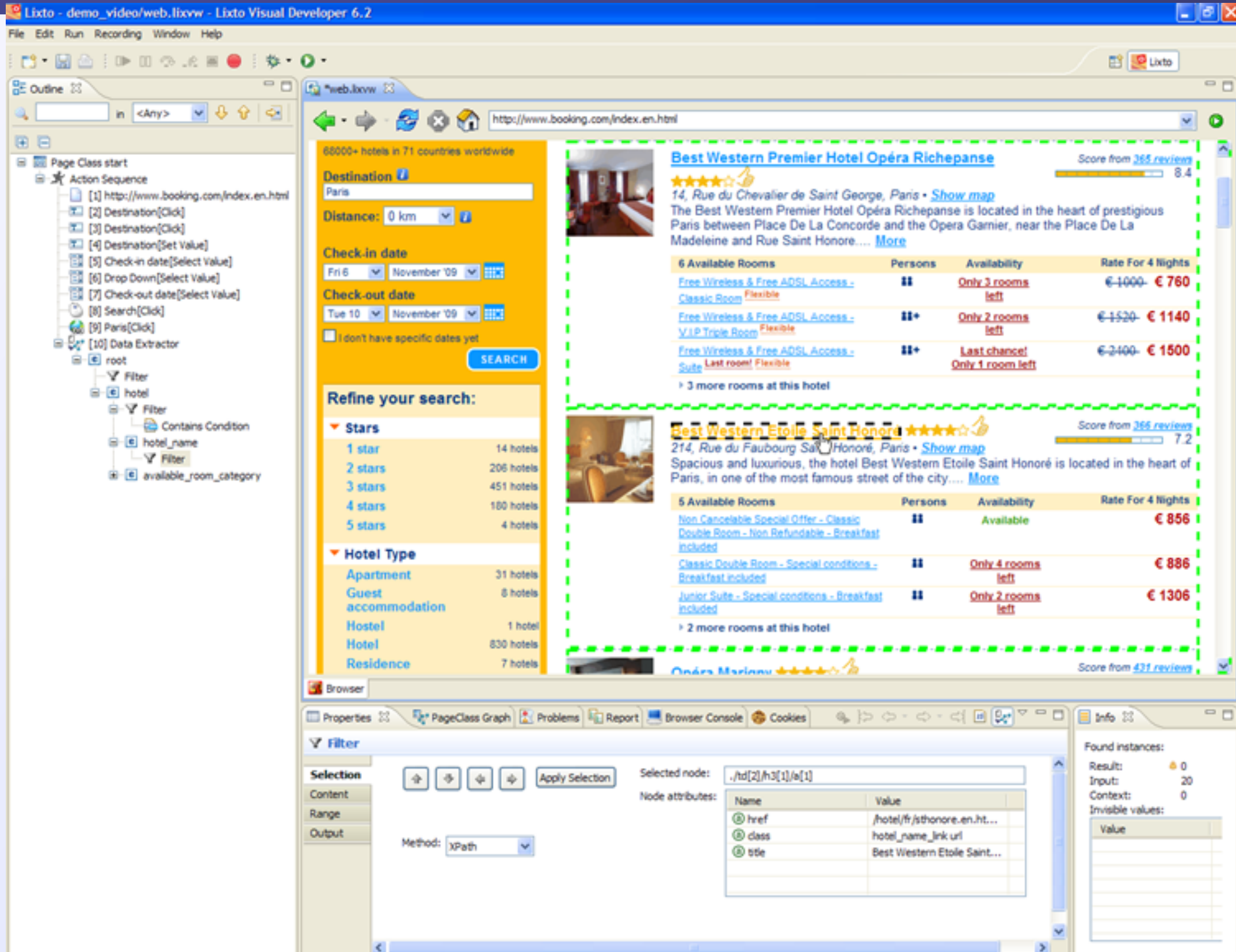


# Lixto Data Extraction Environment





# Lixto Visual Developer



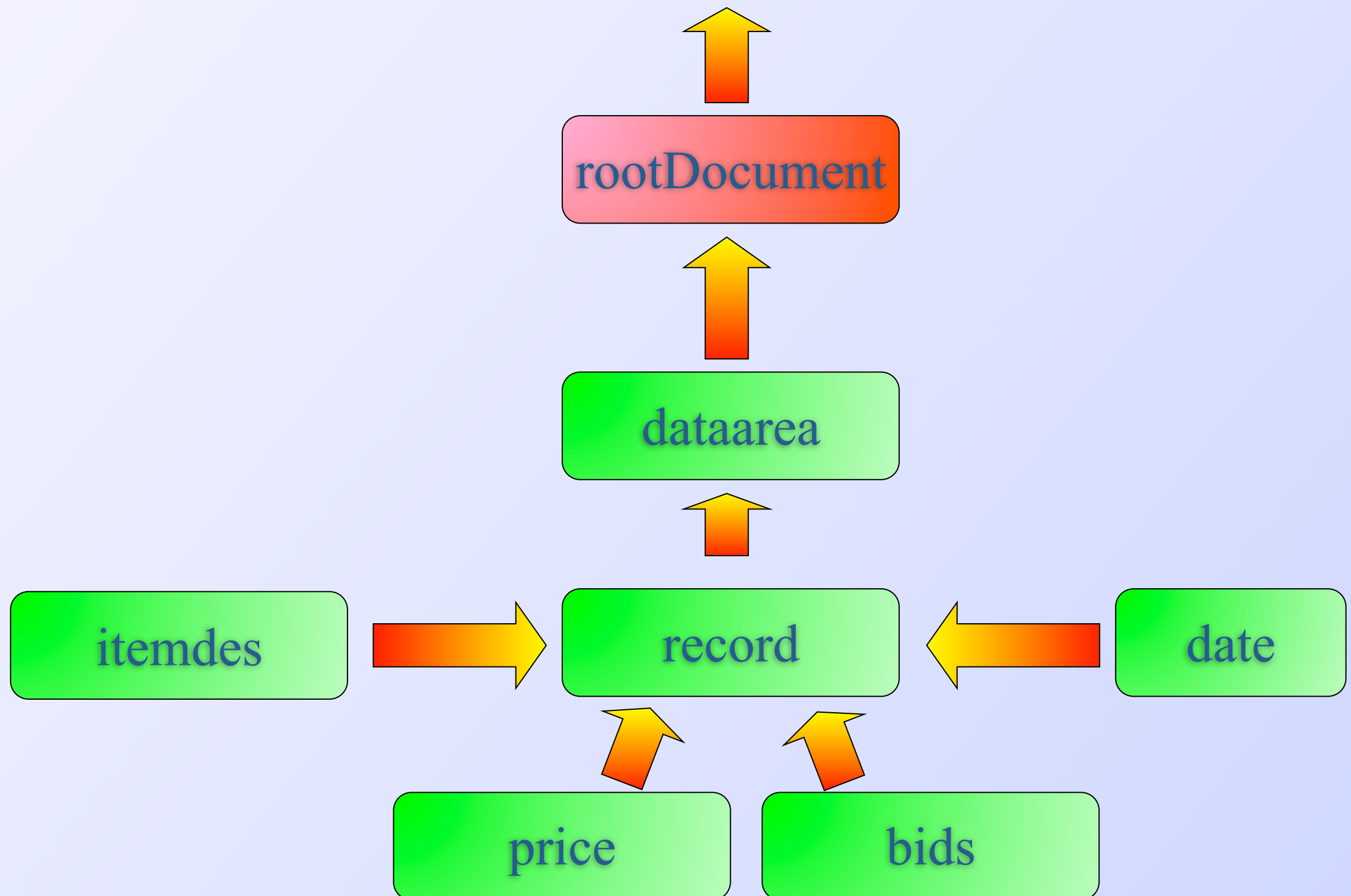


# Lixto Visual Developer

- **Easy** to use
  - Visual, interactive definition of Web extraction pattern (Eclipse-like)
  - Recording of user actions in the browser
    - e.g. mouse and key events (replay simulates human browsing behavior)
- Yet highly **expressive**
  - Support for **complex navigation patterns**
    - including “detail” pages and “next” pages
  - Expressive declarative **logic-based extraction language Elog**
  - **Robustness** to structural changes of the Web page

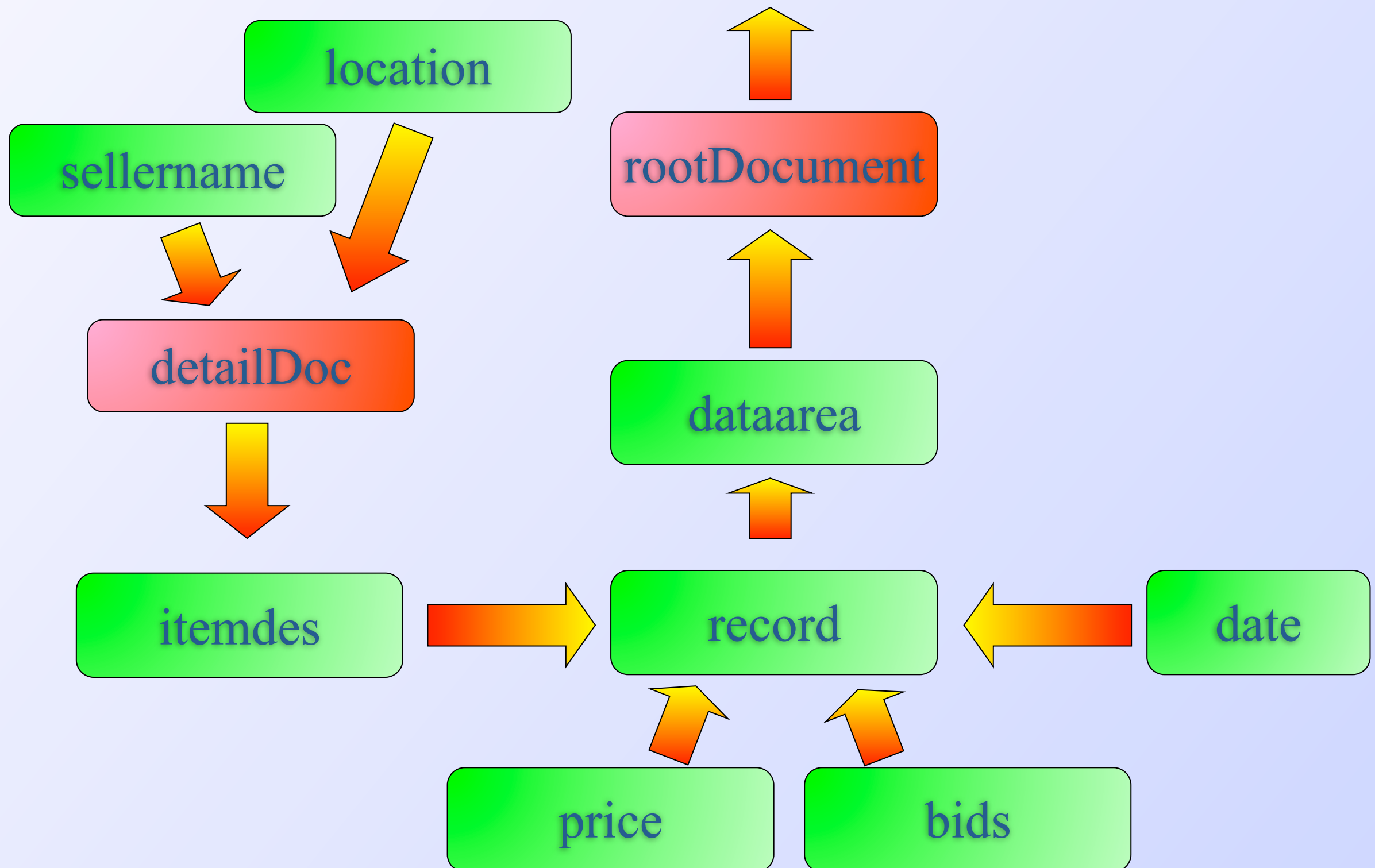


# Data Extractor: Pattern Graph and Recursion



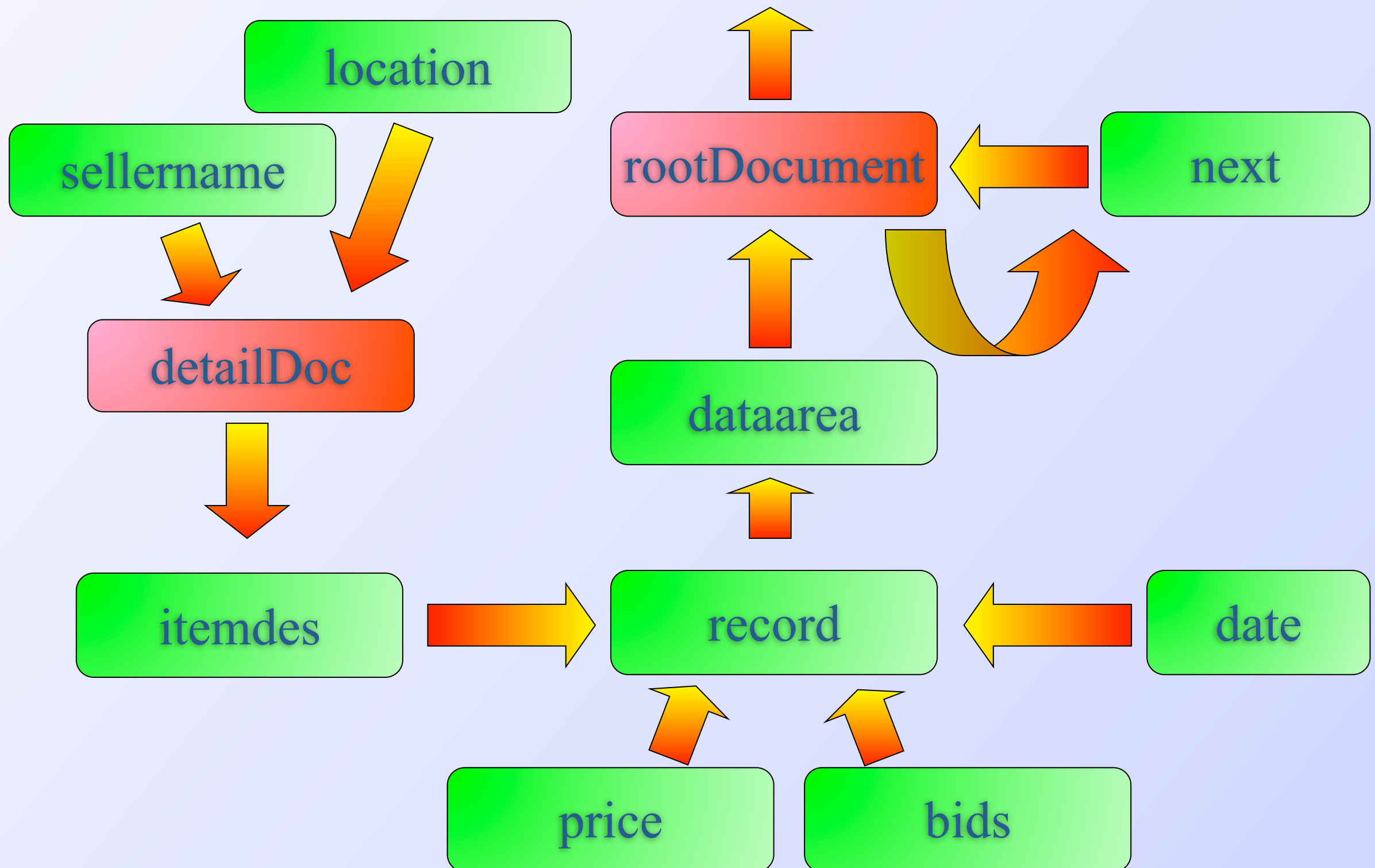


# Data Extractor: Pattern Graph and Recursion



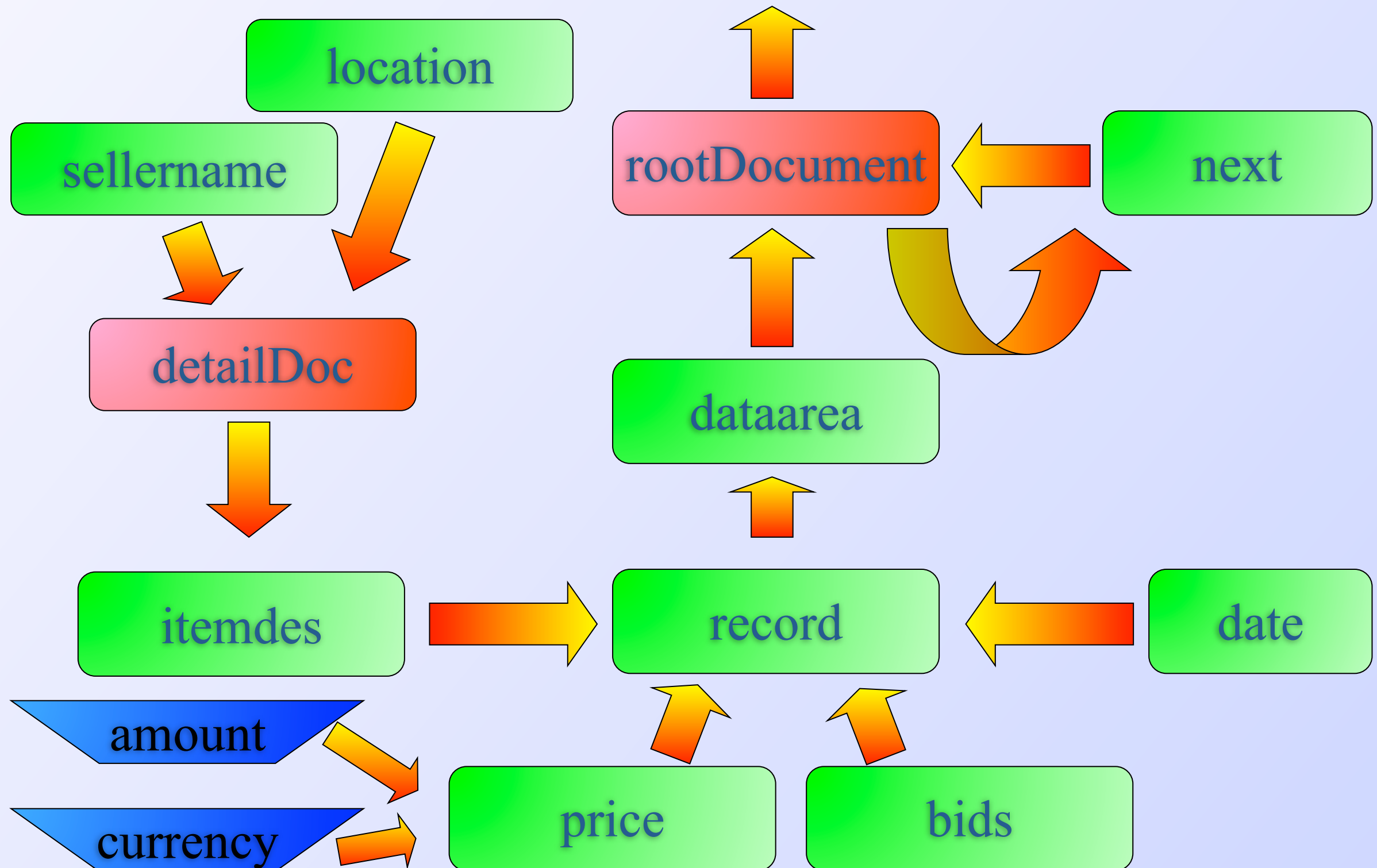


# Data Extractor: Pattern Graph and Recursion



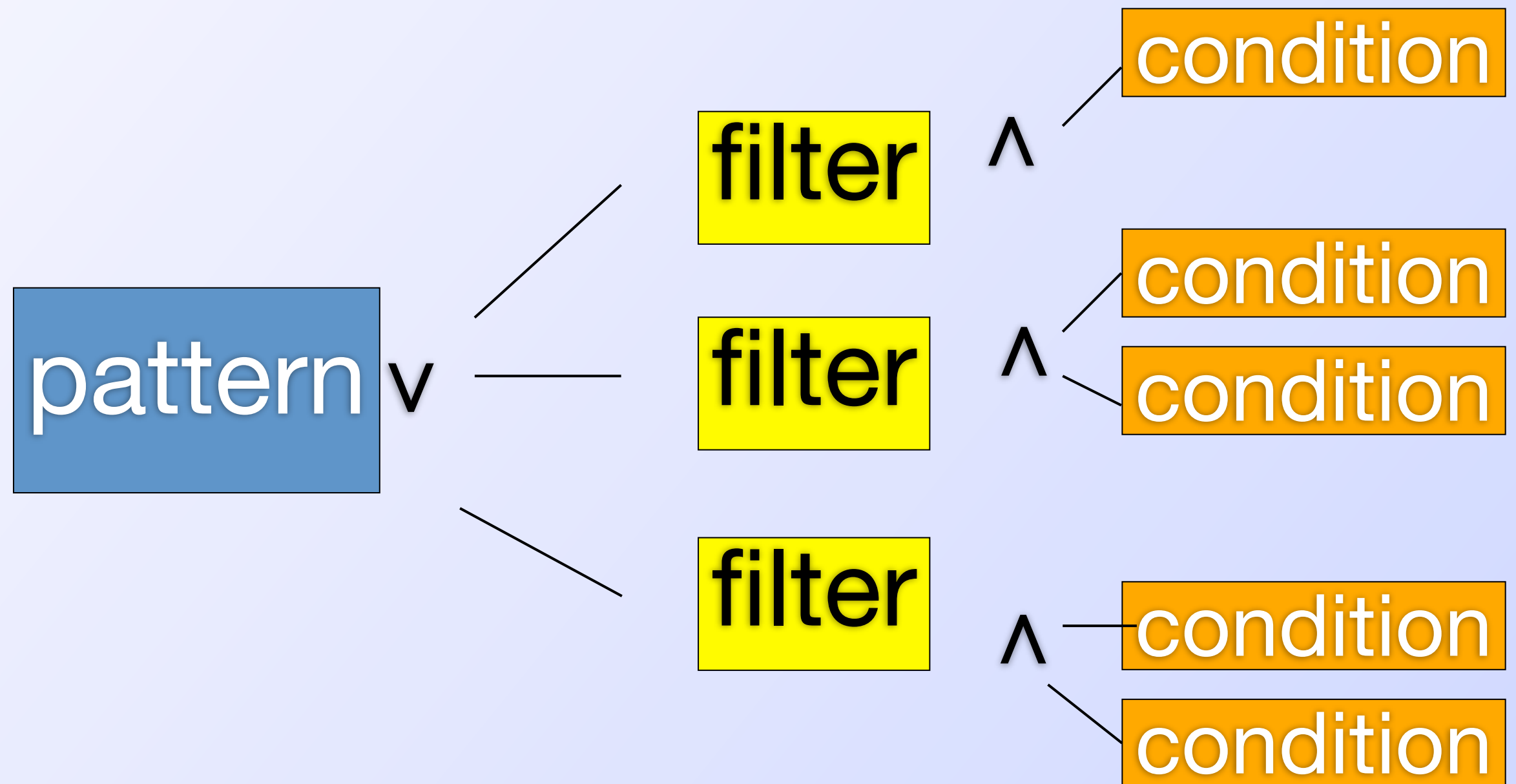


# Data Extractor: Pattern Graph and Recursion





# Pattern Structure



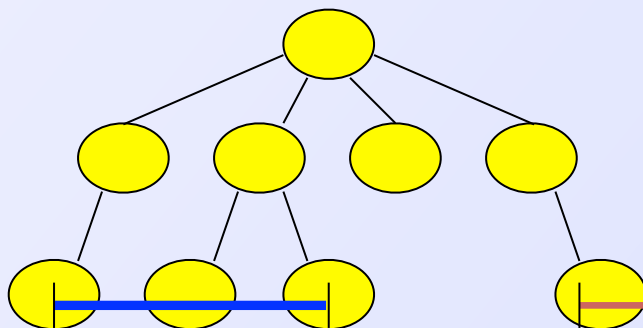
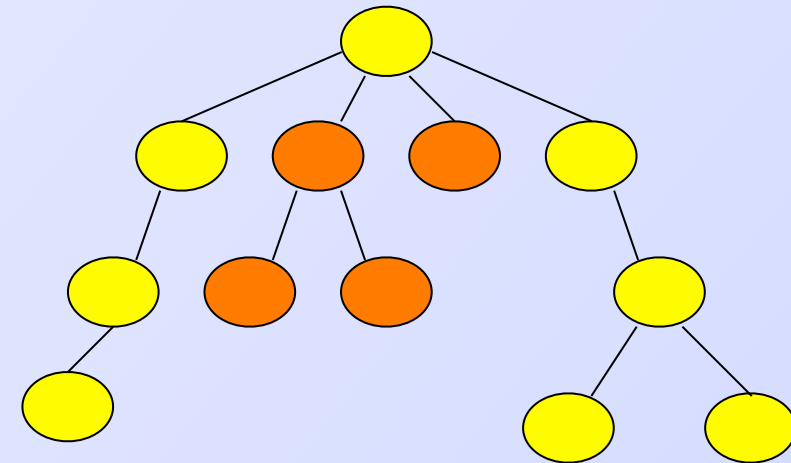
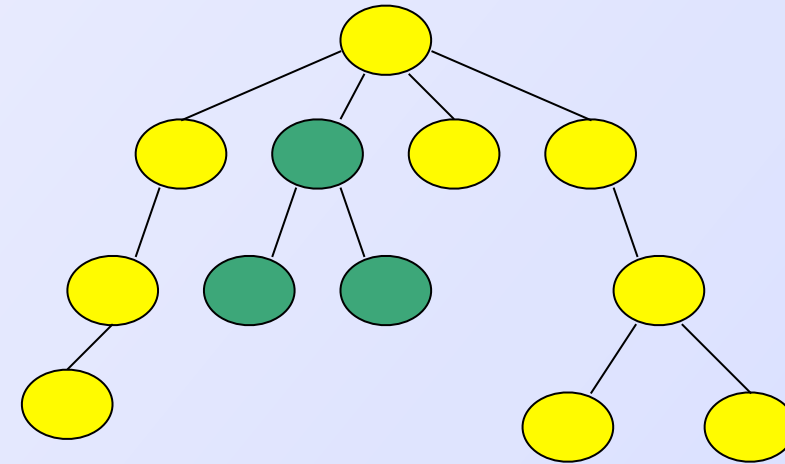
## Regional Algebra (Nodes and Texts)

Basic conditions before, after, contains, n-th child



# Object Model and Objects

- Labeled unranked tree
  - Tree Nodes
  - Complex Tree Regions
- String Representation
  - Strings
  - Attribute Values
  - Output Transformation
- CSS Box Model with spatial relations



Phosphorus (yellow or white)	7723-04-0	Only if it is a yellow or white form
Sulfuric acid (acid solutions including fumes, vapors, gas, fog, and other airborne forms of any particulate acid)	7664-93-9	Only if it is an aerosol for use as directed
Vanadium (except when contained in an alloy)	7440-62-2	Except if it is contained in an alloy
Zinc (dust or fume)	7440-66-6	Only if it is a dust or fume

The qualifiers for the following three chemicals is based on the chemical activity rather than the form of the chemical. These chemicals are subject to EPCRA section 315 reporting requirements only when the indicated activity is performed.

Chemical / Chemical Category	EPA Number	Qualifier
Diisocyanate and diisocyanate-like compounds (isocyanating, and the processing or otherwise use of diisocyanate and diisocyanate-like compounds if the diisocyanate and diisocyanate-like compounds are present as constituents in a chemical and if they were created during the manufacture of that chemical.)	NA	Only if they are manufactured at the facility, or are processed or otherwise used when present as constituents in a chemical but only if they were created during the manufacture of that chemical.
Isopropyl alcohol (only persons who manufacture by the strong acid process are subject; no supplier notification)	67-63-0	Only if it is being manufactured by the strong acid process. Facilities that produce or otherwise use isopropyl alcohol are <u>not</u> required.
Isocyanate (only persons who manufacture are subject; no supplier notification)	81-87-2	Only if it is being manufactured.

There are no supplier notification requirements for isopropyl alcohol and isocyanate since the processors and users of these chemicals are not required to report. Manufacturers of these chemicals do not need to notify their customers that these are reportable EPCRA



# Internal Knowledge Representation

- **Elog:** A Datalog-like language
  - **declarative and/or** semantics of Datalog matches
    - **narrowing and broadening** steps in the pattern generation
  - **maintenance** in Elog wrappers is **local** to the changed criteria
    - simple
  - Elog program itself is **hidden** from the wrapper designer

Lixto	Elog
Wrapper	Elog Program
Pattern	IDB-Predicate
Filter	Rule
Condition	Atom of rule body
Parent pattern	Special body atom
Element path	Special body atom



# Elog Programs

## ■ Homogeneous Programs

- Filters that define **same** IDB predicate refer to same parent pattern
- thus: **pattern graph is a tree**

## ■ Heterogeneous Programs

- Filters that define **same** IDB predicate refer to **different** parent patterns
- thus: **pattern graph may be cyclic**
  - but remains **locally stratified** if not mixing range and pattern references
  - fixed-point evaluation—stop if no new instance can be derived

```
document(S, X) ← getDocument($1, X)
  table(S, X) ← document(_, S), subelem(S, . * .table, X)
  table(S, X) ← table(_, S), subelem(S, . * .table, X)
```



# Example Elog Program Fragment

```
overview(X0,X1) :-  
    getDocumentFromNavigation(X0=$1,X1).
```

```
root(X0,X1) :-  
    overview(_,X0),  
    subelem(X0,(.,[]),X1).
```

```
hotel(X0,X1) :-  
    root(_,X0),  
    subelem(X0,(...lixto:nearest("tr"),[]),X1),  
    contains(X1,(./td[1]/a[1],[]),X2),  
    before(X0,X1,(...strong[1],[('text','Sortieren',CONTAINS)]),0,-1,X3,X4).
```

```
detail(X0,X1) :-  
    link(_,X0), getDocumentbyClick(X0,X1).
```

```
price(X0,X1) :-  
    hotel(_,X0),  
    subelem(X0,(...lixto:nearest("div")/lixto:nearest("strong"),[]),X1)  
    [1,1].
```

```
next(X0,X1) :-  
    root(_,X0),  
    subelem(X0,(...lixto:nearest("a"),[('text','Nächste Seite',CONTAINS)]),X1).
```

```
overview(X0,X1) :-  
    next(_,X0), getDocumentbyClick(X0,X1).
```

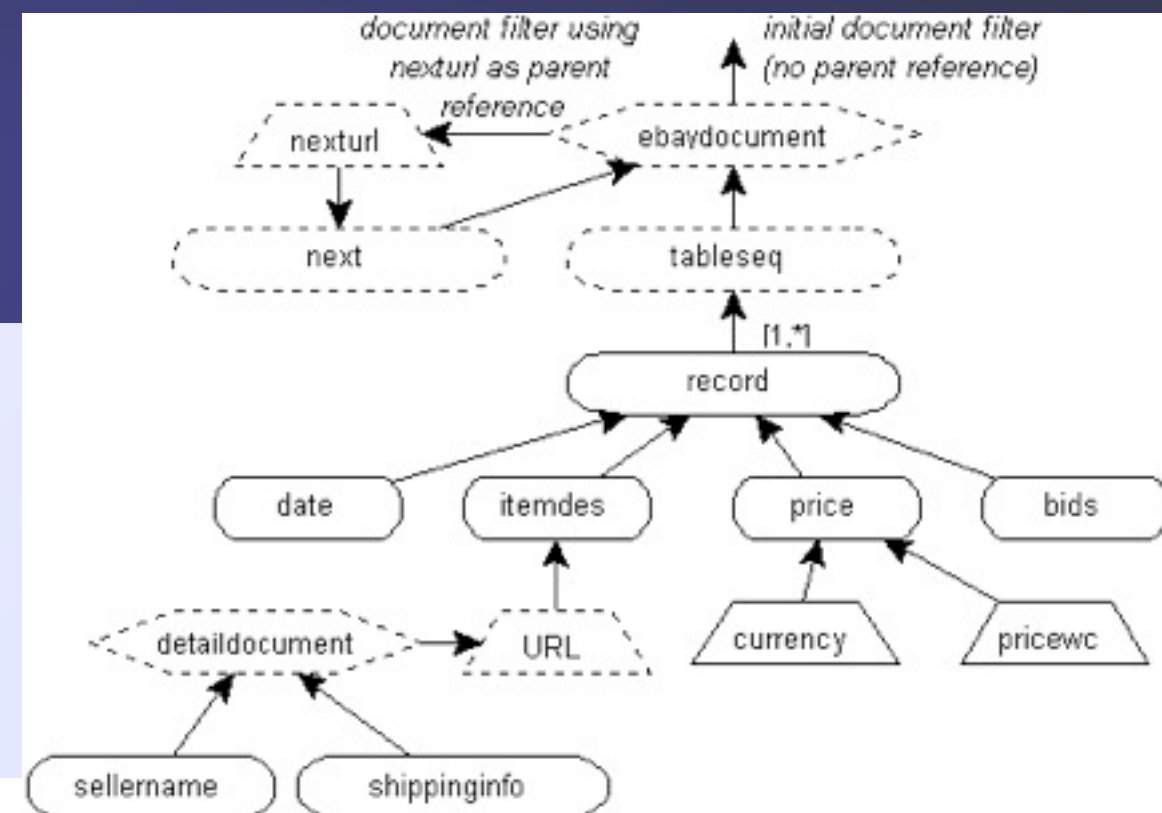


# Built-In Predicates

- Built-in predicates with binding schema
- Structural predicates
  - `subelem(S, path, X)`:
    - $X$  is a subelement connected to  $S$  via  $path$
  - inner structure: `contains`, `nthchild`
  - context: `before(S, X, path, d1, d2, Y, D)`, `after`, `notbefore`, `notafter`
- Textual and concept predicates
  - `subtext`, `subatt`
  - `isDate`, `isNumber`, `isCity`, `< >` on dates, distances, ...
- Pattern references
  - refer to any IDB predicate (e.g., “before there is an instance of price”) **107**



# Recursive Wrapping



```

next(S, X) ← ebaydocument(⌊, S),
             subelem(S, (⌋.content, [(href, ⌋, substr),
                                   (elementtext, (next page), exact)]), X)

```

```

nexturl(S, X) ← next(⌊, S), subatt(S, href, X)

```

```

ebaydocument(S, X) ← getDocument(S = $1, X)

```

```

ebaydocument(S, X) ← nexturl(⌊, S), getDocument(S, X)

```

Or, alternatively:

```

ebaydocument(S, X) ← ebaydocument(⌊, S),
                     subelem(S, (⌋.content, [(href, ⌋, substr),
                                           (elementtext, (next page), exact)]), Y),
                     subatt(Y, href, Z), getDocument(Z, X)

```



# Elog- Core Fragment

## Elog-:

- labeled tree structure via child and nextsibling relations
- omits features of Elog on textual structures, distances, etc.
- Signature:

$$t_{ur} = \langle \text{dom}, \text{root}, \text{leaf}, (\text{label}_a)_{a \in \Sigma}, \\ \text{firstchild}, \text{nextsibling}, \text{lastsibling} \rangle$$

Elog- has combined  
complexity  
 $O(|\text{dom}|^* |\text{program}|)$

Basic Elog relations are derived as

$$\begin{aligned} \text{subelem}_\epsilon(X, Y) &:= X = Y. \\ \text{subelem}_{\text{path}}(X, Y) &:= \text{child}(X, Z), \\ &\quad \text{subelem}_{\text{path}}(Z, Y). \\ \text{subelem}_{a.\text{path}}(X, Y) &:= \text{child}(X, Z), \text{label}_a(Z), \\ &\quad \text{subelem}_{\text{path}}(Z, Y). \end{aligned}$$

$$\begin{aligned} \text{contains}_{\text{path}}(X, Y) &:= \text{subelem}_{\text{path}}(X, Y). \\ \text{before}_{\text{path}}(S, X, Y) &:= \text{subelem}_{\text{path}}(S, Y), \\ &\quad \text{nextsibling}(X, Y). \\ \text{after}_{\text{path}}(S, X, Y) &:= \text{subelem}_{\text{path}}(S, Y), \\ &\quad \text{nextsibling}(Y, X). \\ \text{firstson}(S, X) &:= \text{firstchild}(S, X). \\ \text{lastson}(S, X) &:= \text{lastchild}(S, X). \end{aligned}$$



# Expressiveness of Elog

- Unary queries in **MSO** over trees
  - expressiveness **yardstick** for information extraction functions
  - but: lack of **operational** semantics
  - thus: **hard** to use as wrapping language

## Theorems [Gottlob & Koch, PODS 2002]:

Over  $\tau_{ur}$ , Monadic Datalog = MSO

A unary query is definable in MSO iff it is definable via a monadic datalog program.

ELOG- expresses monadic datalog + child

**Lixto** expresses all MSO wrapping tasks.



# Part II:

# Unsupervised, domain-specific



	Knowledge			
	– (site-specific)	Schema	Instances	Patterns
Visual	<div>SXPath</div> <div>XPath + Visual Relations</div>	<div>SCRAP</div> <div>Robust wrappers by schema</div>	<div>WRAPPO</div> <div>Visual patterns and annotations</div>	
	<div>OXPath</div> <div>XPath + Actions</div>	<div>Lixto</div> <div>Visual designer + expressive extraction language Elog</div>	<div>DIADEM</div> <div>Form understanding, object identification, and exploration driven by domain knowledge</div>	



# Part II: Introduction

# Ontologies in DE



# Why Unsupervised Data Extraction?



- ❖ **Too many** fish in the pond
  - ❖ > **17 000** real estate UK sites
    - ❖ similar for restaurants, travel, car dealers, airlines, pharmacies, retail shops, ...
  - ❖ aggregators cover only a fraction
    - ❖ updated slowly
- ⇒ **per site** manual work **infeasible**
  - ❖ wrapper construction too expensive
    - ❖ tracking changes
  - ❖ excludes manual & (semi-) supervised



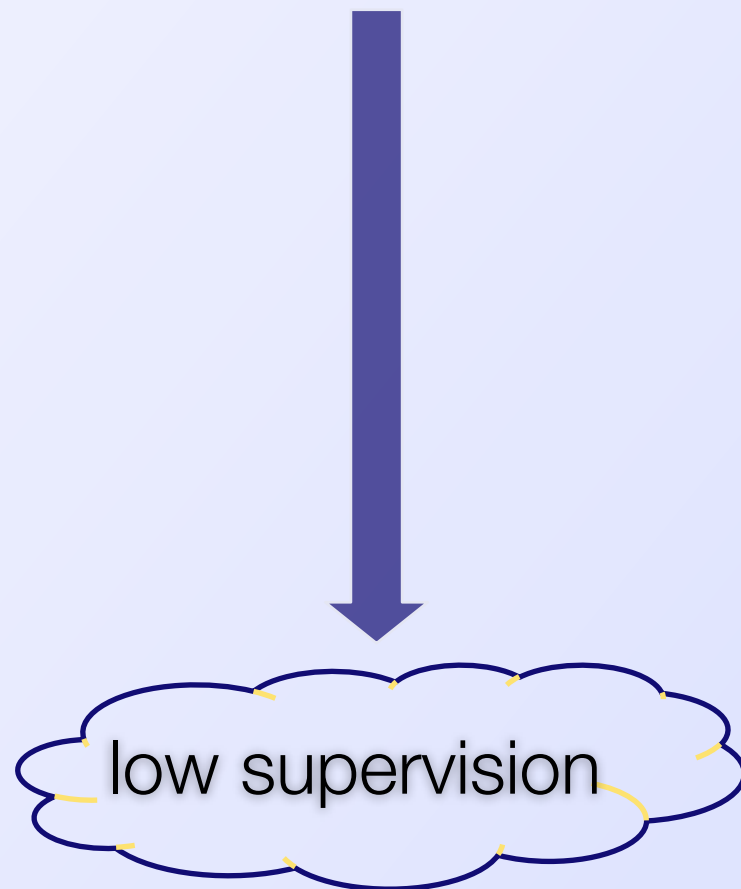
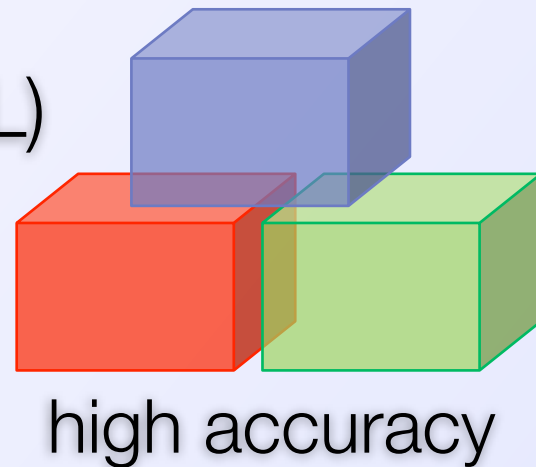
# Unsupervised Web Data Extraction

- Key observation:
  - “database” web sites are generated using **templates**
  - wrapper generators need to automatically identifying templates
- Two major approaches
  - **machine learning** from a few hand-labeled examples
    - high precision only for **simple domains** (single entity type, few attributes)
    - or: per-site/template training required
  - fully automatically exploit the **repeated structure of result pages**
    - good precision needs a lot of data per site (many records per page, many pages)
    - only works for sites with **sites with simple object types**

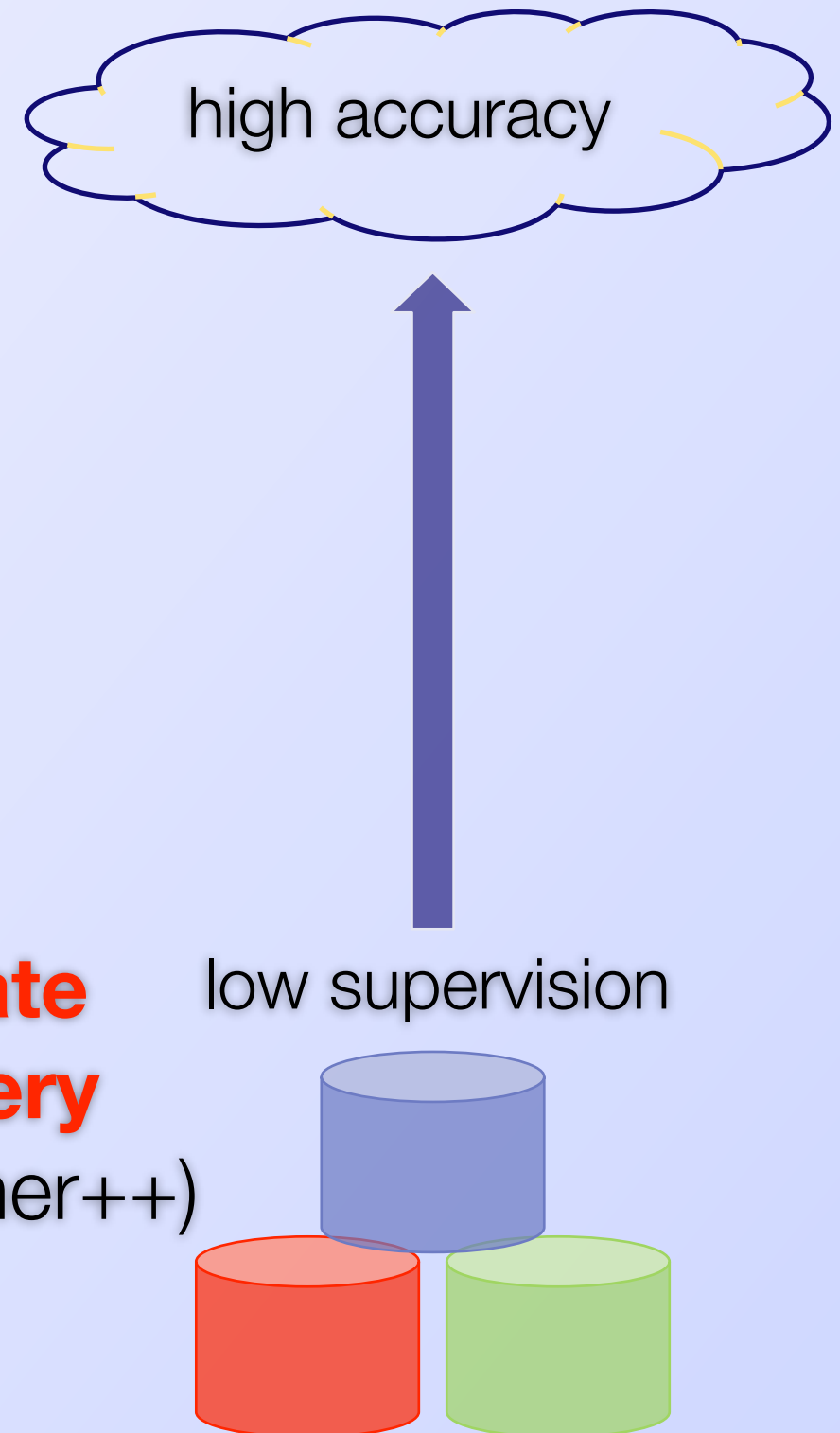


# Bad News: Nobody Can do it Yet

**Wrapper  
Induction (ML)**



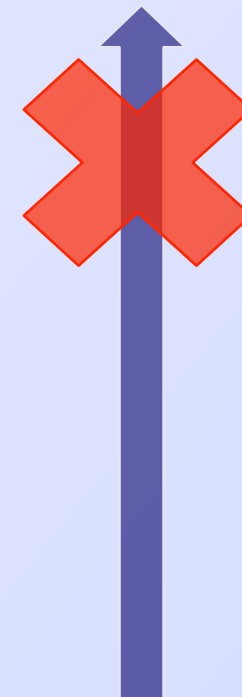
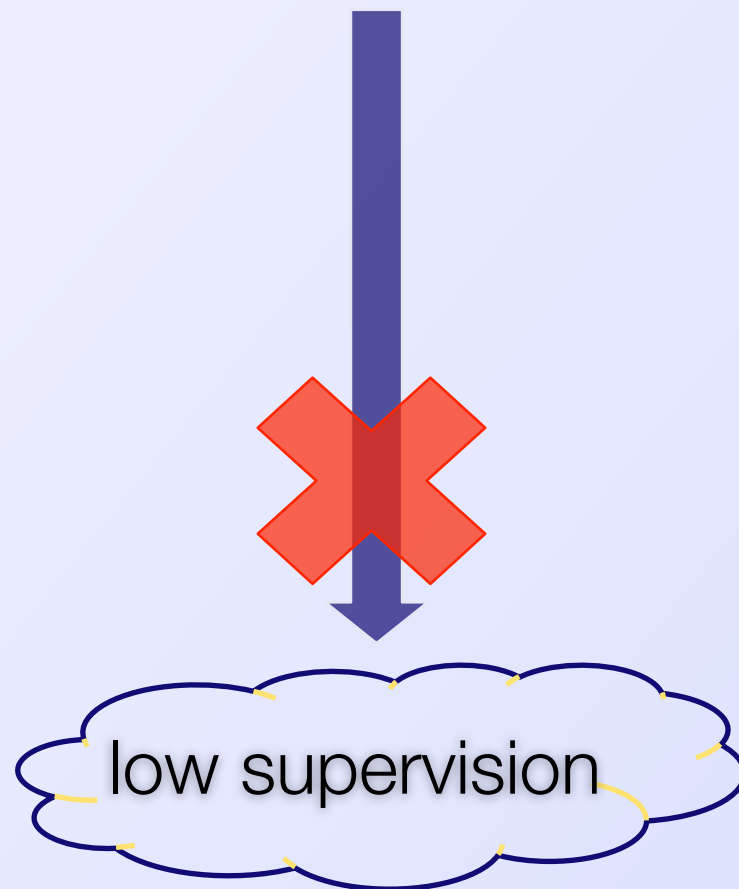
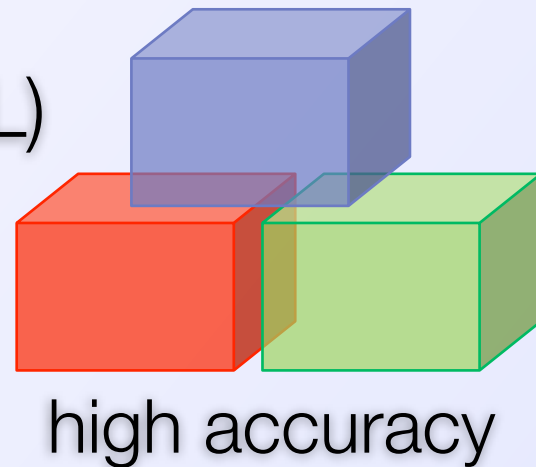
**Template  
Discovery**  
(RoadRunner++)



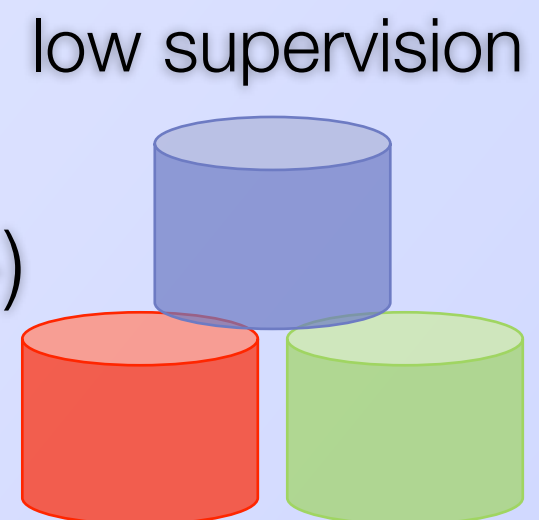


# Bad News: Nobody Can do it Yet

**Wrapper  
Induction (ML)**



**Template  
Discovery**  
(RoadRunner++)





# Knowledge in Data Extraction



# Knowledge in Data Extraction

- what's “**knowledge**” here
  - **observational:** what to observe, annotations
    - that a certain text is highlighted, that a certain keyword appears in it
  - **phenomenological:** how observations become concepts
    - that a text “...:” to the close north-west of a field is that field's label
  - **ontological:** schema, concepts & constraints
    - e.g., “bathroom”, “every property must have a location”
- orthogonal: **script** knowledge for web pages
- both domain-independent and domain-dependent
  - but currently a trend towards domain-dependent knowledge



# Knowledge in Data Extraction

- what's “**knowledge**” here
  - **observational:** what to observe, annotations **phenomenon**
    - that a certain text is highlighted, that a certain keyword appears in it
  - **phenomenological:** how observations become concepts
    - that a text “...:” to the close north-west of a field is that field's label
  - **ontological:** schema, concepts & constraints
    - e.g., “bathroom”, “every property must have a location”
- orthogonal: **script** knowledge for web pages
- both domain-independent and domain-dependent
  - but currently a trend towards domain-dependent knowledge



# Knowledge in Data Extraction

- what's “**knowledge**” here
  - **observational:** what to observe, annotations **phenomenon**
    - that a certain text is highlighted, that a certain keyword appears in it
  - **phenomenological:** how observations become concepts
    - that a text “...:” to the close north-west of a field is that field's label
  - **ontological:** schema, concepts & constraints
    - e.g., “bathroom”, “every property must have a location” **idea/noumenon**
- orthogonal: **script** knowledge for web pages
- both domain-independent and domain-dependent
  - but currently a trend towards domain-dependent knowledge



# Knowledge in Data Extraction

## what's “**knowledge**” here

- **observational:** what to observe, annotations
    - that a certain text is highlighted, that a certain keyword appears in it
  - **phenomenological:** how observations become concepts
    - that a text “...:” to the close north-west of a field is that field's label
  - **ontological:** schema, concepts & constraints
    - e.g., “bathroom”, “every property must have a location”
- phenomenon
- mapping
- idea/**noumenon**
- orthogonal: **script** knowledge for web pages
  - both domain-independent and domain-dependent
    - but currently a trend towards domain-dependent knowledge



# Trend: Towards Domain-Knowledge

## ■ **Observational** only:

- Su, Wang, Lochovsky. **ODE**, TODS 2009

## ■ **Ontological** only:

- SCRAP (see before)

## ■ **Observational & ontological**:

- Dalvi, Kumar, Soliman. **Automatic Wrappers for Large Scale Web Extraction**, VLDB 2011. (AutoWrapper in the following)
- Venetis, Halevy, Madhavan, et al. **Recovering Semantics of Tables on the Web**, VDLB 2011. (SemanticTables in the following)



# Trend: Towards Domain-Knowledge

## ■ **Observational** only:

- Su, Wang, Lochovsky. **ODE**, TODS 2009

## ■ **Ontological** only:

- SCRAP (see before)

## ■ **Observational & ontological:**

**shallow** ontology, tailored  
to **single** attribute extraction

- Dalvi, Kumar, Soliman. **Automatic Wrappers for Large Scale Web Extraction**, VLDB 2011. (AutoWrapper in the following)
- Venetis, Halevy, Madhavan, et al. **Recovering Semantics of Tables on the Web**, VDLB 2011. (SemanticTables in the following)



# Trend: Towards Domain-Knowledge

## ■ **Observational** only:

- Su, Wang, Lochovsky. **ODE**, TODS 2009

## ■ **Ontological** only:

- SCRAP (see before)

## ■ **Observational & ontological:**

**shallow** ontology, tailored  
to **single** attribute extraction

- Dalvi, Kumar, Soliman. **Automatic Wrappers for Large Scale Web Extraction**, VLDB 2011. (AutoWrapper in the following)
- Venetis, Halevy, Madhavan, et al. **Recovering Semantics of Tables on the Web**, VDLB 2011. (SemanticTables in the following)



	Knowledge			
	– (site-specific)	Schema	Instances	Patterns
Visual	<div>SXPath</div> <div>XPath + Visual Relations</div>	<div>SCRAP</div> <div>Robust wrappers by schema</div>	<div>WRAPPO</div> <div>Visual patterns and annotations</div>	
	<div>OXPath</div> <div>XPath + Actions</div>	<div>Lixto</div> <div>Visual designer + expressive extraction language Elog</div>	<div>DIADEM</div> <div>Form understanding, object identification, and exploration driven by domain knowledge</div>	



# WRAPPO

Wrapping Objects from the Web by  
Exploiting  
Domain Knowledge and Spatial  
Information



# WRAPPO

Wrapping Objects from the Web by  
Exploiting  
Domain Knowledge and Spatial

S. Flesca, E. Oro, M. Ruffolo: **WRAPPO—  
Wrapping objects from the web by exploiting  
domain knowledge and spatial information.**  
*Technical Report ICAR-CNR (2012).*



# Motivations

- **Problem:** existing DE mostly based on simple DOM
  - yet web pages use presentation to convey semantics to users
- **Observation:**
  - recent approaches that use spatial arrangement of pages
  - recent approaches that use the semantics of the page content
- **WRAPPO:** higher quality, more robust wrappers by combining
  - semantic annotations and
  - spatial arrangement / visual information



# Presentation-Oriented Documents I

## DOM Portion

Digital Cameras – Compare Prices on Digital Cameras at PriceGrabber

http://cameras.pricegrabber.com/digital/p/48/

pricegrabber

Shop by Department

Digital Cameras Cameras Digital Camera Batteries Reviews & Ratings digital cameras digital camera Buying Guides Local Deals

Digital Cameras 2,493 Matches

Home Cameras Digital Cameras Cameras

Featured Brand

See Products from:

SONY

Grid View List View

ZIP Code: Sort By:

Price Range

☐ Under \$109 (629)

☐ \$109 - \$199 (610)

☐ \$199 - \$353 (642)

☐ \$353+ (641)

Condition

☐ New (~2345)

☐ Used (~56)

☐ Refurbished (~94)

Shipping

☐ Free Shipping (~1104)

Power Score

☐ 7 to 10 (1677)

☐ 5 to 7 (271)

☐ 3 to 5 (33)

☐ Lower than 3 (4)

Top Manufacturers

☐ Canon (266)

☐ Nikon (242)

☐ Sony (334)

☐ Panasonic (236)

☐ Fujifilm (136)

Featured Product

**Sony Cyber-shot DSC-HX100V Black Digital Camera**

POWER SCORE 0.2

DSC-HX100V - 16.2 Megapixel - 4.80 mm-144 mm - 3" LCD - 30x Optical Zoom - Optical IS - 4608 x...  
Renew your passion for photography with the Cyber-shot DSC-HX100V digital camera. A Carl Zeiss... [More](#)

[Add to Favorites](#) [Add to Compare](#)

As low as **\$428.95**

[COMPARE NOW](#)

Price Range **\$428.95 - 449.99**

Sold By 10 Merchants

**Canon PowerShot SX150 IS 14.1 Megapixel Compact Camera - Black**

56648001 - 3" LCD - 12x Optical Zoom - Optical IS - 4320 x 3240 Image - 1280 x 720 Video -...  
Discover how simple and easy it is to be creative and shoot memorable photos and videos of your... [More](#)

[Add to Favorites](#) [Add to Compare](#)

As low as **\$228.88**

[COMPARE NOW](#)

Price Range **\$228.88 - 274.99**

Sold By 15 Merchants

**Nikon COOLPIX S8200 Black Digital Camera**

26286 - 16.1 MP, 14x Opt, 9DiSDHC/SDXC Card Slot  
Nikon's COOLPIX S8200 Digital Camera puts ease-of-use and a remarkable 16.1Mp CMOS sensor at... [More](#)

[Add to Favorites](#) [Add to Compare](#)

As low as **\$326.95**

[COMPARE NOW](#)

Price Range **\$326.95 - 329.99**

Sold By 11 Merchants

Welcome [Sign in or Register](#)

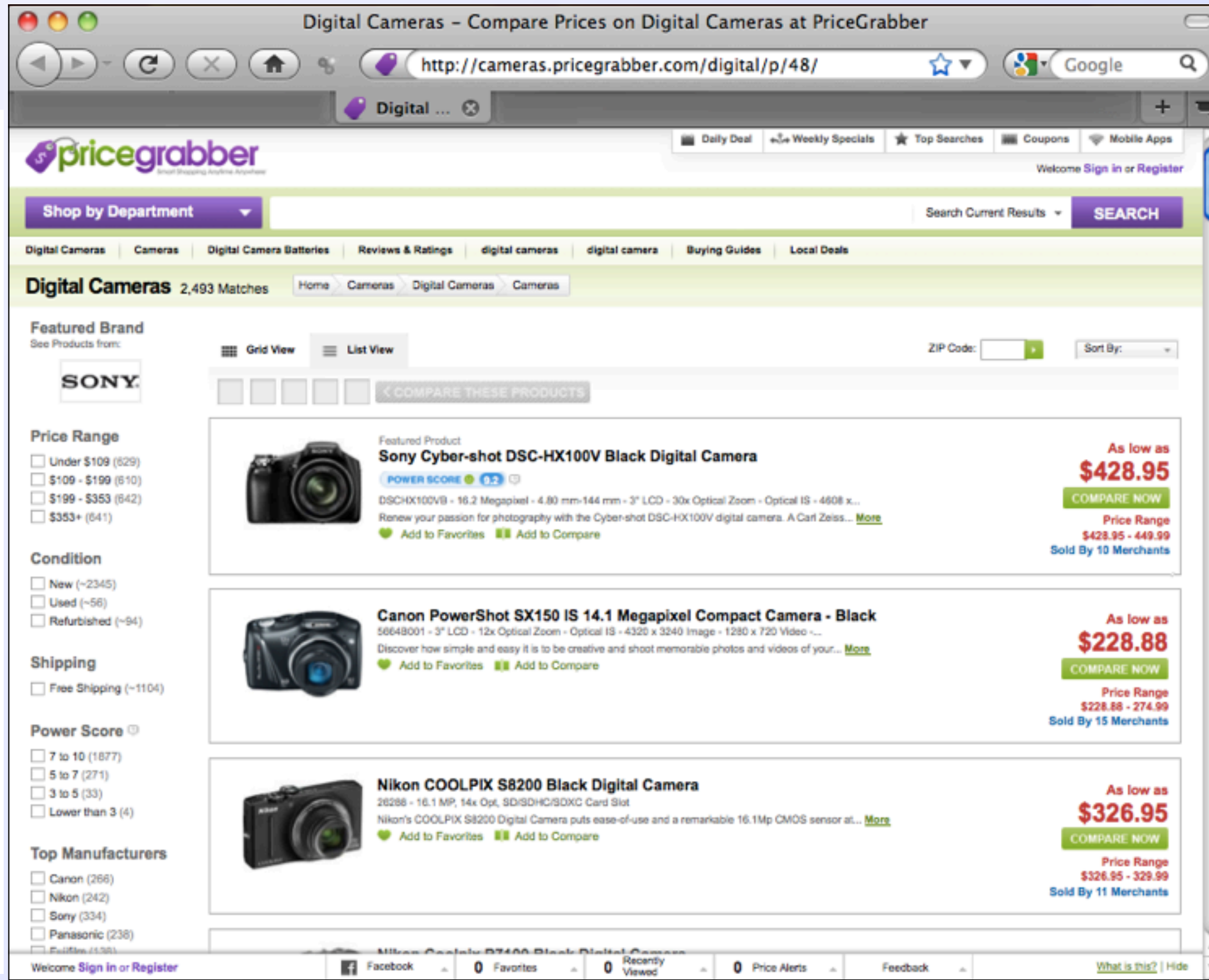
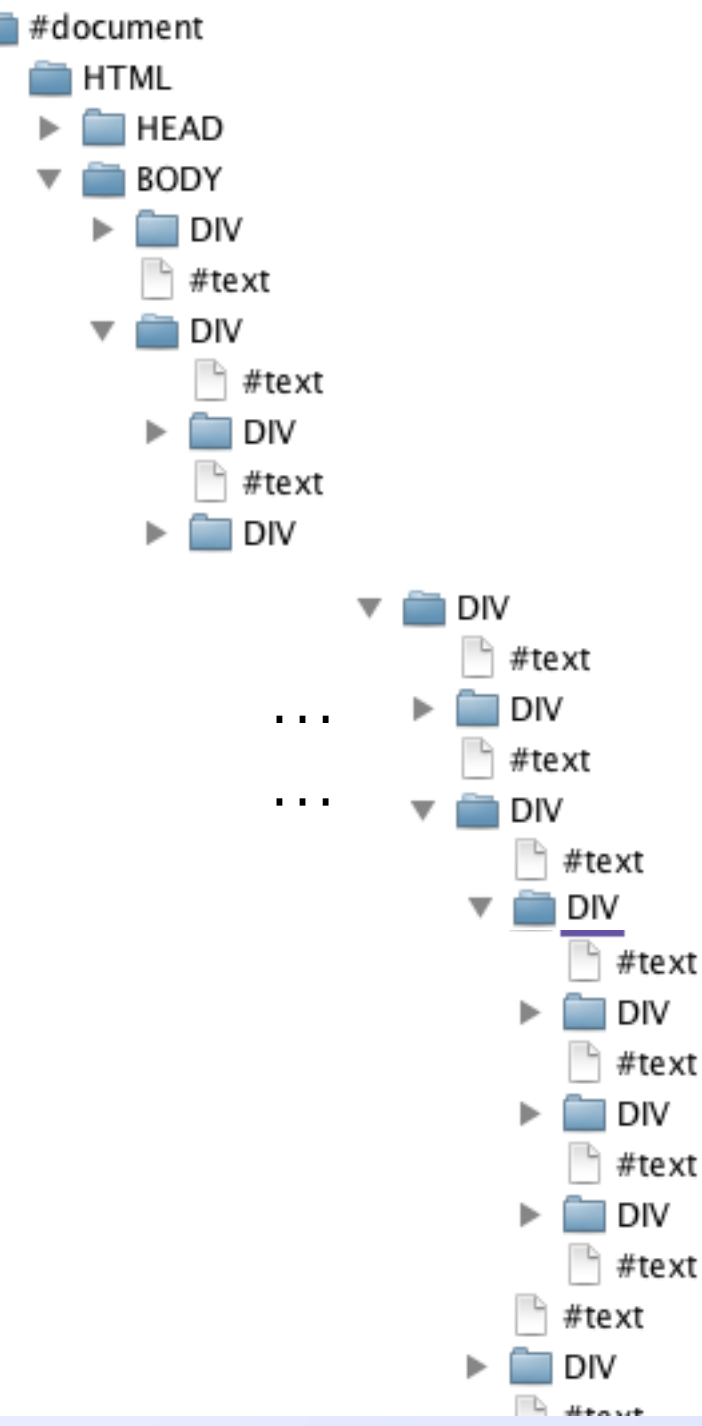
Facebook 0 Favorites 0 Recently Viewed 0 Price Alerts Feedback

[What is this?](#) [Hide](#)



# Presentation-Oriented Documents I

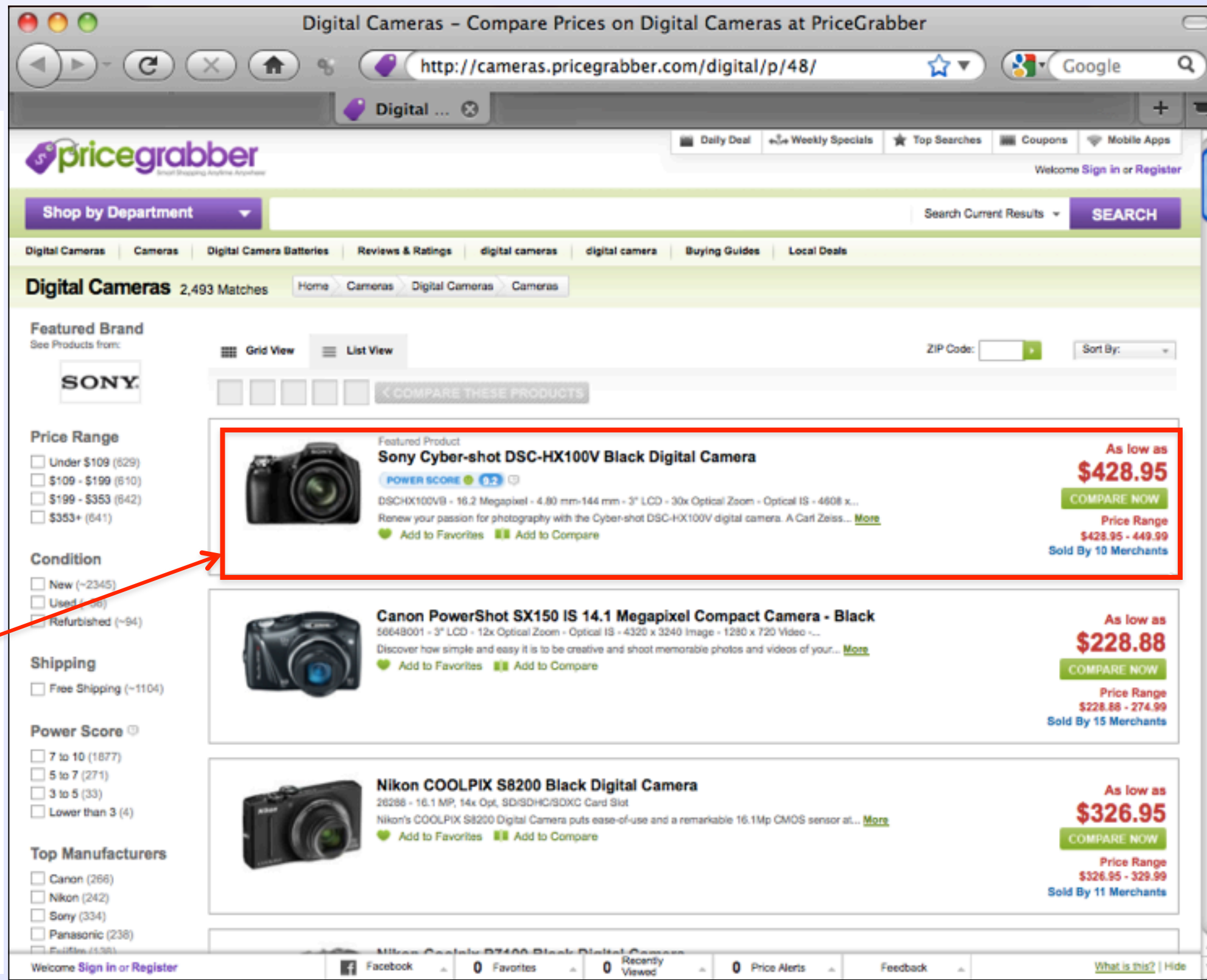
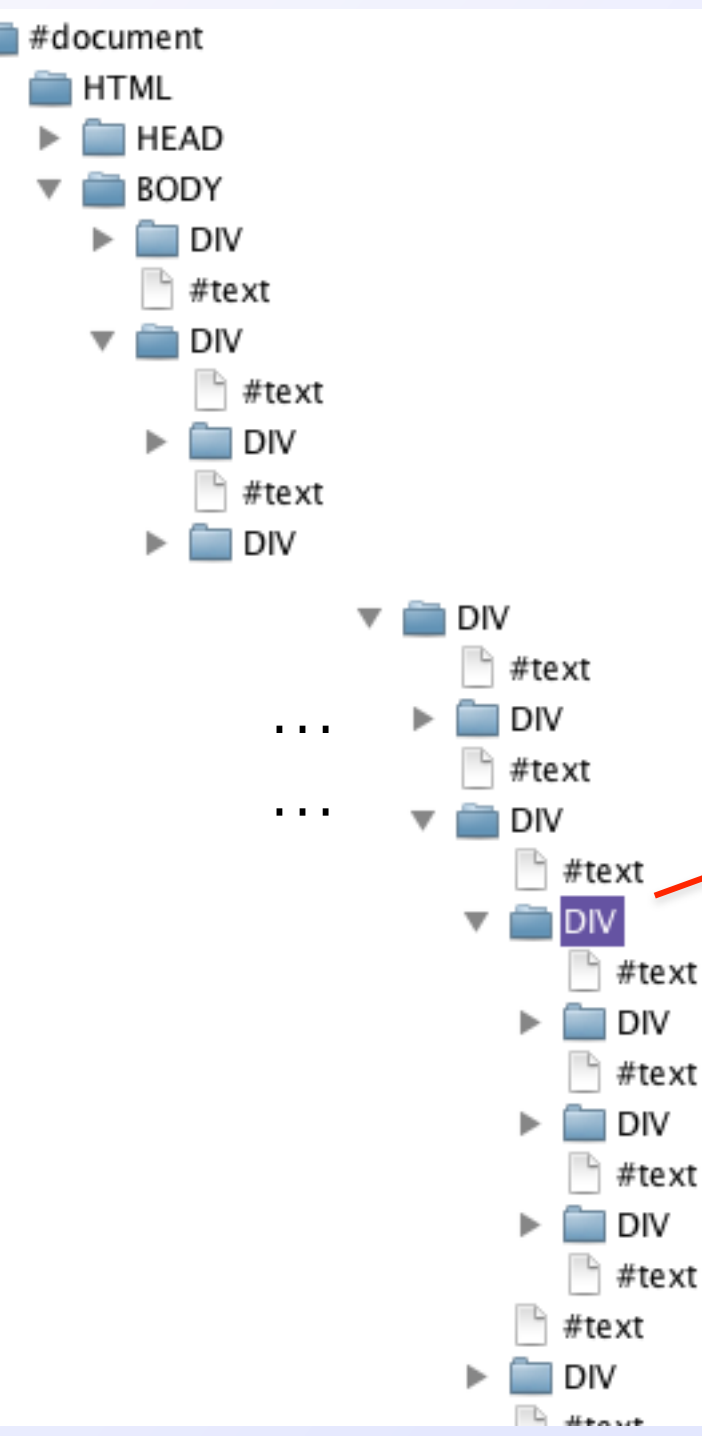
# DOM Portion





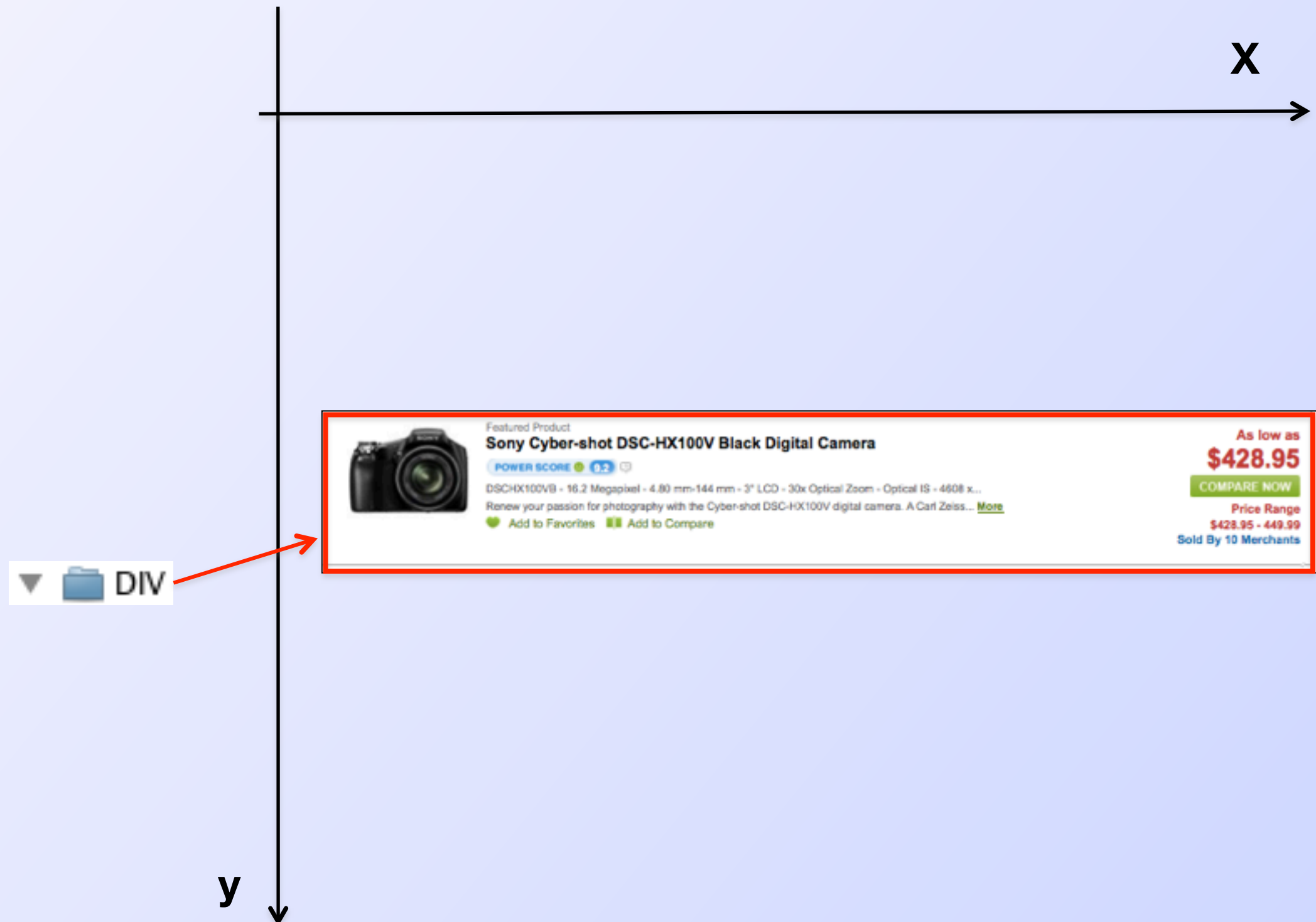
# Presentation-Oriented Documents I

# DOM Portion



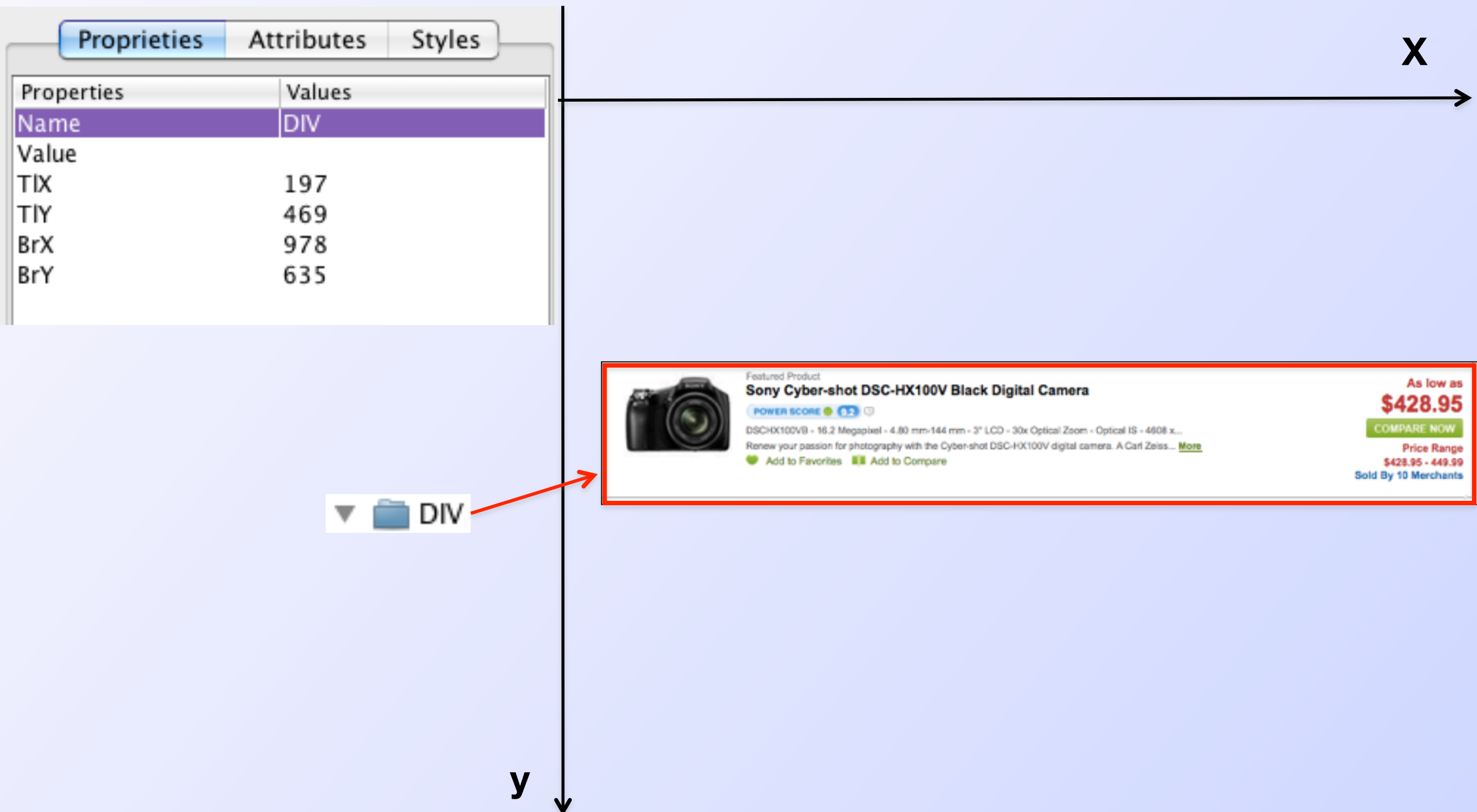


# Presentation-Oriented Documents II





# Presentation-Oriented Documents II





# Presentation-Oriented Documents II

Diagram illustrating the extraction of data from a presentation-oriented document (e.g., a product listing) using a coordinate system (X, Y).

The diagram shows a coordinate system with X and Y axes. A red dot marks the starting point at (197, 469). A red arrow points from a folder icon labeled "DIV" to the product listing area, which is bounded by a red rectangle. The bottom-right corner of this rectangle is marked with a red dot at (978, 635).

**Properties Panel:**

Properties	Values
Name	DIV
Value	
TIX	197
TIY	469
BrX	978
BrY	635

**Product Listing (Sony Cyber-shot DSC-HX100V Black Digital Camera):**

- Featured Product
- Sony Cyber-shot DSC-HX100V Black Digital Camera
- POWER SCORE 3.2
- DSC-HX100VB - 16.2 Megapixel - 4.80 mm-144 mm - 3" LCD - 30x Optical Zoom - Optical IS - 4608 x...
- Renew your passion for photography with the Cyber-shot DSC-HX100V digital camera. A Carl Zeiss... [More](#)
- [Add to Favorites](#) [Add to Compare](#)
- As low as **\$428.95**
- [COMPARE NOW](#)
- Price Range **\$428.95 - 449.99**
- Sold By 10 Merchants






# Presentation-Oriented Documents II

Proprieties	Attributes	Styles
Properties	Values	
background-attach...	scroll	
background-color	transparent	
background-image	none	
background-position	0% 0%	
background-repeat	repeat	
border-bottom-color	rgb(82, 140, 0)	
border-bottom-style	solid	
border-bottom-width	3px	
border-collapse	separate	
border-left-color	rgb(82, 140, 0)	

(197,  
469)

▼  DIV

Featured Product  
**Sony Cyber-shot DSC-HX100V Black Digital Camera**  
POWER SCORE  3.2  
DSCHX100VB - 16.2 Megapixel - 4.80 mm-144 mm - 3" LCD - 30x Optical Zoom - Optical IS - 4608 x...  
Renew your passion for photography with the Cyber-shot DSC-HX100V digital camera. A Carl Zeiss... [More](#)  
 Add to Favorites  Add to Compare  
As low as **\$428.95**  
[COMPARE NOW](#)  
Price Range  
\$428.95 - 449.99  
Sold By 10 Merchants

(978,  
635)



# Presentation-Oriented Documents II

Digital Cameras - Compare Prices on Digital Cameras at PriceGrabber

http://cameras.pricegrabber.com/digital/p/48/

pricegrabber

Shop by Department

Digital Cameras 2,493 Matches

Featured Brand: SONY

Price Range:

- ☐ Under \$109 (629)
- ☐ \$109 - \$199 (610)
- ☐ \$199 - \$353 (642)
- ☐ \$353+ (641)

Condition:

- ☐ New (~2345)
- ☐ Used (~56)
- ☐ Refurbished (~94)

Shipping:

- ☐ Free Shipping (~1104)

Power Score:

- ☐ 7 to 10 (1677)
- ☐ 5 to 7 (271)
- ☐ 3 to 5 (33)
- ☐ Lower than 3 (4)

Top Manufacturers:

- ☐ Canon (266)
- ☐ Nikon (242)
- ☐ Sony (334)
- ☐ Panasonic (236)
- ☐ Fujifilm (138)

Grid View List View

COMPARE THESE PRODUCTS

Featured Product: Sony Cyber-shot DSC-HX100V Black Digital Camera

As low as \$428.95

POWER SCORE 8.2

DSC-HX100VB - 16.2 Megapixel - 4.80 mm-144 mm - 3" LCD - 30x Optical Zoom - Optical IS - 4608 x...  
Renew your passion for photography with the Cyber-shot DSC-HX100V digital camera. A Carl Zeiss... [More](#)

[Add to Favorites](#) [Add to Compare](#)

Price Range \$428.95 - 449.99  
Sold By 10 Merchants

COMPARE NOW

Canon PowerShot SX150 IS 14.1 Megapixel Compact Camera - Black

As low as \$228.88

56648001 - 3" LCD - 12x Optical Zoom - Optical IS - 4320 x 3240 Image - 1280 x 720 Video -...  
Discover how simple and easy it is to be creative and shoot memorable photos and videos of your... [More](#)

[Add to Favorites](#) [Add to Compare](#)

Price Range \$228.88 - 274.99  
Sold By 15 Merchants

COMPARE NOW

Nikon COOLPIX S8200 Black Digital Camera

As low as \$326.95

26286 - 16.1 MP, 14x Opt, 3D/SDHC/SDXC Card Slot  
Nikon's COOLPIX S8200 Digital Camera puts ease-of-use and a remarkable 16.1Mp CMOS sensor at... [More](#)

[Add to Favorites](#) [Add to Compare](#)

Price Range \$326.95 - 329.99  
Sold By 11 Merchants

COMPARE NOW

Welcome Sign in or Register

Facebook 0 Favorites 0 Recently Viewed 0 Price Alerts Feedback

What is this? | Hide



# Presentation-Oriented Documents II

camera

Digital Cameras - Compare Prices on Digital Cameras at PriceGrabber

http://cameras.pricegrabber.com/digital/p/48/

pricegrabber

Shop by Department

Digital Cameras | Cameras | Digital Camera Batteries | Reviews & Ratings | digital cameras | digital camera | Buying Guides | Local Deals

Digital Cameras 2,493 Matches

Home | Cameras | Digital Cameras | Cameras

Featured Brand

See Products from:

SONY

Grid View | List View

ZIP Code: Sort By:

Price Range

- ☐ Under \$109 (629)
- ☐ \$109 - \$199 (610)
- ☐ \$199 - \$353 (642)
- ☐ \$353+ (641)

Condition

- ☐ New (~2345)
- ☐ Used (~56)
- ☐ Refurbished (~94)

Shipping

- ☐ Free Shipping (~1104)

Power Score

- ☐ 7 to 10 (1677)
- ☐ 5 to 7 (271)
- ☐ 3 to 5 (33)
- ☐ Lower than 3 (4)

Top Manufacturers

- ☐ Canon (266)
- ☐ Nikon (242)
- ☐ Sony (334)
- ☐ Panasonic (236)
- ☐ Fujifilm (138)

Featured Product

**Sony Cyber-shot DSC-HX100V Black Digital Camera**

POWER SCORE 0.2

DSC-HX100VB - 16.2 Megapixel - 4.80 mm-144 mm - 3" LCD - 30x Optical Zoom - Optical IS - 4608 x...  
Renew your passion for photography with the Cyber-shot DSC-HX100V digital camera. A Carl Zeiss... [More](#)

[Add to Favorites](#) [Add to Compare](#)

As low as **\$428.95**

[COMPARE NOW](#)

Price Range **\$428.95 - 449.99**

Sold By 10 Merchants

**Canon PowerShot SX150 IS 14.1 Megapixel Compact Camera - Black**

56648001 - 3" LCD - 12x Optical Zoom - Optical IS - 4320 x 3240 Image - 1280 x 720 Video -...  
Discover how simple and easy it is to be creative and shoot memorable photos and videos of your... [More](#)

[Add to Favorites](#) [Add to Compare](#)

As low as **\$228.88**

[COMPARE NOW](#)

Price Range **\$228.88 - 274.99**

Sold By 15 Merchants

**Nikon COOLPIX S8200 Black Digital Camera**

26286 - 16.1 MP, 14x Opt, 9DiSDHC/SDXC Card Slot  
Nikon's COOLPIX S8200 Digital Camera puts ease-of-use and a remarkable 16.1Mp CMOS sensor at... [More](#)

[Add to Favorites](#) [Add to Compare](#)

As low as **\$326.95**

[COMPARE NOW](#)

Price Range **\$326.95 - 329.99**

Sold By 11 Merchants

Welcome [Sign in or Register](#)

Facebook | 0 Favorites | 0 Recently Viewed | 0 Price Alerts | Feedback | [What is this?](#) | Hide



# Presentation-Oriented Documents II

camera

Digital Cameras - Compare Prices on Digital Cameras at PriceGrabber

http://cameras.pricegrabber.com/digital/p/48/

pricegrabber

Shop by Department

Digital Cameras 2,493 Matches

Featured Brand: SONY

Price Range:

- ☐ Under \$109 (629)
- ☐ \$109 - \$199 (610)
- ☐ \$199 - \$353 (642)
- ☐ \$353+ (641)

Condition:

- ☐ New (~2345)
- ☐ Used (~56)
- ☐ Refurbished (~94)

Shipping:

- ☐ Free Shipping (~1104)

Power Score:

- ☐ 7 to 10 (1677)
- ☐ 5 to 7 (271)
- ☐ 3 to 5 (33)
- ☐ Lower than 3 (4)

Top Manufacturers:

- ☐ Canon (266)
- ☐ Nikon (242)
- ☐ Sony (334)
- ☐ Panasonic (236)
- ☐ Fujifilm (138)

Grid View List View

COMPARE THESE PRODUCTS

**Featured Product**

**Sony Cyber-shot DSC-HX100V Black Digital Camera**

POWER SCORE 8.2

DSC-HX100VB - 16.2 Megapixel - 4.80 mm-144 mm - 3" LCD - 30x Optical Zoom - Optical IS - 4608 x...  
Renew your passion for photography with the Cyber-shot DSC-HX100V digital camera. A Carl Zeiss... [More](#)

As low as **\$428.95**

COMPARE NOW

Price Range **\$428.95 - 449.99**  
Sold By 10 Merchants

**Canon PowerShot SX150 IS 14.1 Megapixel Compact Camera - Black**

56648001 - 3" LCD - 12x Optical Zoom - Optical IS - 4320 x 3240 Image - 1280 x 720 Video -...  
Discover how simple and easy it is to be creative and shoot memorable photos and videos of your... [More](#)

As low as **\$228.88**

COMPARE NOW

Price Range **\$228.88 - 274.99**  
Sold By 15 Merchants

**Nikon COOLPIX S8200 Black Digital Camera**

26286 - 16.1 MP, 14x Opt, 3D/SDHC/SDXC Card Slot  
Nikon's COOLPIX S8200 Digital Camera puts ease-of-use and a remarkable 16.1Mp CMOS sensor at... [More](#)

As low as **\$326.95**

COMPARE NOW

Price Range **\$326.95 - 329.99**  
Sold By 11 Merchants

Welcome [Sign in](#) or [Register](#)

Facebook 0 Favorites 0 Recently Viewed 0 Price Alerts Feedback

[What is this?](#) | [Hide](#)



# Presentation-Oriented Documents II

camera

Its brand

Digital Cameras - Compare Prices on Digital Cameras at PriceGrabber

http://cameras.pricegrabber.com/digital/p/48/

pricegrabber

Shop by Department

Digital Cameras | Cameras | Digital Camera Batteries | Reviews & Ratings | digital cameras | digital camera | Buying Guides | Local Deals

Digital Cameras 2,493 Matches

Featured Brand

See Products from:

SONY

Grid View | List View

Price Range

- ☐ Under \$109 (629)
- ☐ \$109 - \$199 (610)
- ☐ \$199 - \$353 (642)
- ☐ \$353+ (641)

Condition

- ☐ New (~2345)
- ☐ Used (~56)
- ☐ Refurbished (~94)

Shipping

- ☐ Free Shipping (~1104)

Power Score

- ☐ 7 to 10 (1677)
- ☐ 5 to 7 (271)
- ☐ 3 to 5 (33)
- ☐ Lower than 3 (4)

Top Manufacturers

- ☐ Canon (266)
- ☐ Nikon (242)
- ☐ Sony (334)
- ☐ Panasonic (236)
- ☐ Fujifilm (136)

Compare these products

**Sony Cyber-shot DSC-HX100V Black Digital Camera**

As low as **\$428.95**

COMPARE NOW

Price Range **\$428.95 - 449.99**

Sold By 10 Merchants

**Canon PowerShot SX150 IS 14.1 Megapixel Compact Camera - Black**

As low as **\$228.88**

COMPARE NOW

Price Range **\$228.88 - 274.99**

Sold By 15 Merchants

**Nikon COOLPIX S8200 Black Digital Camera**

As low as **\$326.95**

COMPARE NOW

Price Range **\$326.95 - 329.99**

Sold By 11 Merchants



# Presentation-Oriented Documents II

camera

Its brand

Digital Cameras – Compare Prices on Digital Cameras at PriceGrabber

http://cameras.pricegrabber.com/digital/p/48/

pricegrabber

Shop by Department

Digital Cameras | Cameras | Digital Camera Batteries | Reviews & Ratings | digital cameras | digital camera | Buying Guides | Local Deals

Digital Cameras 2,493 Matches

Home | Cameras | Digital Cameras | Cameras

Featured Brand

See Products from:

SONY

Grid View | List View

ZIP Code: Sort By:

Price Range

- ☐ Under \$109 (629)
- ☐ \$109 - \$199 (610)
- ☐ \$199 - \$353 (642)
- ☐ \$353+ (641)

Condition

- ☐ New (~2345)
- ☐ Used (~56)
- ☐ Refurbished (~94)

Shipping

- ☐ Free Shipping (~1104)

Power Score

- ☐ 7 to 10 (1677)
- ☐ 5 to 7 (271)
- ☐ 3 to 5 (33)
- ☐ Lower than 3 (4)

Top Manufacturers

- ☐ Canon (266)
- ☐ Nikon (242)
- ☐ Sony (334)
- ☐ Panasonic (236)
- ☐ Fujifilm (136)

Featured Product

**Sony Cyber-shot DSC-HX100V Black Digital Camera**

POWER SCORE 8.2

DSC-HX100V - 16.2 Megapixel - 4.80 mm-144 mm - 3" LCD - 30x Optical Zoom - Optical IS - 4608 x...  
Renew your passion for photography with the Cyber-shot DSC-HX100V digital camera. A Carl Zeiss... [More](#)

As low as **\$428.95**

COMPARE NOW

Price Range **\$428.95 - 449.99**  
Sold By 10 Merchants

**Canon PowerShot SX150 IS 14.1 Megapixel Compact Camera - Black**

56648001 - 3" LCD - 12x Optical Zoom - Optical IS - 4320 x 3240 Image - 1280 x 720 Video -...  
Discover how simple and easy it is to be creative and shoot memorable photos and videos of your... [More](#)

As low as **\$228.88**

COMPARE NOW

Price Range **\$228.88 - 274.99**  
Sold By 15 Merchants

**Nikon COOLPIX S8200 Black Digital Camera**

26286 - 16.1 MP, 14x Opt. 3D/SDHC/SDXC Card Slot  
Nikon's COOLPIX S8200 Digital Camera puts ease-of-use and a remarkable 16.1Mp CMOS sensor at... [More](#)

As low as **\$326.95**

COMPARE NOW

Price Range **\$326.95 - 329.99**  
Sold By 11 Merchants

Welcome Sign in or Register

Facebook | 0 Favorites | 0 Recently Viewed | 0 Price Alerts | Feedback | [What is this?](#) | Hide



# Presentation-Oriented Documents II

camera

Its brand

Its price

Digital Cameras - Compare Prices on Digital Cameras at PriceGrabber

http://cameras.pricegrabber.com/digital/p/48/

pricegrabber

Shop by Department

Digital Cameras 2,493 Matches

Featured Brand: SONY

Price Range:

- ☐ Under \$109 (629)
- ☐ \$109 - \$199 (610)
- ☐ \$199 - \$353 (642)
- ☐ \$353+ (641)

Condition:

- ☐ New (~2345)
- ☐ Used (~56)
- ☐ Refurbished (~94)

Shipping:

- ☐ Free Shipping (~1104)

Power Score

- ☐ 7 to 10 (1677)
- ☐ 5 to 7 (271)
- ☐ 3 to 5 (33)
- ☐ Lower than 3 (4)

Top Manufacturers:

- ☐ Canon (266)
- ☐ Nikon (242)
- ☐ Sony (334)
- ☐ Panasonic (236)
- ☐ Fujifilm (136)

Featured Product: Sony Cyber-shot DSC-HX100V Black Digital Camera

As low as **\$428.95**

Price Range: \$428.95 - 449.99

Sold By 10 Merchants

Canon PowerShot SX150 IS 14.1 Megapixel Compact Camera - Black

As low as **\$228.88**

Price Range: \$228.88 - 274.99

Sold By 15 Merchants

Nikon COOLPIX S8200 Black Digital Camera

As low as **\$326.95**

Price Range: \$326.95 - 329.99

Sold By 11 Merchants



# Presentation-Oriented Documents II

camera

Its brand

Its price

Digital Cameras – Compare Prices on Digital Cameras at PriceGrabber

http://cameras.pricegrabber.com/digital/p/48/

pricegrabber

Shop by Department

Digital Cameras | Cameras | Digital Camera Batteries | Reviews & Ratings | digital cameras | digital camera | Buying Guides | Local Deals

Digital Cameras 2,493 Matches

Home | Cameras | Digital Cameras | Cameras

Featured Brand

See Products from:

SONY

Grid View | List View

ZIP Code: Sort By:

Price Range

- ☐ Under \$109 (629)
- ☐ \$109 - \$199 (610)
- ☐ \$199 - \$353 (642)
- ☐ \$353+ (641)

Condition

- ☐ New (~2345)
- ☐ Used (~56)
- ☐ Refurbished (~94)

Shipping

- ☐ Free Shipping (~1104)

Power Score

- ☐ 7 to 10 (1677)
- ☐ 5 to 7 (271)
- ☐ 3 to 5 (33)
- ☐ Lower than 3 (4)

Top Manufacturers

- ☐ Canon (266)
- ☐ Nikon (242)
- ☐ Sony (334)
- ☐ Panasonic (236)
- ☐ Fujifilm (136)

Featured Product

**Sony Cyber-shot DSC-HX100V Black Digital Camera**

POWER SCORE 8.2

DSC-HX100VB - 16.2 Megapixel - 4.80 mm-144 mm - 3" LCD - 30x Optical Zoom - Optical IS - 4608 x...  
Renew your passion for photography with the Cyber-shot DSC-HX100V digital camera. A Carl Zeiss... [More](#)

As low as **\$428.95**

COMPARE NOW

Price Range **\$428.95 - 449.99**  
Sold By 10 Merchants

**Canon PowerShot SX150 IS 14.1 Megapixel Compact Camera - Black**

56648001 - 3" LCD - 12x Optical Zoom - Optical IS - 4320 x 3240 Image - 1280 x 720 Video -...  
Discover how simple and easy it is to be creative and shoot memorable photos and videos of your... [More](#)

As low as **\$228.88**

COMPARE NOW

Price Range **\$228.88 - 274.99**  
Sold By 15 Merchants

**Nikon COOLPIX S8200 Black Digital Camera**

26286 - 16.1 MP, 14x Opt. 3D/SDHC/SDXC Card Slot  
Nikon's COOLPIX S8200 Digital Camera puts ease-of-use and a remarkable 16.1Mp CMOS sensor at... [More](#)

As low as **\$326.95**

COMPARE NOW

Price Range **\$326.95 - 329.99**  
Sold By 11 Merchants

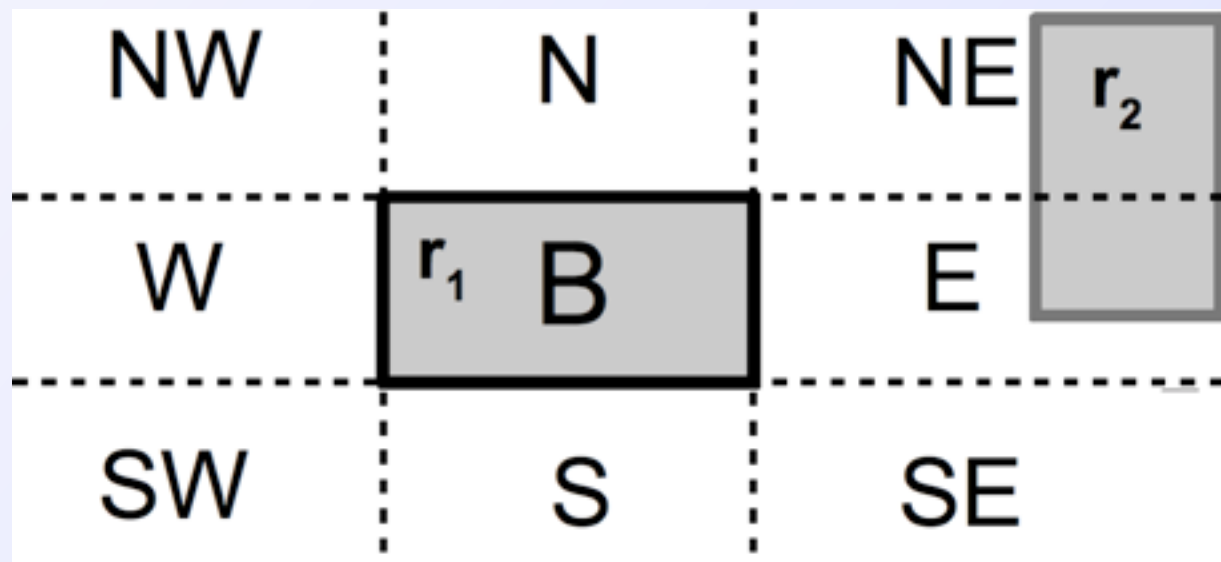
Welcome Sign in or Register

Facebook | 0 Favorites | 0 Recently Viewed | 0 Price Alerts | Feedback | [What is this?](#) | Hide

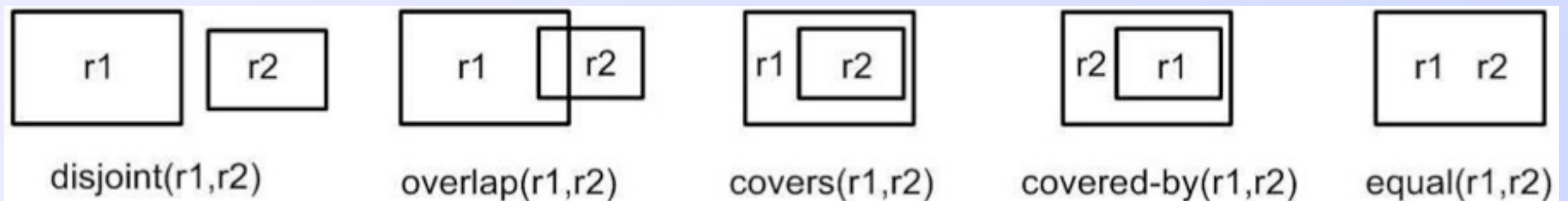


# Spatial Models

## Rectangular Cardinal Relations (RCR)



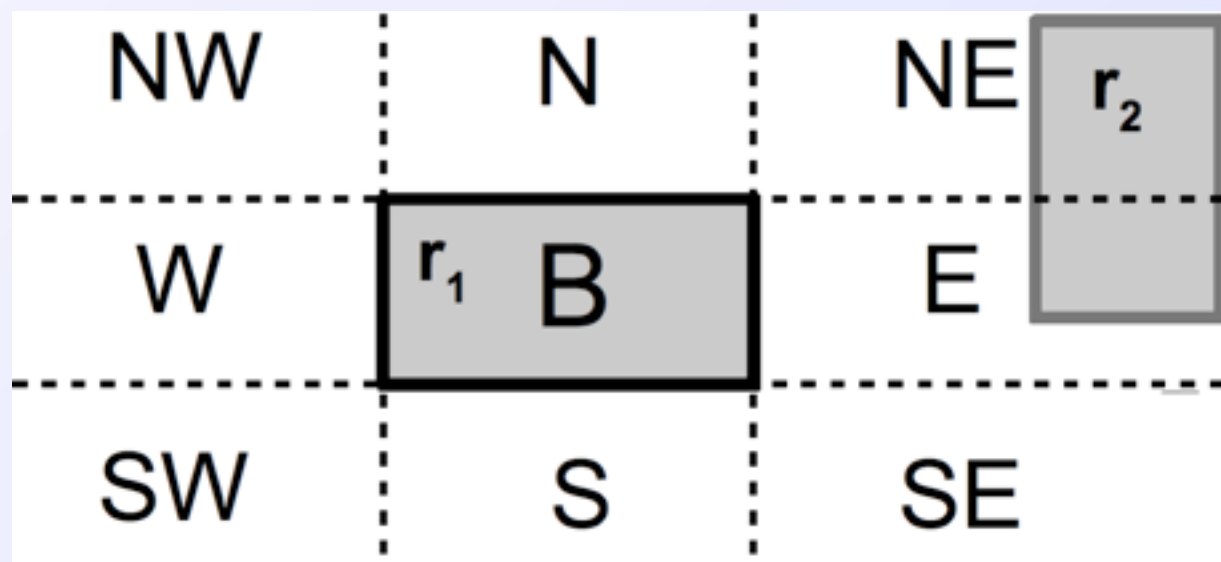
## Topological Relations





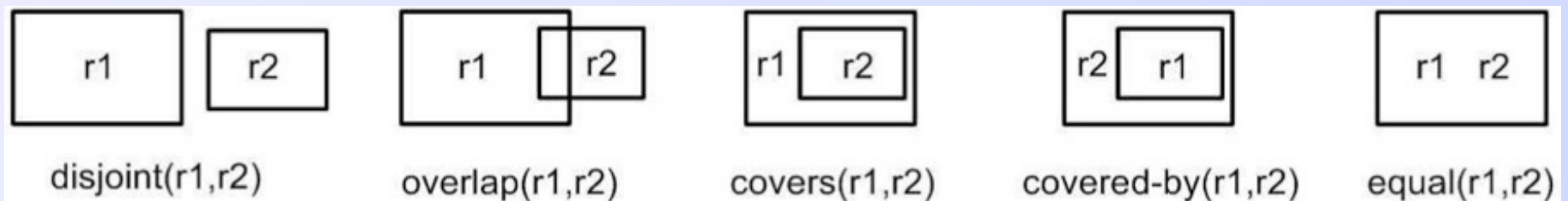
# Spatial Models

## Rectangular Cardinal Relations (RCR)



$r_1$  **E:NE**  $r_2$

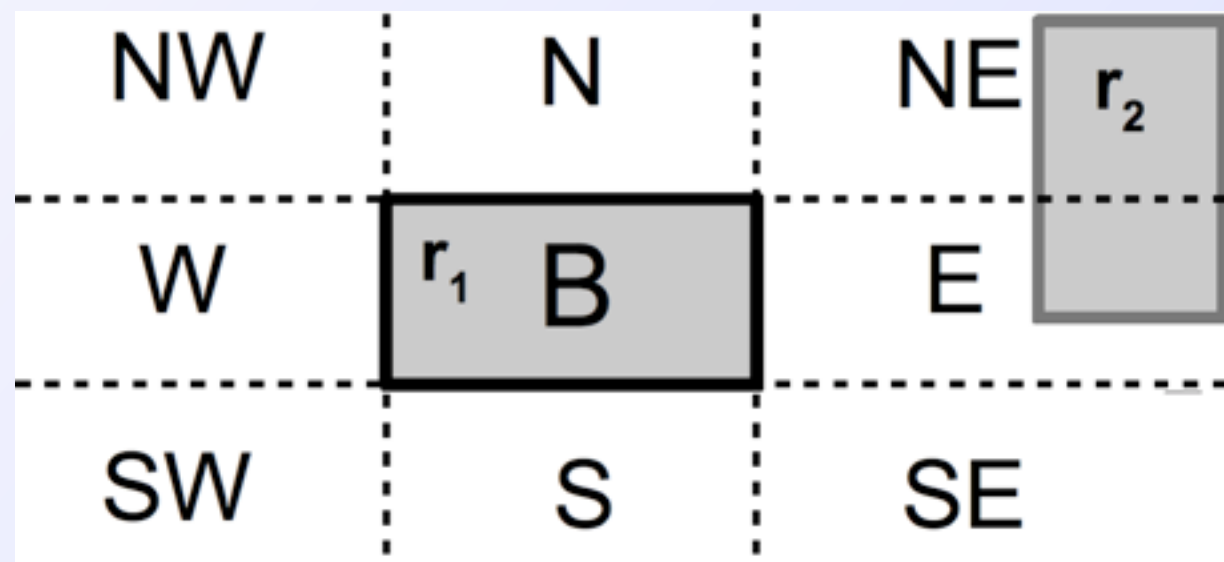
## Topological Relations





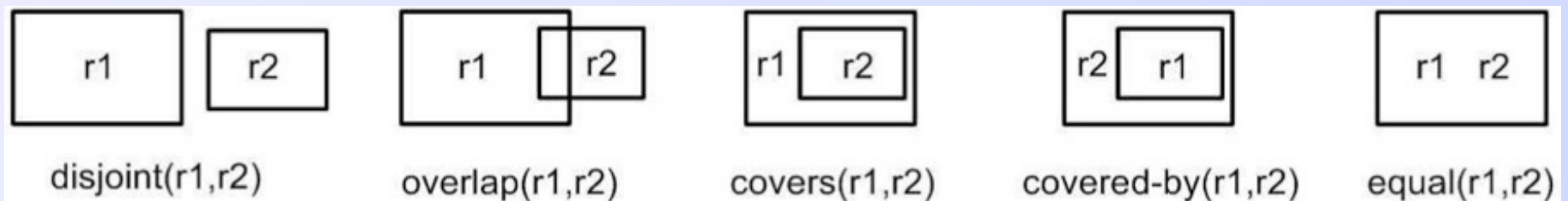
# Spatial Models

## Rectangular Cardinal Relations (RCR)



$r_1$  **E:NE**  $r_2$   
 eastAndNorthEast( $r_1, r_2$ )

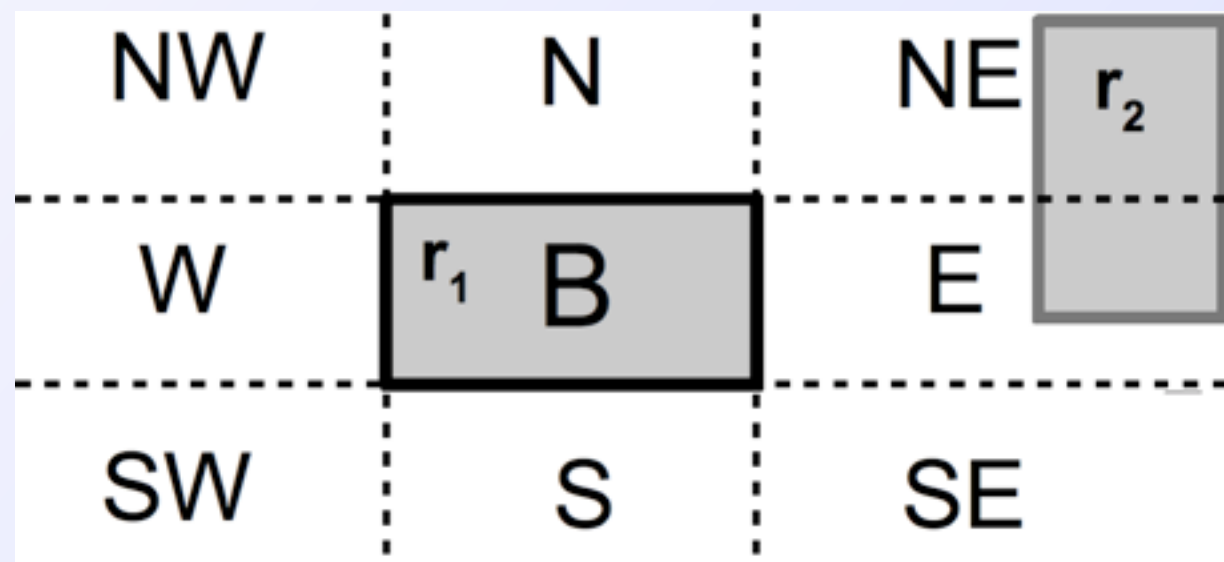
## Topological Relations





# Spatial Models

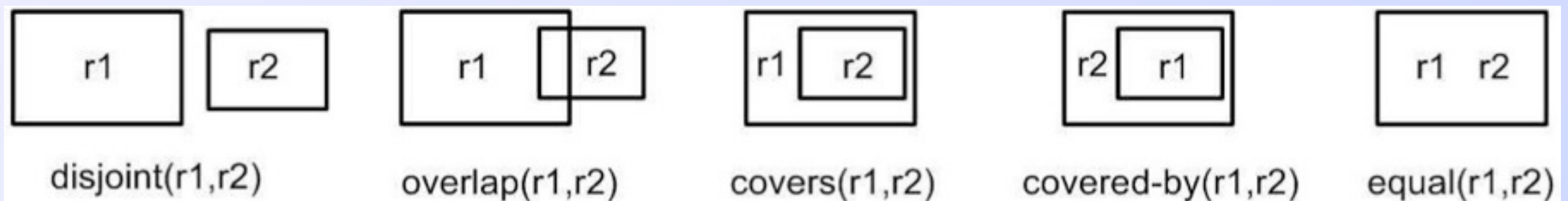
## Rectangular Cardinal Relations (RCR)



$r_1$  **E:NE**  $r_2$   
`eastAndNorthEast( $r_1, r_2$ )`

Spatial models allows for expressing disjunctive relations among regions

## Topological Relations





# Spatial Models: Example

To express spatial preferences



Featured Product

## Sony Cyber-shot DSC-HX100V Black Digital Camera

POWER SCORE 9.2

DSCHX100VB - 16.2 Megapixel - 4.80 mm-144 mm - 3" LCD - 30x Optical Zoom - Optical IS - 4608 x...

Renew your passion for photography with the Cyber-shot DSC-HX100V digital camera. A Carl Zeiss... [More](#)

Add to Favorites Add to Compare

As low as

**\$428.95**

[COMPARE NOW](#)

Price Range

\$428.95 - 449.99

Sold By 10 Merchants



# Spatial Models: Example

To express spatial preferences



Featured Product

**Sony Cyber-shot DSC-HX100V Black Digital Camera**

POWER SCORE 9.2

DSCHX100VB - 16.2 Megapixel - 4.80 mm-144 mm - 3" LCD - 30x Optical Zoom - Optical IS - 4608 x...

Renew your passion for photography with the Cyber-shot DSC-HX100V digital camera. A Carl Zeiss... [More](#)

Add to Favorites Add to Compare

As low as

**\$428.95**

[COMPARE NOW](#)

Price Range

\$428.95 - 449.99

Sold By 10 Merchants

IMG **E** brand



# Spatial Models: Example

To express spatial preferences



Featured Product

**Sony Cyber-shot DSC-HX100V Black Digital Camera**

POWER SCORE 9.2

DSCHX100VB - 16.2 Megapixel - 4.80 mm-144 mm - 3" LCD - 30x Optical Zoom - Optical IS - 4608 x...

Renew your passion for photography with the Cyber-shot DSC-HX100V digital camera. A Carl Zeiss... [More](#)

Add to Favorites Add to Compare

As low as

**\$428.95**

[COMPARE NOW](#)

Price Range

\$428.95 - 449.99

Sold By 10 Merchants

IMG **E** brand

IMG **E** price



# Spatial Models: Example

In Fotocamere Digitali

▼ **Marca**

- ☐ Canon (1.788)
- ☐ Nikon (1.701)
- ☐ Samsung (1.454)
- ☐ Sony (1.232)
- Scegli altro...

▼ **Modello**

- ☐ EOS 5D (103)
- ☐ D90 (96)
- ☐ CYBER-SHOT DSC-T100 (28)
- ☐ MICRO (23)
- Scegli altro...

▼ **Condizione**

- ☐ Nuovo (9.787)
- ☐ Nuovo: altro (vedi dettagli) (45)
- ☐ Ricondizionato dal produttore (1)
- ☐ Ricondizionato dal venditore (3)
- Scegli altro...

▼ **Prezzo**

€  a €  >>

► **Tipo di fotocamera digitale**

► **Zoom ottico**

► **Formato scheda di memoria**

► **Venditore**

▼ **Categorie**

Fotocamere digitali

- 6 Megapixels e oltre (8.475)
- Da 5.0 a 5.9 Megapixels (166)
- Da 4.0 a 4.9 Megapixels (28)
- Reflex Digitali, SLR (1.577)

Tutti gli oggetti Solo Aste online Compralo Subito


Vedi come:

Mostra prima: Inserzioni più rilevanti

Spedizione a ITA


Pagina 1 di 434

Oggetto In Vetrina




**Fotocamera Digitale V527 FUCSIA**  
12 MP 8x Zoom LCD 2,7"

Compralo Subito EUR 47,00




**FOTOCAMERA DIGITALE BENQ**  
DCE 300

11m EUR 19,00



Ingrandisci  
**Fotocamera Digitale SUBACQUEA**  
fino a 3 MT 12 MegaPixel


Compralo Subito EUR 52,90



Ingrandisci  
**FOTOCAMERA DIGITALE ★**  
KODAK M575 14 MPX ★  
EASYSHARE ZOO...


Compralo Subito EUR 85,90  
*Proposta d'acquisto*

Promuovi i tuoi oggetti




Ingrandisci  
**Fotocamera Digitale SUBACQUEA**  
x bambini fino a 3 Metri

Compralo Subito EUR 54,90



**Fotocamera digitale Panasonic DMC**  
fz30


9m EUR 100,00



**FOTOCAMERA DIGITALE x bambini**  
8 Mega Pixel rosa V8690

Compralo Subito EUR 33,90

Affidabilità Top



Ingrandisci  
**Fotocamera Digitale V527 NERA**  
12 MP 8x Zoom LCD 2,7"

Compralo Subito EUR 47,00

DIRETTAMENTE DALL'IMPORTATORE



# Spatial Models: Example

In Fotocamere Digitali

▼ Marca

- ☐ Canon (1.788)
- ☐ Nikon (1.701)
- ☐ Samsung (1.454)
- ☐ Sony (1.232)
- Scegli altro...

▼ Modello

- ☐ EOS 5D (103)
- ☐ D90 (98)
- ☐ CYBER-SHOT DSC-T100 (28)
- ☐ MICRO (23)
- Scegli altro...

▼ Condizione

- ☐ Nuovo (9.787)
- ☐ Nuovo: altro (vedi dettagli) (45)
- ☐ Ricondizionato dal produttore (1)
- ☐ Ricondizionato dal venditore (3)
- Scegli altro...

▼ Prezzo

€  a €

► Tipo di fotocamera digitale

► Zoom ottico

► Formato scheda di memoria

► Venditore

▼ Categorie

Fotocamere digitali

- 6 Megapixels e oltre (8.475)
- Da 5.0 a 5.9 Megapixels (166)
- Da 4.0 a 4.9 Megapixels (28)
- Reflex Digitali, SLR (1.577)

Tutti gli oggetti Solo Aste online Compralo Subito

Vedi come:

Mostra prima: Inserzioni più rilevanti Pagina 1 di 434

Spedizione a ITA

Oggetto In Vetrina

Fotocamera Digitale V527 FUCSIA 12 MP 8x Zoom LCD 2,7"

Compralo Subito EUR 47,00

FOTOCAMERA DIGITALE BENQ DCE 300

11m EUR 19,00

Promuovi i tuoi oggetti

Ingrandisci Fotocamera Digitale SUBACQUEA x bambini fino a 3 Metri

Compralo Subito EUR 54,90

Ingrandisci Fotocamera digitale Panasonic DMC fz30

9m EUR 100,00

Ingrandisci FOTOCAMERA DIGITALE x bambini 8 Mega Pixel rosa V8690

Compralo Subito EUR 33,90

Ingrandisci Fotocamera Digitale V527 NERA 12 MP 8x Zoom LCD 2,7"

Compralo Subito EUR 47,00

DIRETTAMENTE DALL'IMPORTATORE

Affidabilità Top

Which is the correct price?

IMG S price



# Spatial Models: Example

In Fotocamere Digitali

▼ Marca

- ☐ Canon (1.788)
- ☐ Nikon (1.701)
- ☐ Samsung (1.454)
- ☐ Sony (1.232)
- Scegli altro...

▼ Modello

- ☐ EOS 5D (103)
- ☐ D90 (98)
- ☐ CYBER-SHOT DSC-T100 (28)
- ☐ MICRO (23)
- Scegli altro...

▼ Condizione

- ☐ Nuovo (9.787)
- ☐ Nuovo: altro (vedi dettagli) (45)
- ☐ Ricondizionato dal produttore (1)
- ☐ Ricondizionato dal venditore (3)
- Scegli altro...

▼ Prezzo

€  a €

► Tipo di fotocamera digitale

► Zoom ottico

► Formato scheda di memoria

► Venditore

▼ Categorie

Fotocamere digitali

- 6 Megapixels e oltre (8.475)
- Da 5.0 a 5.9 Megapixels (166)
- Da 4.0 a 4.9 Megapixels (28)
- Reflex Digitali, SLR (1.577)

Tutti gli oggetti Solo Aste online Compralo Subito

Vedi come:

Mostra prima: Inserzioni più rilevanti Pagina 1 di 434

Spedizione a ITA

Oggetto In Vetrina

Fotocamera Digitale V527 FUCSIA 12 MP 8x Zoom LCD 2,7"

Compralo Subito EUR 47,00

FOTOCAMERA DIGITALE BENQ DCE 300

11m **EUR 19,00**

Promuovi i tuoi oggetti

Ingrandisci Fotocamera Digitale SUBACQUEA x bambini fino a 3 Metri

Compralo Subito EUR 54,90

Ingrandisci Fotocamera digitale Panasonic DMC fz30

9m **EUR 100,00**

FOTOCAMERA DIGITALE x bambini 8 Mega Pixel rosa V8690

Compralo Subito EUR 33,90

Ingrandisci Fotocamera Digitale V527 NERA 12 MP 8x Zoom LCD 2,7"

Compralo Subito EUR 47,00

DIRETTAMENTE DALL'IMPORTATORE

Affidabilità Top

Which is the correct price?

IMG IGS price



# Content Semantics

- Basic data types defined through
  - dictionaries (Gazetteers)

## CAMERA BRAND DICTIONARY

benq	olympus
bushnell	panasonic
canon	pentax
casio	polaroid
cobra digital	protax
flip	ricoh
fuji	rollei
fujifilm	samsung
hasselblad	sanyo
kodak	sealife
leica	sigma
mamiya	sony
megxon	voigtlander
minox	vtech
nikkor	...
nikon	



# Content Semantics

- **Basic data types** defined through
  - dictionaries (Gazetteers)
  - XOnto rules (Extraction rules + ontologies)
  - Dom nodes



# Content Semantics

- **Basic data types** defined through
  - dictionaries (Gazetteers)
  - XOnto rules (Extraction rules + ontologies)
  - Dom nodes

brand  $\leftarrow$  entity(dictionaryBrand)

price  $\leftarrow$  number, currency

price  $\leftarrow$  currency, number

currency  $\leftarrow$  regex("\u20AC|([eE][uU][rR][oO]?)\\$")



# Content Semantics

## Basic data types defined through

- dictionaries (Gazetteers)
- XOnto rules (Ext)
- Dom nodes

Oro, Ruffolo. 2008. **XONTO: An Ontology-Based System for Semantic Information Extraction from PDF Documents.** *ICTAI*.

brand ← entity(dictionaryBrand)

price ← number, currency


price ← currency, number

currency ← regex("\u20AC|([eE][uU][rR][oO]?)|\\$")





# Content Semantics

- Spatial and visual relations among contents





Featured Product

**Sony Cyber-shot DSC-HX100V Black Digital Camera**

POWER SCORE  **9.2** 

DSCHX100VB - 16.2 Megapixel - 4.80 mm-144 mm - 3" LCD - 30x Optical Zoom - Optical IS - 4608 x...

Renew your passion for photography with the Cyber-shot DSC-HX100V digital camera. A Carl Zeiss... [More](#)

 Add to Favorites  Add to Compare

As low as

**\$428.95**

[COMPARE NOW](#)

Price Range

**\$428.95 - 449.99**


Sold By 10 Merchants



# Content Semantics

- Spatial and visual relations among contents

A brand



Featured Product

**Sony Cyber-shot DSC-HX100V Black Digital Camera**

POWER SCORE 9.2

DSCHX100VB - 16.2 Megapixel - 4.80 mm-144 mm - 3" LCD - 30x Optical Zoom - Optical IS - 4608 x...

Renew your passion for photography with the Cyber-shot DSC-HX100V digital camera. A Carl Zeiss... [More](#)

♥ Add to Favorites ■ Add to Compare

As low as  
**\$428.95**

[COMPARE NOW](#)

Price Range  
**\$428.95 - 449.99**

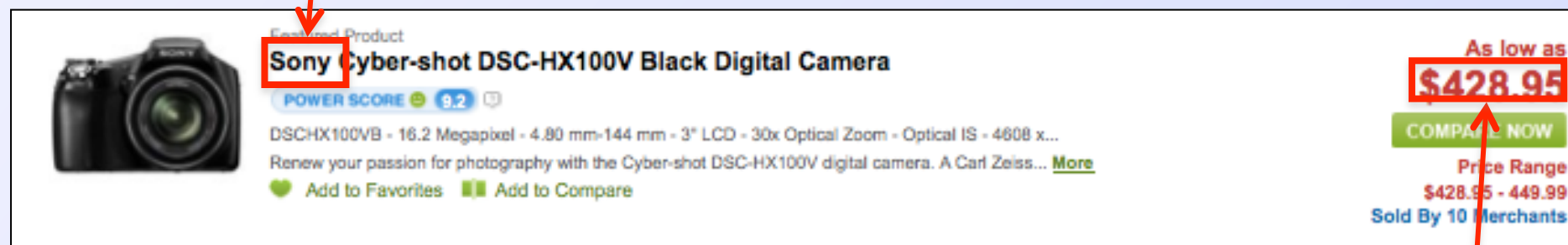
Sold By 10 Merchants



# Content Semantics

- Spatial and visual relations among contents

A brand



Featured Product

**Sony Cyber-shot DSC-HX100V Black Digital Camera**

POWER SCORE 9.2

DSCHX100VB - 16.2 Megapixel - 4.80 mm-144 mm - 3" LCD - 30x Optical Zoom - Optical IS - 4608 x...

Renew your passion for photography with the Cyber-shot DSC-HX100V digital camera. A Carl Zeiss... [More](#)

[Add to Favorites](#) [Add to Compare](#)

As low as **\$428.95**

[COMPARE NOW](#)

Price Range  
\$428.95 - 449.99  
Sold By 10 Merchants

A price



# Content Semantics

- Spatial and visual relations among contents

A brand

An IMG

A price

Featured Product

**Sony Cyber-shot DSC-HX100V Black Digital Camera**

POWER SCORE 9.2

DSCHX100VB - 16.2 Megapixel - 4.80 mm-144 mm - 3" LCD - 30x Optical Zoom - Optical IS - 4608 x...

Renew your passion for photography with the Cyber-shot DSC-HX100V digital camera. A Carl Zeiss... [More](#)

[Add to Favorites](#) [Add to Compare](#)

As low as **\$428.95**

[COMPARE NOW](#)

Price Range  
\$428.95 - 449.99  
Sold By 10 Merchants



# Concept Descriptors

camera ← photo, brand, price    [photo IGE|IGS price]

The diagram illustrates the components of concept descriptors. For the descriptor 'camera', 'photo' is identified as the **object**, while 'brand' and 'price' are grouped as **Its attributes**. For the descriptor '[photo IGE|IGS price]', 'photo' is the **object**, and 'IGE|IGS' and 'price' are grouped as **Constraints and preferences**.




# Preferences and Probabilities

- Toward Best-effort Information Extraction
- Null attributes
- Multi-value attributes



# Preferences and Probabilities

- Toward Best-effort Information Extraction
- Null attributes
- Multi-value attributes


 <p>Reviewed on June 17, 2011</p>	<p><b>Sony Cyber-shot DSC-HX9V (Black)</b></p> <p>CNET Editor rating: ★★★★★ Excellent</p> <p>Average user rating: ★★★★★ out of 24 reviews</p> <p>The Bottom Line: Feature junkies in search of a compact megazoom should get the Sony Cyber-shot DSC-HX9V. <a href="#">Read review</a></p>	<p><b>\$322 to \$395</b> at 15 stores</p> <p><b>\$349.99</b> - Sony Electronics, Inc. <b>\$395.87</b> - eBay <b>\$348.00</b> - Buydig.com</p>
--	--	---



# Preferences and Probabilities

- Toward Best-effort Information Extraction
- Null attributes
- Multi-value attributes

Which price is  
the desired  
answer?

 <p>Reviewed on June 17, 2011</p>	<h3>Sony Cyber-shot DSC-HX9V (Black)</h3> <p>CNET Editor rating: ★★★★★ Excellent</p> <p>Average user rating: ★★★★★ out of 24 reviews</p> <p>The Bottom Line: Feature junkies in search of a compact megazoom should get the Sony Cyber-shot DSC-HX9V. <a href="#">Read review</a></p>	<p><b>\$322 to \$395</b> at 15 stores</p> <p><b>\$349.99</b> - Sony Electronics, Inc.</p> <p><b>\$395.87</b> - eBay</p> <p><b>\$348.00</b> - Buydig.com</p>
--	---	---



# Wrappo Architecture



# Wrappo Architecture

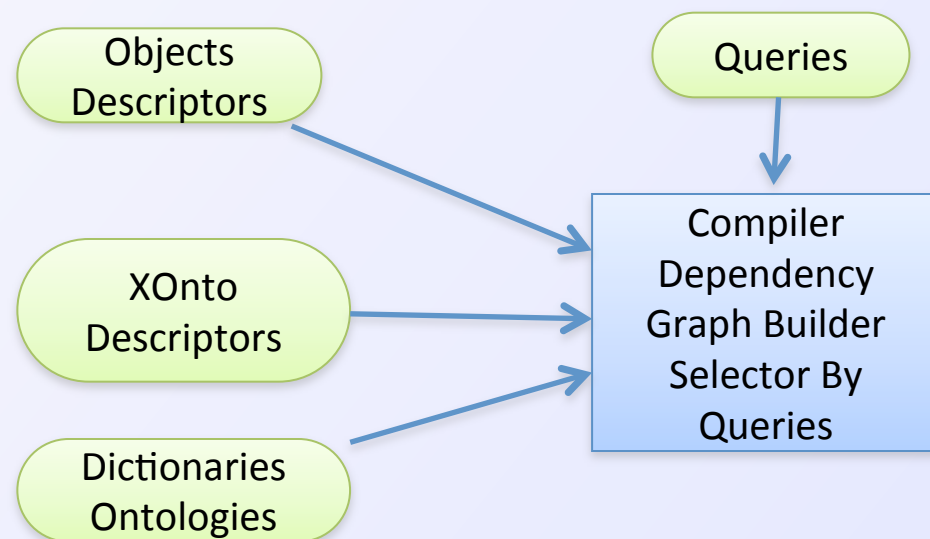
Objects  
Descriptors

XOnto  
Descriptors

Dictionaries  
Ontologies

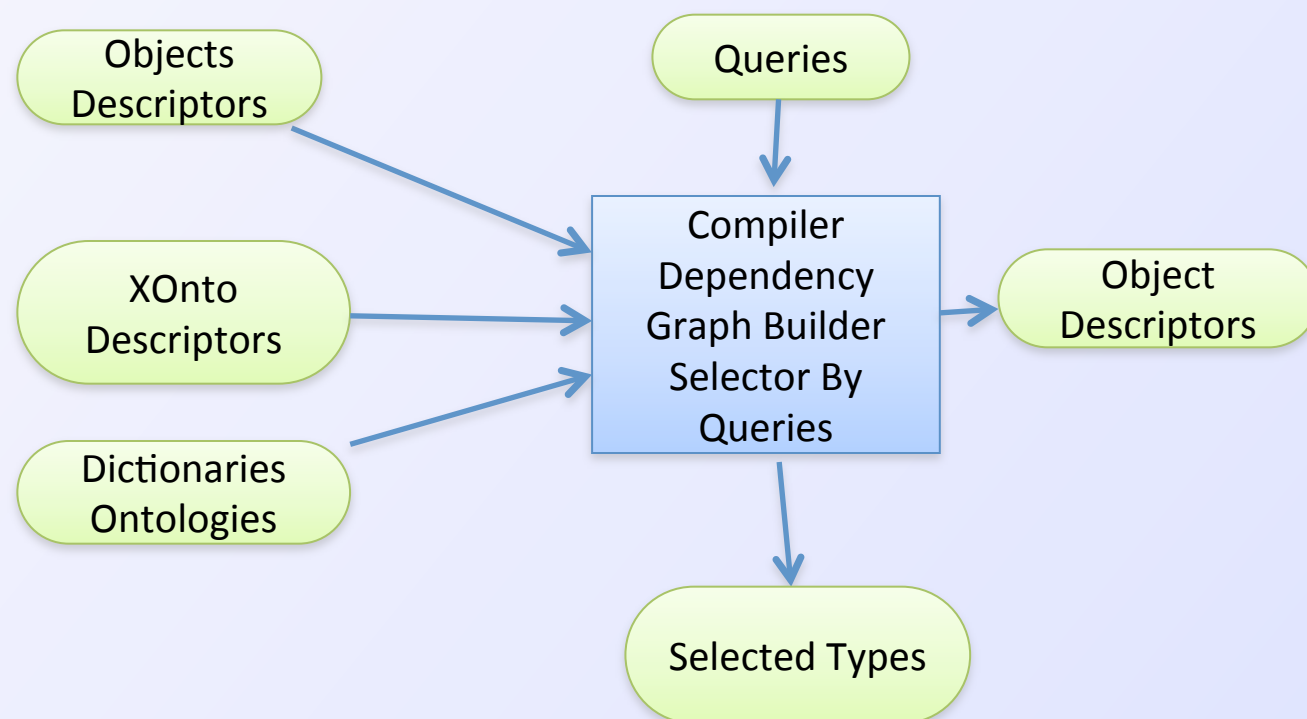


# Wrappo Architecture



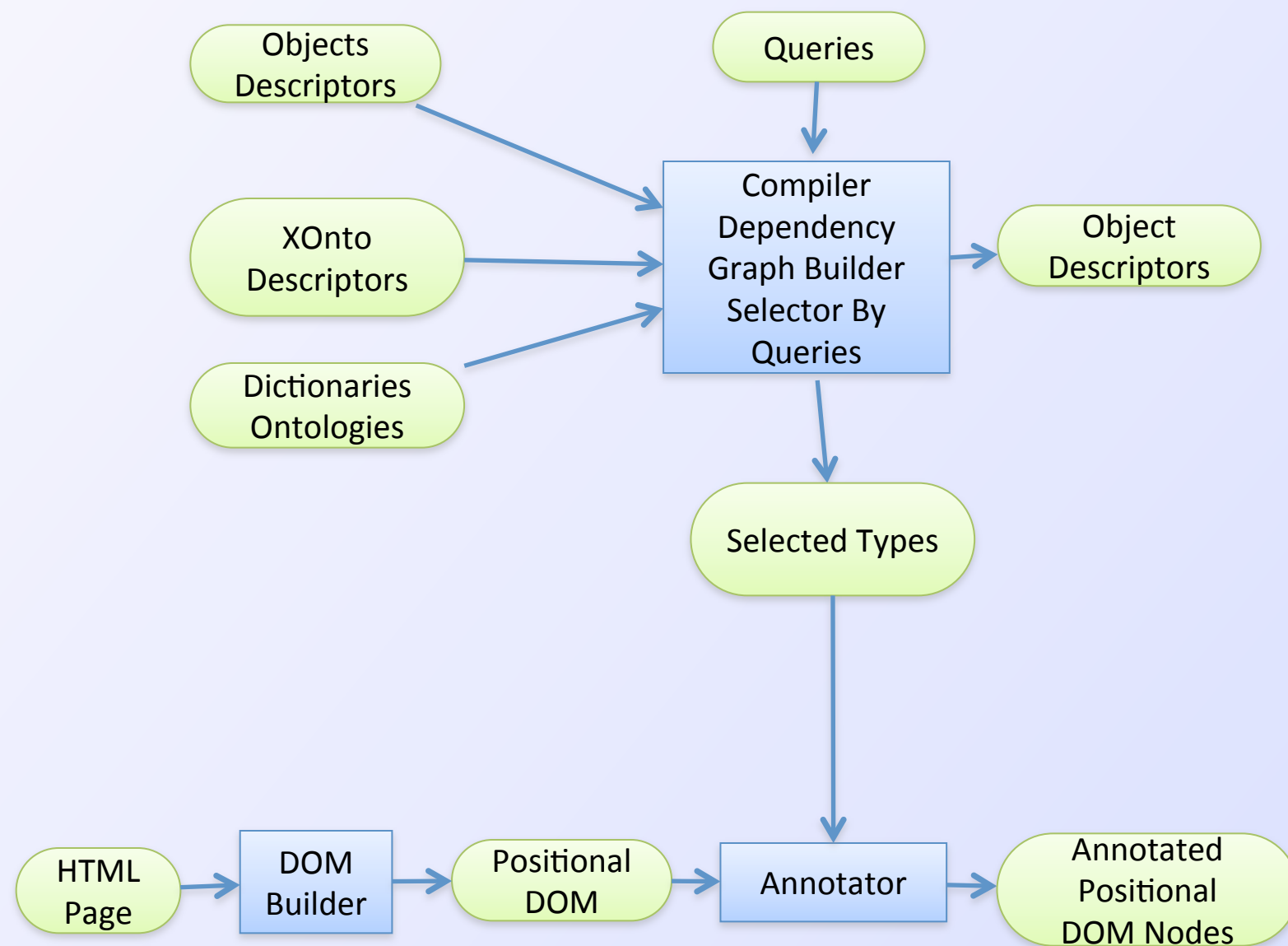


# Wrappo Architecture



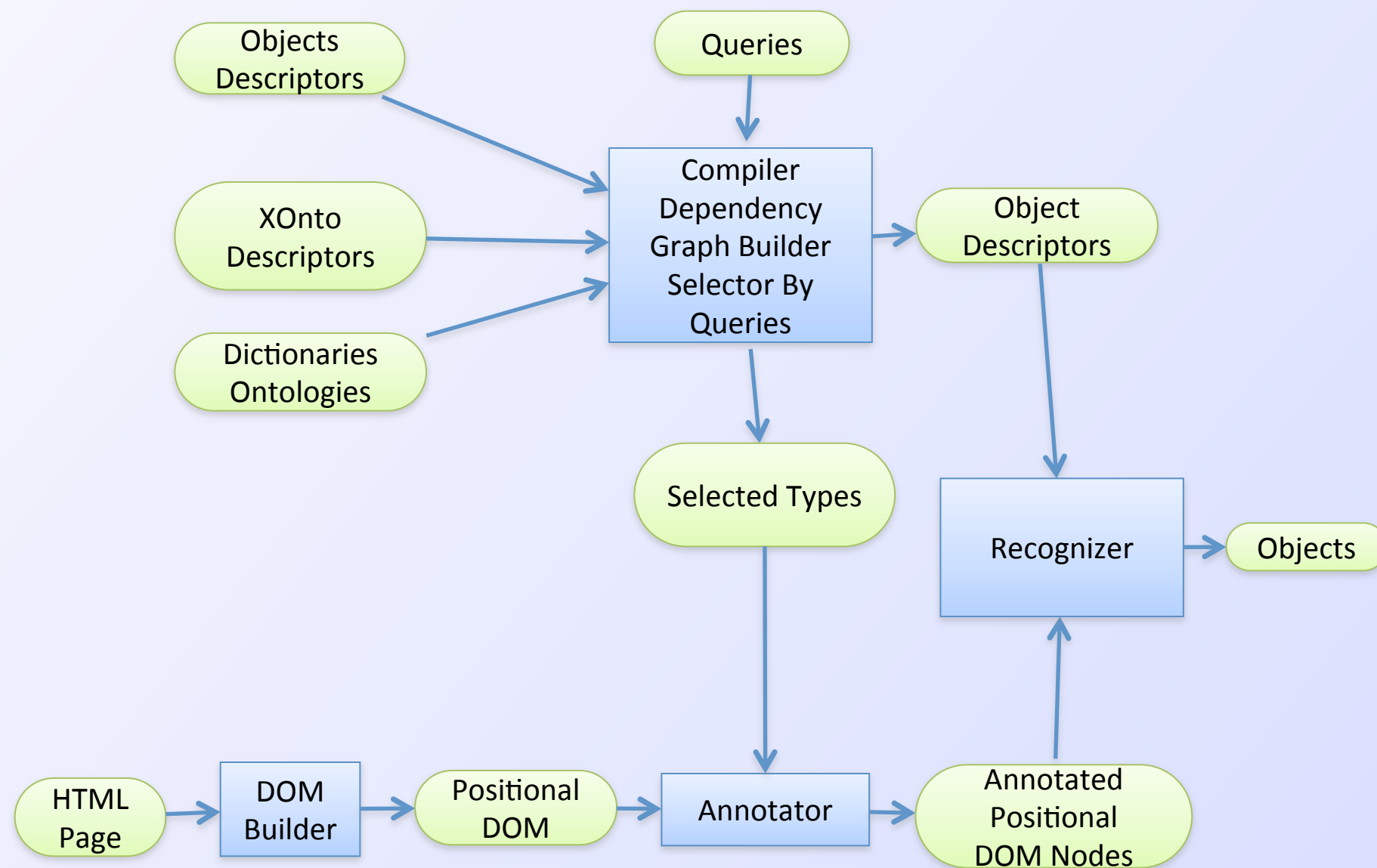


# Wrappo Architecture



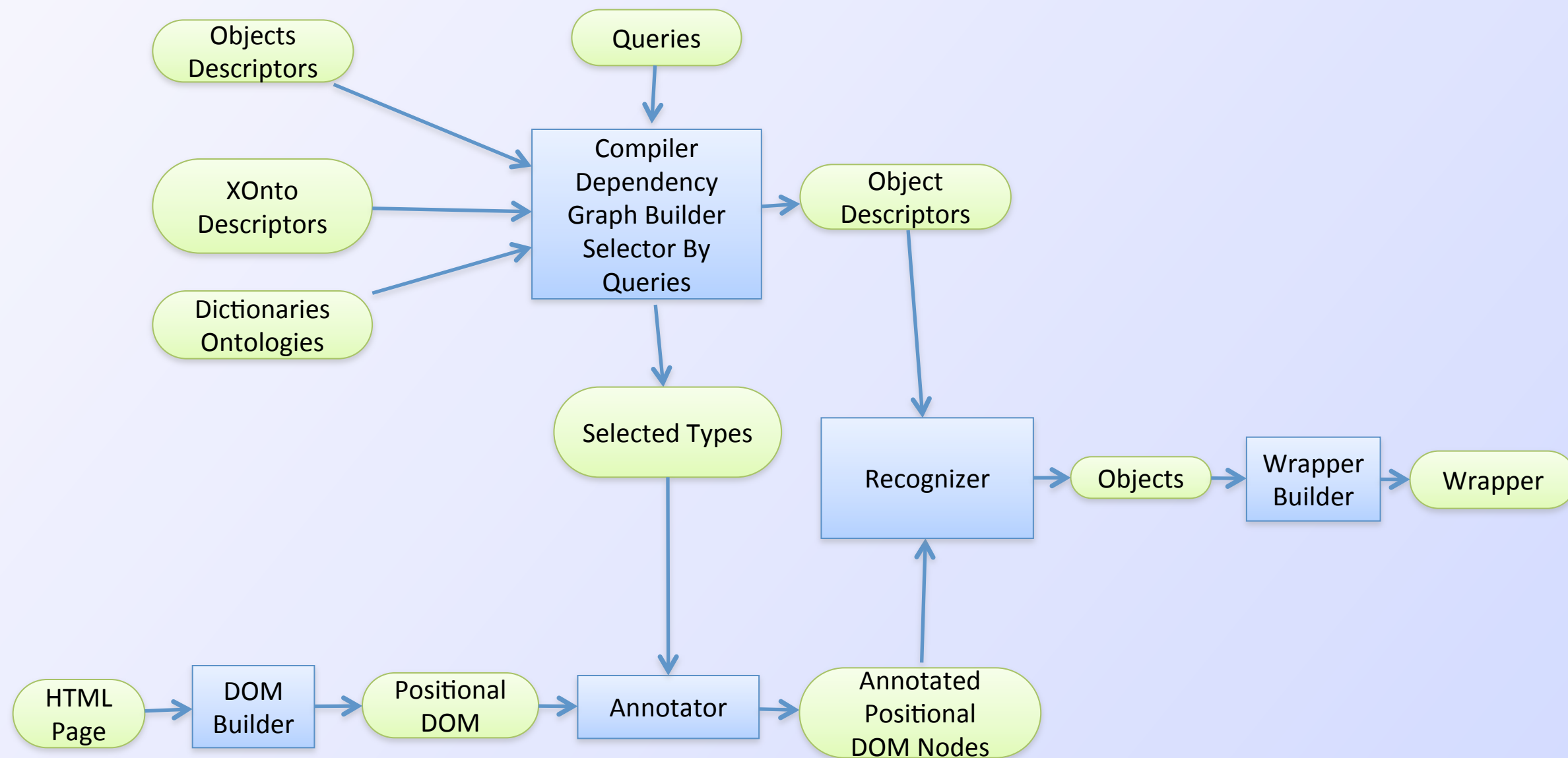


# Wrappo Architecture



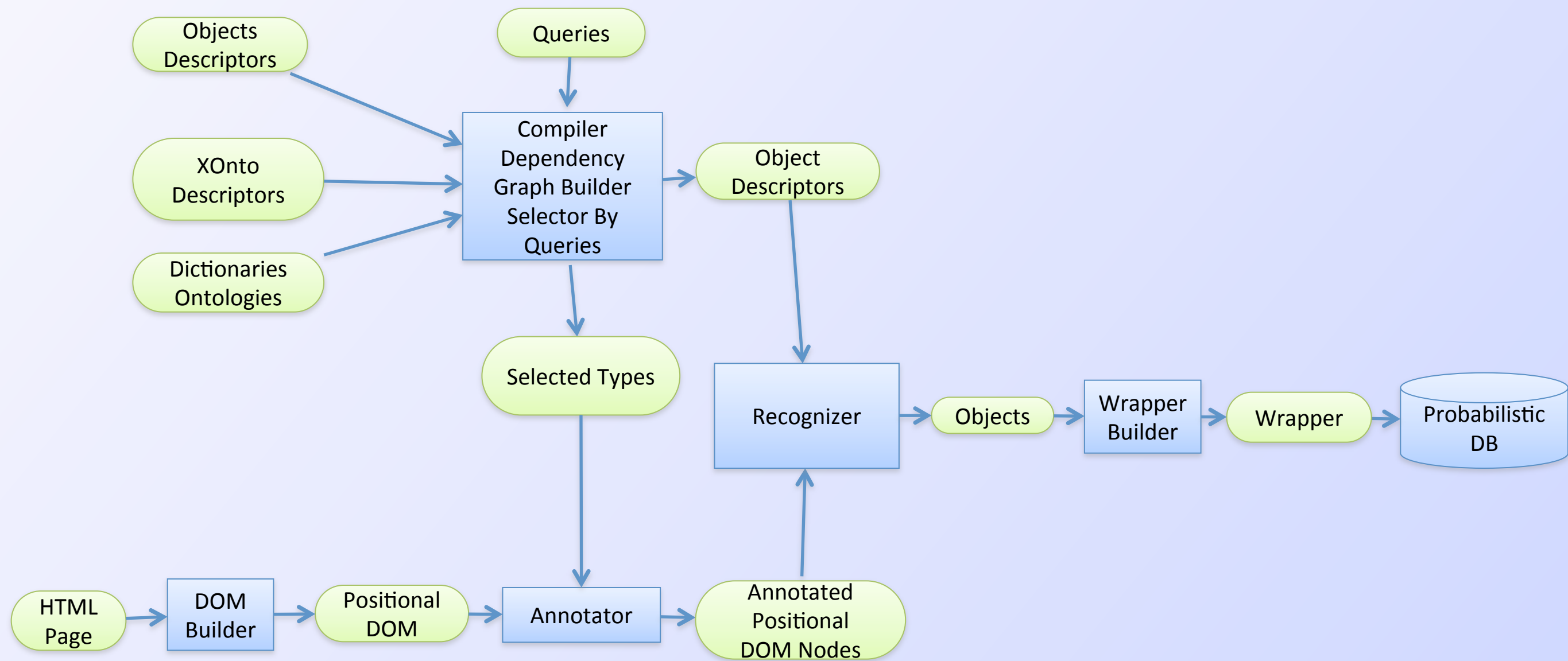


# Wrappo Architecture



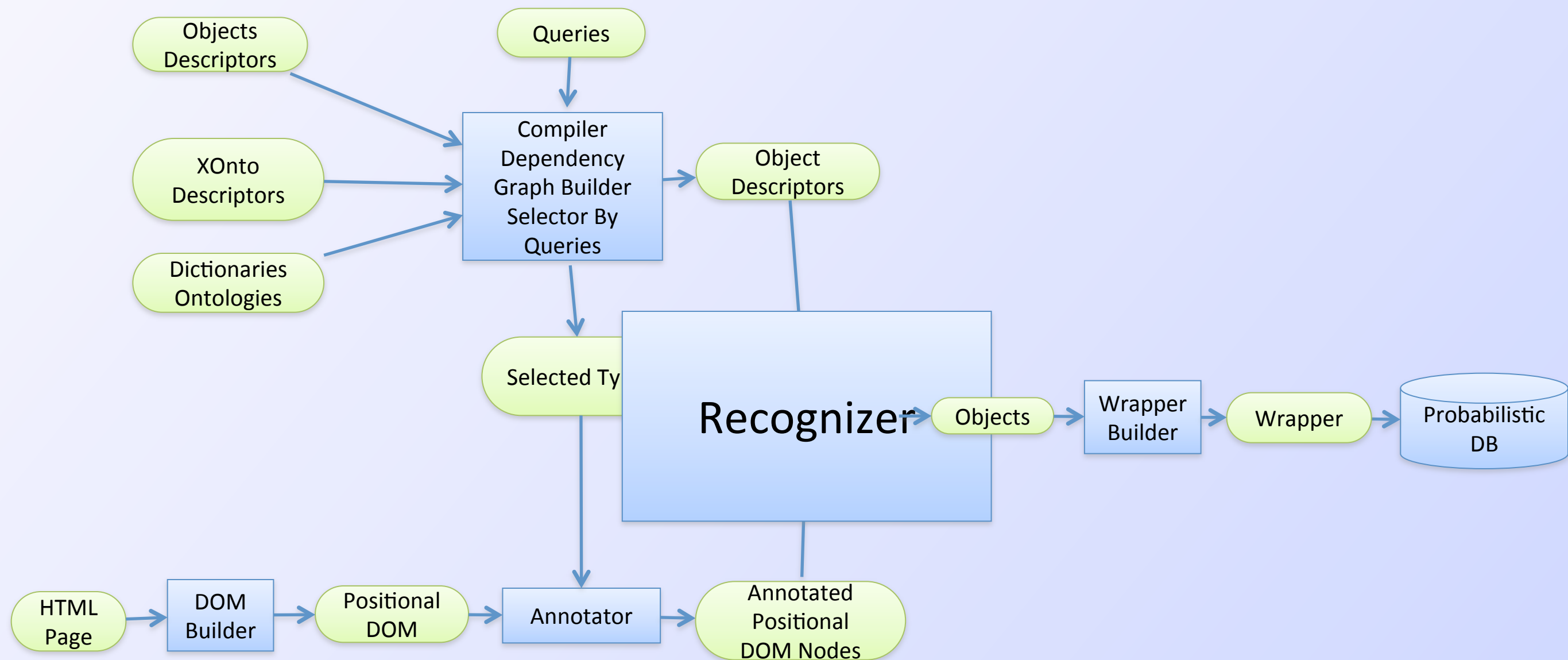


# Wrappo Architecture



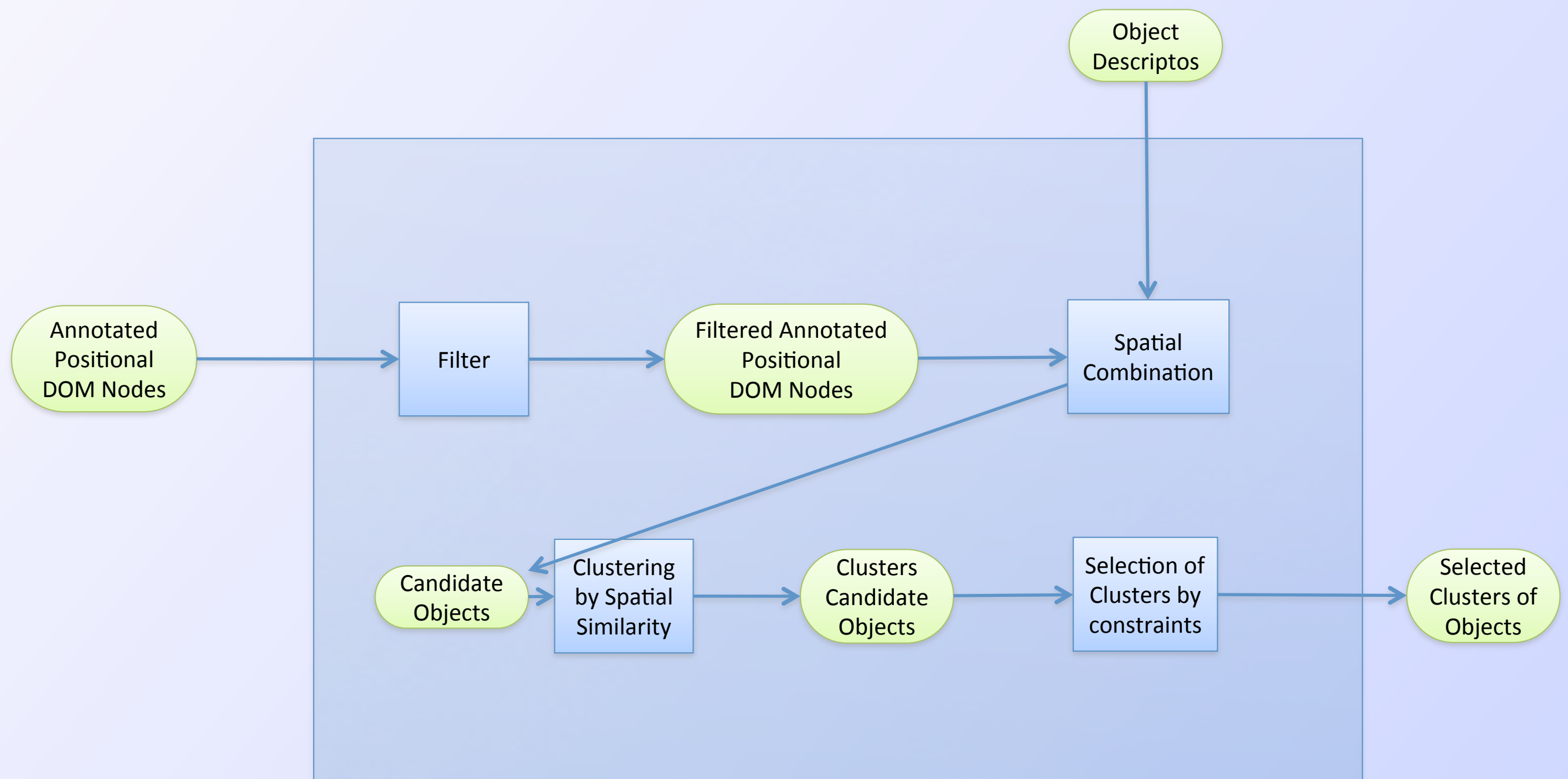


# Wrappo Architecture



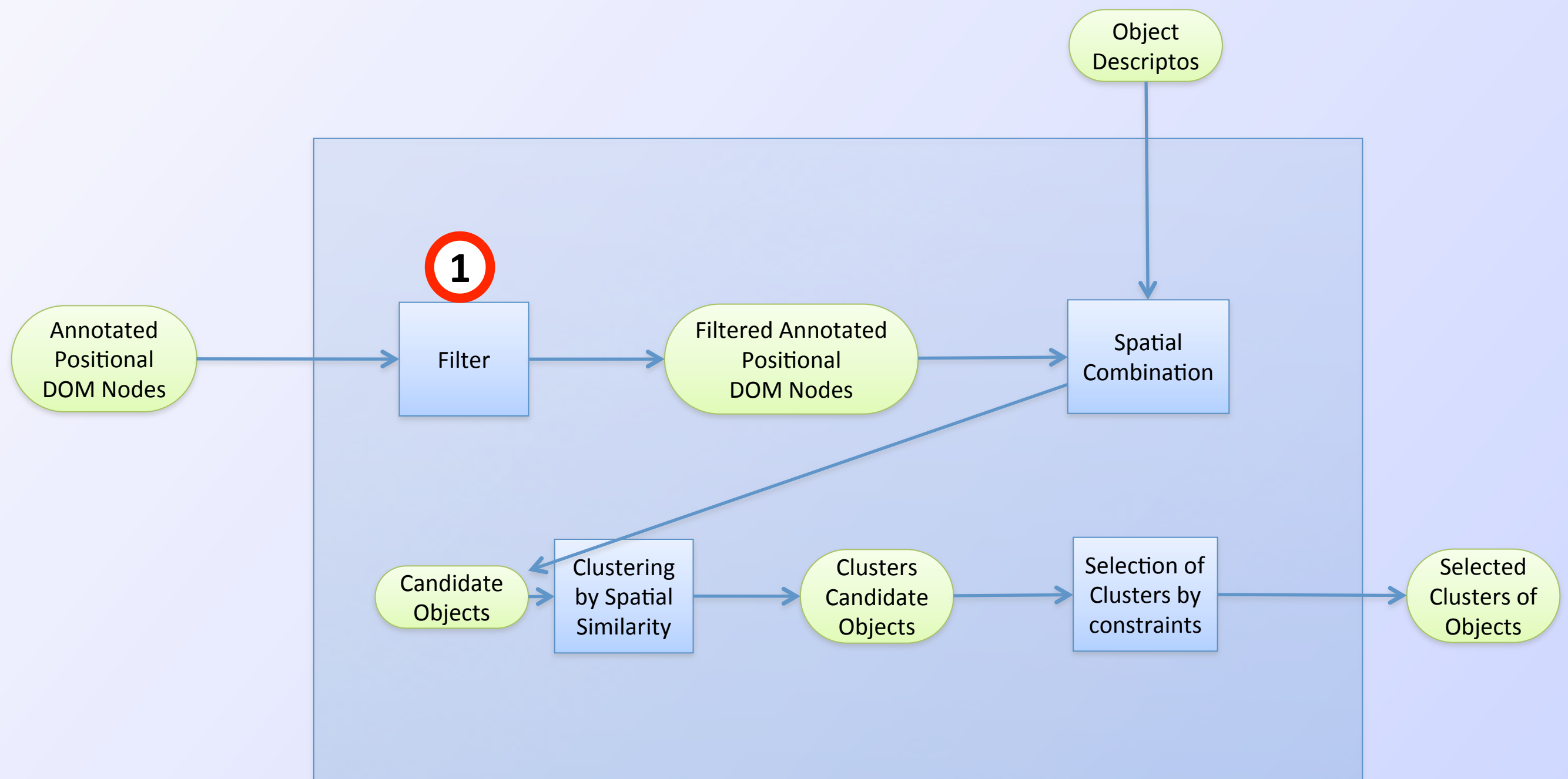


# Recognizer



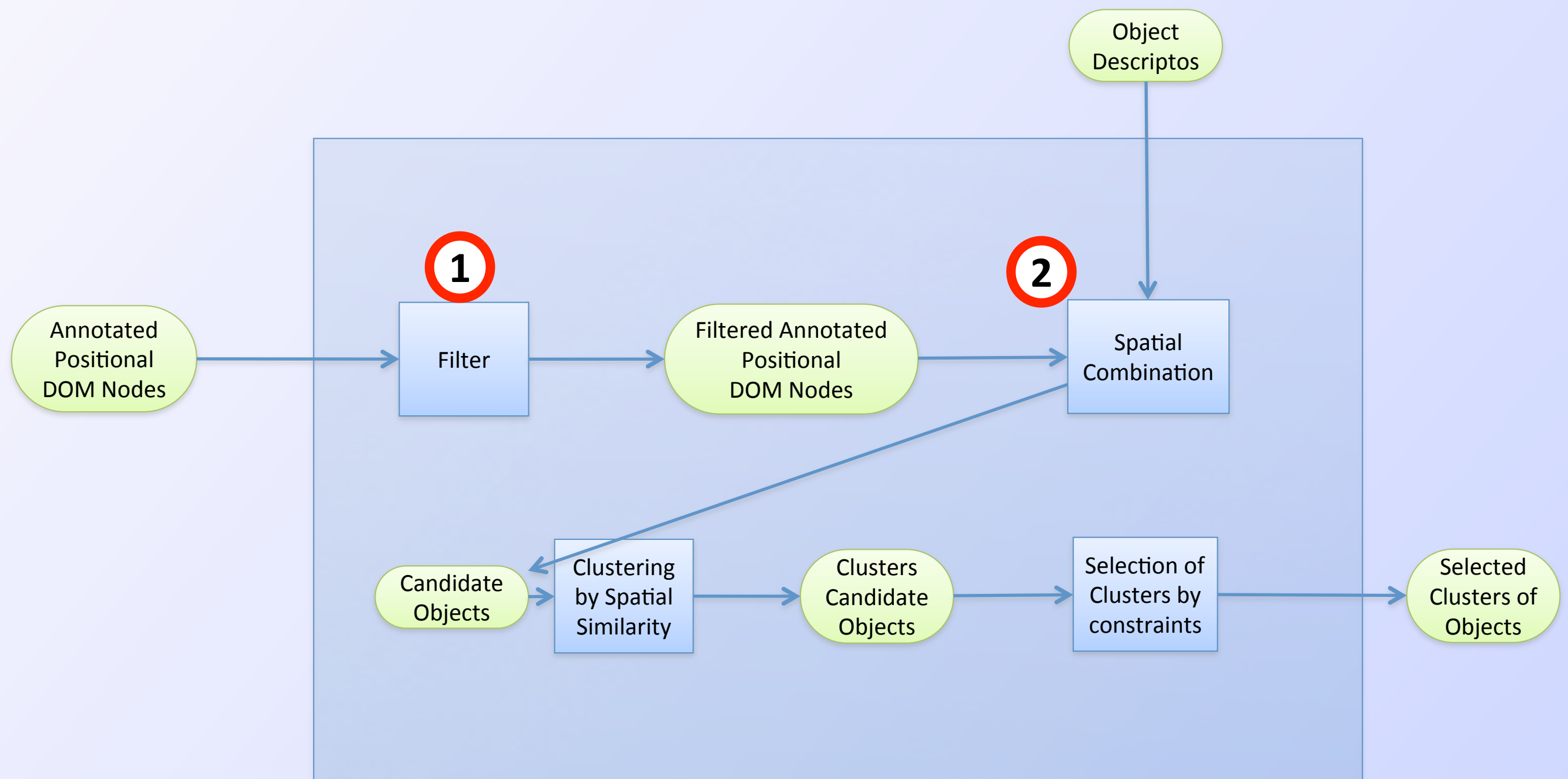


# Recognizer



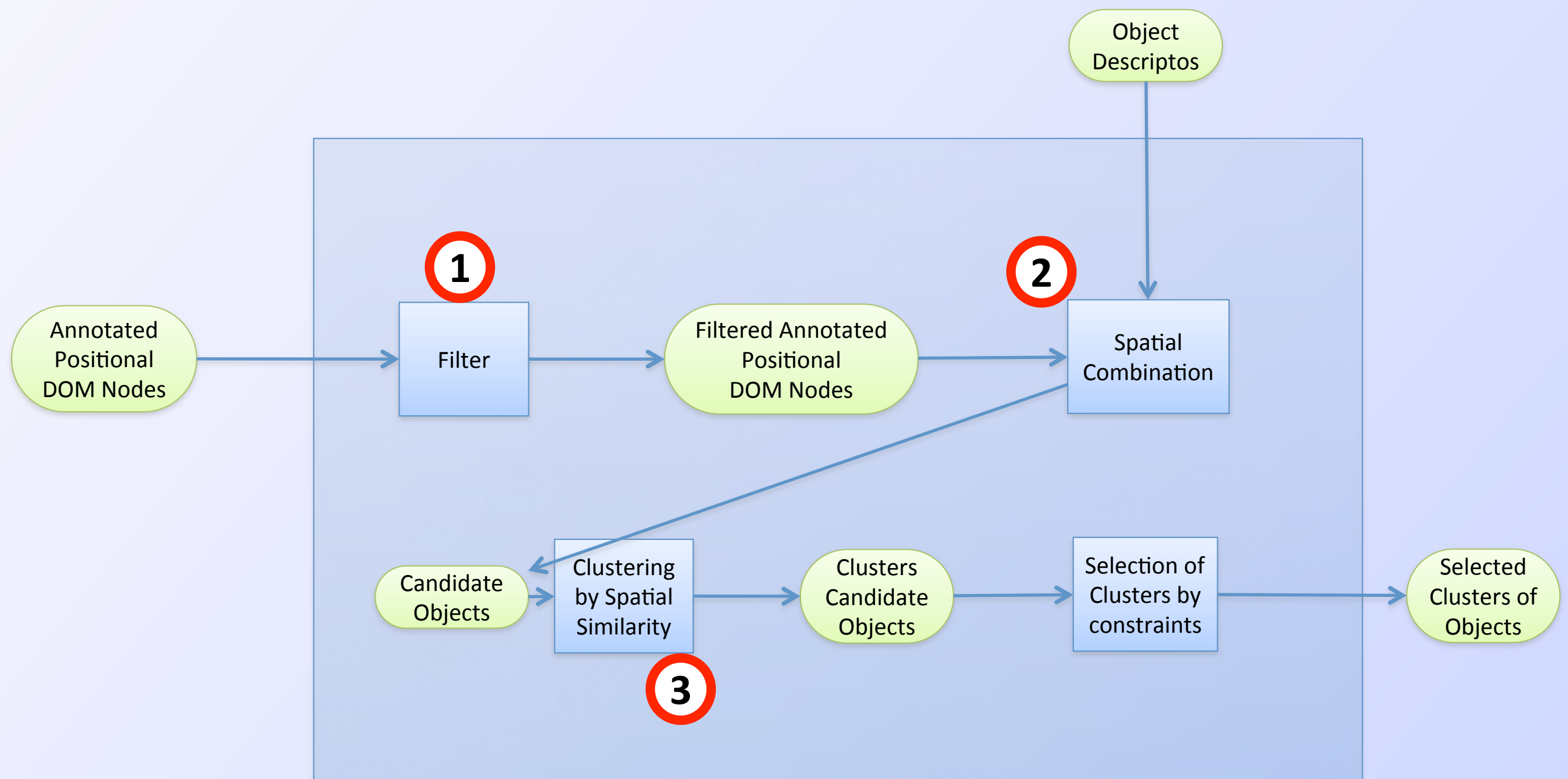


# Recognizer



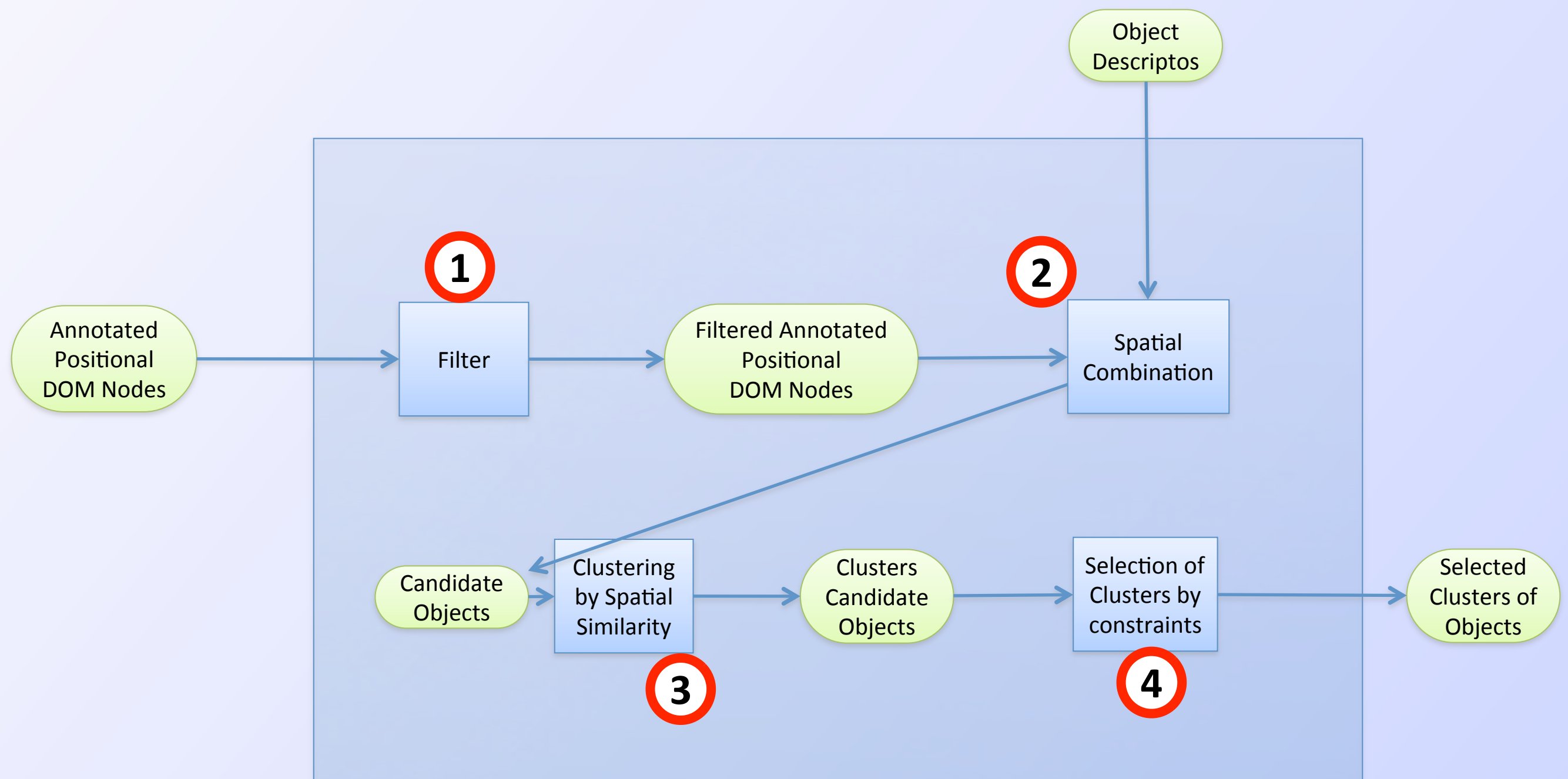


# Recognizer





# Recognizer





# Annotated Nodes

1

Output

idenziazioni Crea screenshot Calcola distanza Trova combinazione Crea Wrapper Esporta in XHTML

MatchingTable

Marca  
img  
Currency

Digital Cameras 2,493 Matches Home Cameras Digital Cameras Cameras

Panasonic GF3 Camera Check Out The New LUMIX® GF3 With Intelligent Auto Plus. Learn More! [www.panasonic.com/LUMIX](http://www.panasonic.com/LUMIX)

Canon Digital Camera Sale Get Great Savings on Canon Cameras. Once you know, you Newegg. [www.newegg.com](http://www.newegg.com) ★★★★★

Nikon Digital Camera Save 5% And More On Select Digital Cameras, Plus Free Shipping. [www.bestbuy.com](http://www.bestbuy.com)

Featured Brand  
See Products from:

SONY

Welcome Sign In or Register

Grid View List View ZIP Code: Sort By:

Facebook 0 Favorites 0 Recently Viewed 0 Price Alerts Feedback What is this? | Hide

Price Range

Under \$109 (629)  
\$109 - \$199 (610)  
\$199 - \$353 (642)  
\$353+ (641)

Condition

New (~2345)  
Used (~56)  
Refurbished (~94)

Shipping

Free Shipping (~1104)

Power Score

7 to 10 (1877)  
5 to 7 (271)  
3 to 5 (33)  
Lower than 3 (4)

Featured Product

**Sony Cyber-shot DSC-HX100V Black Digital Camera**

DSCHX100VB - 16.2 Megapixel - 4.80 mm-144 mm - 3" LCD - 30x Optical Zoom - Optical IS - 4608 x...  
Renew your passion for photography with the Cyber-shot DSC-HX100V digital camera. A Carl Zeiss... [More](#)

Add to Favorites Add to Compare

As low as **\$428.95**  
COMPARE NOW  
Price Range **\$428.95 - 449.99**  
Sold By 10 Merchants

**Canon PowerShot SX150 IS 14.1 Megapixel Compact Camera - Black**

5664B001 - 3" LCD - 12x Optical Zoom - Optical IS - 4320 x 3240 Image - 1280 x 720 Video ~...  
Discover how simple and easy it is to be creative and shoot memorable photos and videos of your... [More](#)

Add to Favorites Add to Compare

As low as **\$228.88**  
COMPARE NOW  
Price Range **\$228.88 - 274.99**  
Sold By 15 Merchants

**Nikon COOLPIX S8200 Black Digital Camera**

26288 - 16.1 MP, 14x Opt, SD/SDHC/SDXC Card Slot  
Nikon's COOLPIX S8200 Digital Camera puts ease-of-use and a remarkable 16.1Mp CMOS sensor at... [More](#)

Add to Favorites Add to Compare

As low as **\$326.95**  
COMPARE NOW



# Annotated Nodes

1

Annotation of nodes containing camera's brands

Output

Identificazioni Crea screenshot Calcola distanza Trova combinazione Crea Wrapper Esporta in XHTML

Home Cameras Digital Cameras Cameras

New LUMIX® GF3 With Intelligent Auto Plus. Learn More! [www.panasonic.com/LUMIX](http://www.panasonic.com/LUMIX)

Savings on Canon Cameras. Once you know, you Newegg. [www.newegg.com](http://www.newegg.com) ★★★★★

Nikon Digital Camera Save 5% And More On Select Digital Cameras, Plus Free Shipping. [www.bestbuy.com](http://www.bestbuy.com)

Sponsored Links

Featured Brand

See Products from:

SONY

Welcome [Sign In](#) or [Register](#)

Grid View List View ZIP Code: Sort By:

Facebook 0 Favorites 0 Recently Viewed 0 Price Alerts Feedback [What is this?](#) [Hide](#)

Price Range

Under \$109 (629)

\$109 - \$199 (610)

\$199 - \$353 (642)

\$353+ (641)

Condition

New (~2345)

Used (~56)

Refurbished (~94)

Shipping

Free Shipping (~1104)

Power Score

7 to 10 (1877)

5 to 7 (271)

3 to 5 (33)

Lower than 3 (4)

Featured Product

**Sony Cyber-shot DSC-HX100V Black Digital Camera**

As low as **\$428.95**

COMPARE NOW

DSCHX100VB - 16.2 Megapixel - 4.80 mm-144 mm - 3" LCD - 30x Optical Zoom - Optical IS - 4608 x...

Renew your passion for photography with the Cyber-shot DSC-HX100V digital camera. A Carl Zeiss... [More](#)

[Add to Favorites](#) [Add to Compare](#)

Price Range **\$428.95 - 449.99**

Sold By 10 Merchants

**Canon PowerShot SX150 IS 14.1 Megapixel Compact Camera - Black**

As low as **\$228.88**

COMPARE NOW

5664B001 - 3" LCD - 12x Optical Zoom - Optical IS - 4320 x 3240 Image - 1280 x 720 Video ~...

Discover how simple and easy it is to be creative and shoot memorable photos and videos of your... [More](#)

[Add to Favorites](#) [Add to Compare](#)

Price Range **\$228.88 - 274.99**

Sold By 15 Merchants

**Nikon COOLPIX S8200 Black Digital Camera**

As low as **\$326.95**

COMPARE NOW

26288 - 16.1 MP, 14x Opt, SD/SDHC/SDXC Card Slot

Nikon's COOLPIX S8200 Digital Camera puts ease-of-use and a remarkable 16.1Mp CMOS sensor at... [More](#)

[Add to Favorites](#) [Add to Compare](#)



# Spatial Combination

2

Output

Seleziona evidenziazioni Crea screenshot Calcola distanza Trova combinazione Crea Wrapper Esporta in XHTML

◀ Risultati Combinazione (337) ▶

**Digital Cameras** 2,493 Matches Home Cameras Digital Cameras Cameras

[Panasonic GF3 Camera](#) Check Out The New LUMIX® GF3 With Intelligent Auto Plus. Learn More! [www.panasonic.com/LUMIX](http://www.panasonic.com/LUMIX)

[Canon Digital Camera Sale](#) Get Great Savings on Canon Cameras. Once you know, you Newegg. [www.newegg.com](http://www.newegg.com) ★★★★★

[Nikon Digital Camera](#) Save 5% And More On Select Digital Cameras, Plus Free Shipping. [www.bestbuy.com](http://www.bestbuy.com)

**Featured Brand**  
See Products from:

**SONY**

Welcome [Sign in](#) or [Register](#)

Grid View List View ZIP Code:  Sort By:

Facebook 0 Favorites 0 Recently Viewed 0 Price Alerts Feedback [What is this?](#) | [Hide](#)

**Price Range**

- Under \$109 (629)
- \$109 - \$199 (610)
- \$199 - \$353 (642)
- \$353+ (641)

**Condition**

- New (~2345)
- Used (~56)
- Refurbished (~94)

**Shipping**

- Free Shipping (~1104)

**Power Score**

- 7 to 10 (1877)
- 5 to 7 (271)
- 3 to 5 (33)
- Lower than 3 (4)

**Featured Product**

**Sony Cyber-shot DSC-HX100V Black Digital Camera**

As low as **\$428.95**

COMPARE NOW

DSCHX100VB - 16.2 Megapixel - 4.80 mm-144 mm - 3" LCD - 30x Optical Zoom - Optical IS - 4608 x...  
Renew your passion for photography with the Cyber-shot DSC-HX100V digital camera. A Carl Zeiss... [More](#)

[Add to Favorites](#) [Add to Compare](#)

**Canon PowerShot SX150 IS 14.1 Megapixel Compact Camera - Black**

As low as **\$228.88**

COMPARE NOW

5664B001 - 3" LCD - 12x Optical Zoom - Optical IS - 4320 x 3240 Image - 1280 x 720 Video -...  
Discover how simple and easy it is to be creative and shoot memorable photos and videos of your... [More](#)

[Add to Favorites](#) [Add to Compare](#)

**Nikon COOLPIX S8200 Black Digital Camera**

As low as **\$326.95**

COMPARE NOW

26288 - 16.1 MP, 14x Opt, SD/SDHC/SDXC Card Slot  
Nikon's COOLPIX S8200 Digital Camera puts ease-of-use and a remarkable 16.1MP CMOS sensor at... [More](#)

[Add to Favorites](#) [Add to Compare](#)



# Spatial Combination

2

Output

eventuazioni Crea screenshot Calcola distanza Trova combinazione W Crea Wrapper Esporta in XHTML

Risultati Combinazione (337)

fotocamera2(img: , marca: Canon, ...)

**Digital Cameras** 2,493 Matches Home Cameras Digital Cameras Cameras

[Panasonic GF3 Camera](#) Check Out The New LUMIX® GF3 With Intelligent Auto Plus. Learn More! [www.panasonic.com/LUMIX](http://www.panasonic.com/LUMIX)

[Canon Digital Camera Sale](#) Get Great Savings on Canon Cameras. Once you know, you Newegg. [www.newegg.com](http://www.newegg.com) ★★★★★

[Nikon Digital Camera](#) Save 5% And More On Select Digital Cameras, Plus Free Shipping. [www.bestbuy.com](http://www.bestbuy.com)

**Featured Brand**  
See Products from:

**SONY.**

Welcome [Sign In](#) or [Register](#)

Grid View List View ZIP Code: Sort By:

Facebook 0 Favorites 0 Recently Viewed 0 Price Alerts Feedback [What is this?](#) [Hide](#)

**Price Range**

- Under \$109 (629)
- \$109 - \$199 (610)
- \$199 - \$353 (642)
- \$353+ (641)

**Condition**

- New (~2345)
- Used (~56)
- Refurbished (~94)

**Shipping**

- Free Shipping (~1104)

**Power Score**

- 7 to 10 (1877)
- 5 to 7 (271)
- 3 to 5 (33)
- Lower than 3 (4)

**Featured Product**

**Sony Cyber-shot DSC-HX100V Black Digital Camera**

As low as **\$428.95**

COMPARE NOW

DSCHX100VB - 16.2 Megapixel - 4.80 mm-144 mm - 3" LCD - 30x Optical Zoom - Optical IS - 4608 x...  
Renew your passion for photography with the Cyber-shot DSC-HX100V digital camera.  
A Carl Zeiss... [More](#)

[Add to Favorites](#) [Add to Compare](#)

**Canon PowerShot SX150 IS 14.1 Megapixel Compact Camera - Black**

As low as **\$228.88**

COMPARE NOW

5664B001 - 3" LCD - 12x Optical Zoom - Optical IS - 4320 x 3240 Image - 1280 x 720 Video -...  
Discover how simple and easy it is to be creative and shoot memorable photos and videos of your... [More](#)

[Add to Favorites](#) [Add to Compare](#)

**Nikon COOLPIX S8200 Black Digital Camera**

As low as **\$326.95**

COMPARE NOW

26288 - 16.1 MP, 14x Opt, SD/SDHC/SDXC Card Slot  
Nikon's COOLPIX S8200 Digital Camera puts ease-of-use and a remarkable 16.1Mp CMOS sensor at... [More](#)

[Add to Favorites](#) [Add to Compare](#)



# Clustering by Spatial Similarity

3

Output

Sevidenziamenti Crea screenshot Calcola distanza Trova combinazione Crea Wrapper Esporta in XHTML

RisultatiClassific (433)

(2) Area fotocamera2(img: , ma  
(2) Area fotocamera2(img: , ma  
(2) Area fotocamera2(img: , ma  
(2) Area fotocamera2(img: , ma  
(2) Area fotocamera2(img: , ma  
(2) Area fotocamera2(img: , ma  
(2) Area fotocamera2(img: , ma  
(2) Area fotocamera2(img: , ma  
(1) Area fotocamera2(img: , ma  
(9) Area fotocamera2(img: , ma  
(5) Area fotocamera2(img: , ma  
(2) Area fotocamera2(img: , ma  
(2) Area fotocamera2(img: , ma  
(142) Area fotocamera2(img: , ma  
(24) Area fotocamera2(img: , ma  
(6) Area fotocamera2(img: , ma  
(36) Area fotocamera2(img: , ma  
(95) Area fotocamera2(img: , ma  
(16) Area fotocamera2(img: , ma  
(24) Area fotocamera2(img: , ma  
(2) Area fotocamera2(img: , ma  
(1) Area fotocamera2(img: , ma  
(2) Area fotocamera2(img: , ma  
(4) Area fotocamera2(img: , ma  
(2) Area fotocamera2(img: , ma  
(1) Area fotocamera2(img: , ma  
(1) Area fotocamera2(img: , ma  
(2) Area fotocamera2(img: , ma  
(6) Area fotocamera2(img: , ma  
(9) Area fotocamera2(img: , ma  
(1) Area fotocamera2(img: , ma  
(1) Area fotocamera2(img: , ma  
(2) Area fotocamera2(img: , ma  
(2) Area fotocamera2(img: , ma

7 to 10 (1877)  
5 to 7 (271)  
3 to 5 (33)  
Lower than 3 (4)

**Top Manufacturers**

- Canon (266)
- Nikon (242)
- Sony (334)
- Panasonic (238)
- Fujifilm (138)
- Kodak (186)
- Olympus (261)
- More Brands +

**Camera Type**

- Digital SLR (192)
- Mirrorless Camera (80)
- Digital Compact (71)

**Megapixels**

- 14MP or More (465)
- 10MP to 13MP (800)
- 7MP to 9MP (354)
- 4MP to 6MP (222)
- 3MP or Less (23)

**Optical Zoom**

- 11x or More (126)
- 7x to 10x (159)
- 3x to 6x (1423)

**Nikon COOLPIX S8200 Black Digital Camera**  
26288 - 16.1 MP, 14x Opt, SD/SDHC/SDXC Card Slot  
Nikon's COOLPIX S8200 Digital Camera puts ease-of-use and a remarkable 16.1Mp CMOS sensor at... [More](#)  
Add to Favorites Add to Compare  
As low as **\$326.95**  
COMPARE NOW  
Price Range **\$326.95 - 329.99**  
Sold By 11 Merchants

**Nikon Coolpix P7100 Black Digital Camera**  
26286 - 10.1 MP, 7.1x Opt, SD/SDHC/SDXC Card Slot  
The Nikon COOLPIX P7100 Digital Camera lets you take your creativity wherever you go. Whether... [More](#)  
Add to Favorites Add to Compare  
As low as **\$496.95**  
COMPARE NOW  
Price Range **\$496.95 - 499.99**  
Sold By 7 Merchants

**Sony Alpha NEX-5N Black Digital Camera Kit w/ 18-55mm Lens**  
NEX5NKB - 16.1 MP, 3.1x Opt, Memory Stick Pro DUO/SD/SDHC/SD... Card Slot  
It's all the performance of a DSLR in about half the size and half the weight. Get a perfectly... [More](#)  
Add to Favorites Add to Compare  
As low as **\$699.00**  
COMPARE NOW  
Price Range **\$699.00 - 699.99**  
Sold By 7 Merchants

**Nikon COOLPIX S3100 Silver Digital Camera**  
26262 - 14MP, 5x Opt, SD Card Slot  
Slip this ultra-slim, ultra-light COOLPIX S3100 camera into your pocket and you'll never miss... [More](#)  
Add to Favorites Add to Compare  
As low as **\$109.99**  
COMPARE NOW  
Price Range **\$109.99 - 139.99**



# Clustering by Spatial Similarity

3

Output

Idenziamenti Crea screenshot Calcola distanza Trova combinazione W Crea Wrapper Esporta in XHTML

RisultatiClassific (430)

Digital Cameras 2,493 Matches Home Cameras Digital Cameras Cameras

Panasonic GF3 Camera Check Out The New LUMIX® GF3 With Intelligent Auto Plus. Learn More! [www.panasonic.com/LUMIX](http://www.panasonic.com/LUMIX)

Canon Digital Camera Sale Get Great Savings on Canon Cameras. Once you know, you Newegg. [www.newegg.com](http://www.newegg.com) ★★★★★

Nikon Digital Camera Save 5% And More On Select Digital Cameras, Plus Free Shipping. [www.bestbuy.com](http://www.bestbuy.com)

Featured Brand

See Products from:

SONY

Welcome [Sign In](#) or [Register](#)

Grid View List View ZIP Code: Sort By:

Facebook 0 Favorites 0 Recently Viewed 0 Price Alerts Feedback [What is this?](#) | [Hide](#)

Price Range

Under \$109 (629)

\$109 - \$199 (610)

\$199 - \$353 (642)

\$353+ (641)

Condition

New (~2345)

Used (~56)

Refurbished (~94)

Shipping

Free Shipping (~1104)

Power Score

7 to 10 (1877)

5 to 7 (271)

3 to 5 (33)

Lower than 3 (4)

Featured Product

**Sony Cyber-shot DSC-HX100V Black Digital Camera**

As low as **\$428.95**

COMPARE NOW

DSCHX100VB - 16.2 Megapixel - 4.80 mm-144 mm - 3" LCD - 30x Optical Zoom - Optical IS - 4608 x...  
Renew your passion for photography with the Cyber-shot DSC-HX100V digital camera.  
A Carl Zeiss... [More](#)

Add to Favorites Add to Compare

Price Range **\$428.95 - 449.99**  
Sold By 10 Merchants

**Canon PowerShot SX150 IS 14.1 Megapixel Compact Camera - Black**

As low as **\$228.88**

COMPARE NOW

5664B001 - 3" LCD - 12x Optical Zoom - Optical IS - 4320 x 3240 Image - 1280 x 720 Video -...  
Discover how simple and easy it is to be creative and shoot memorable photos and videos of your... [More](#)

Add to Favorites Add to Compare

Price Range **\$228.88 - 274.99**  
Sold By 15 Merchants

**Nikon COOLPIX S8200 Black Digital Camera**

As low as **\$326.95**

COMPARE NOW

26288 - 16.1 MP, 14x Opt, SD/SDHC/SDXC Card Slot  
Nikon's COOLPIX S8200 Digital Camera puts ease-of-use and a remarkable 16.1Mp CMOS sensor at... [More](#)



# Cluster Selection by Constraints

4

Output

Ala evidenziazioni Crea screenshot Calcola distanza Trova combinazione W Crea Wrapper Esporta in XHTML

RisultatiClassific (430)

**Digital Cameras** 2,493 Matches Home Cameras Digital Cameras Cameras

[Panasonic GF3 Camera](#) Check Out The New LUMIX® GF3 With Intelligent Auto Plus. Learn More! [www.panasonic.com/LUMIX](http://www.panasonic.com/LUMIX)

[Canon Digital Camera Sale](#) Get Great Savings on Canon Cameras. Once you know, you Newegg. [www.newegg.com](http://www.newegg.com) ★★★★★

[Nikon Digital Camera](#) Save 5% And More On Select Digital Cameras, Plus Free Shipping. [www.bestbuy.com](http://www.bestbuy.com)

**Featured Brand**  
See Products from:

**SONY.**

Welcome [Sign In](#) or [Register](#)

Grid View List View ZIP Code: Sort By:

Facebook 0 Favorites 0 Recently Viewed 0 Price Alerts Feedback [What is this?](#) | [Hide](#)

**Price Range**

- Under \$109 (629)
- \$109 - \$199 (610)
- \$199 - \$353 (642)
- \$353+ (641)

**Condition**

- New (~2345)
- Used (~56)
- Refurbished (~94)

**Shipping**

- Free Shipping (~1104)

**Power Score**

- 7 to 10 (1877)
- 5 to 7 (271)
- 3 to 5 (33)
- Lower than 3 (4)

**Featured Product**  
**Sony Cyber-shot DSC-HX100V Black Digital Camera**  
As low as **\$428.95**  
COMPARE NOW  
DSCHX100VB - 16.2 Megapixel - 4.80 mm-144 mm - 3" LCD - 30x Optical Zoom - Optical IS - 4608 x...  
Renew your passion for photography with the Cyber-shot DSC-HX100V digital camera. A Carl Zeiss... [More](#)  
Add to Favorites Add to Compare  
Price Range **\$428.95 - 449.99**  
Sold By 10 Merchants

**Canon PowerShot SX150 IS 14.1 Megapixel Compact Camera - Black**  
As low as **\$228.88**  
COMPARE NOW  
5664B001 - 3" LCD - 12x Optical Zoom - Optical IS - 4320 x 3240 Image - 1280 x 720 Video ~...  
Discover how simple and easy it is to be creative and shoot memorable photos and videos of your... [More](#)  
Add to Favorites Add to Compare  
Price Range **\$228.88 - 274.99**  
Sold By 15 Merchants

**Nikon COOLPIX S8200 Black Digital Camera**  
As low as **\$326.95**  
COMPARE NOW  
26288 - 16.1 MP, 14x Opt, SD/SDHC/SDXC Card Slot  
Nikon's COOLPIX S8200 Digital Camera puts ease-of-use and a remarkable 16.1Mp CMOS sensor at... [More](#)



# Cluster Selection by Constraints

4

Output

alla evidenziazioni Crea screenshot Calcola distanza Trova combinazione W Crea Wrapper Esporta in XHTML

**Digital Cameras** 2,493 Matches Home Cameras Digital Cameras Cameras

[Panasonic GF3 Camera](#) Check Out The New LUMIX® GF3 With Intelligent Auto Plus. Learn More! [www.panasonic.com/LUMIX](http://www.panasonic.com/LUMIX)

[Canon Digital Camera Sale](#) Get Great Savings on Canon Cameras. Once you know, you Newegg. [www.newegg.com](http://www.newegg.com) ★★★★★

[Nikon Digital Camera](#) Save 5% And More On Select Digital Cameras, Plus Free Shipping. [www.bestbuy.com](http://www.bestbuy.com)

**Featured Brand**  
See Products from:

**SONY.**

Welcome [Sign In](#) or [Register](#)

Grid View List View ZIP Code: Sort By:

Facebook 0 Favorites 0 Recently Viewed 0 Price Alerts Feedback [What is this?](#) | [Hide](#)

**Price Range**

- Under \$109 (629)
- \$109 - \$199 (610)
- \$199 - \$353 (642)
- \$353+ (641)

**Condition**

- New (~2345)
- Used (~56)
- Refurbished (~94)

**Shipping**

- Free Shipping (~1104)

**Power Score**

- 7 to 10 (1877)
- 5 to 7 (271)
- 3 to 5 (33)
- Lower than 3 (4)

**Featured Product**  
**Sony Cyber-shot DSC-HX100V Black Digital Camera**  
As low as **\$428.95**  
COMPARE NOW  
DSCHX100VB - 16.2 Megapixel - 4.80 mm-144 mm - 3" LCD - 30x Optical Zoom - Optical IS - 4608 x...  
Renew your passion for photography with the Cyber-shot DSC-HX100V digital camera. A Carl Zeiss... [More](#)  
Add to Favorites Add to Compare  
Price Range \$428.95 - 449.99 Sold By 10 Merchants

**Canon PowerShot SX150 IS 14.1 Megapixel Compact Camera - Black**  
As low as **\$228.88**  
COMPARE NOW  
5664B001 - 3" LCD - 12x Optical Zoom - Optical IS - 4320 x 3240 Image - 1280 x 720 Video -...  
Discover how simple and easy it is to be creative and shoot memorable photos and videos of your... [More](#)  
Add to Favorites Add to Compare  
Price Range \$228.88 - 274.99 Sold By 15 Merchants

**Nikon COOLPIX S8200 Black Digital Camera**  
As low as **\$326.95**  
COMPARE NOW  
26288 - 16.1 MP, 14x Opt, SD/SDHC/SDXC Card Slot  
Nikon's COOLPIX S8200 Digital Camera puts ease-of-use and a remarkable 16.1Mp CMOS sensor at... [More](#)



# Cluster Selection by Constraints

4

Output

alla evidenziazioni Crea screenshot Calcola distanza Trova combinazione Crea Wrapper Esporta in XHTML

RisultatiClassific (430)

(2) Area fotocamera2(img: , ma  
(2) Area fotocamera2(img: , ma  
(2) Area fotocamera2(img: , ma  
(2) Area fotocamera2(img: , ma  
(1) Area fotocamera2(img: , ma  
(1) Area fotocamera2(img: , ma  
(2) Area fotocamera2(img: , ma  
(3) Area fotocamera2(img: , ma  
(10) Are fotocamera2(img: , ma  
(3) Area fotocamera2(img: , ma  
(5) Area fotocamera2(img: , ma  
(2) Area fotocamera2(img: , ma  
(2) Area fotocamera2(img: , ma  
(2) Area fotocamera2(img: , ma  
(3) Area fotocamera2(img: , ma  
(6) Area fotocamera2(img: , ma  
(130) Ar fotocamera2(img: , ma  
(45) Are fotocamera2(img: , ma  
(23) Are fotocamera2(img: , ma  
(2) Area fotocamera2(img: , ma  
(1) Area fotocamera2(img: , ma  
(1) Area fotocamera2(img: , ma  
(4) Area fotocamera2(img: , ma  
(3) Area fotocamera2(img: , ma  
(11) Are fotocamera2(img: , ma  
(9) Area fotocamera2(img: , ma  
(9) Area fotocamera2(img: , ma  
(2) Area fotocamera2(img: , ma  
(30) Are fotocamera2(img: , ma  
(2) Area fotocamera2(img: , ma  
(2) Area fotocamera2(img: , ma  
(2) Area fotocamera2(img: , ma  
(1) Area fotocamera2(img: , ma  
(36) Are fotocamera2(img: , ma  
(36) Are fotocamera2(img: , ma  
(132) Ar fotocamera2(img: , ma  
(88) Are fotocamera2(img: , ma

## Digital Cameras 2,493 Matches

Home Cameras Digital Cameras Cameras

[Panasonic GF3 Camera](#) Check Out The New LUMIX® GF3 With Intelligent Auto Plus. Learn More! [www.panasonic.com/LUMIX](http://www.panasonic.com/LUMIX)

[Canon Digital Camera Sale](#) Get Great Savings on Canon Cameras. Once you know, you Newegg. [www.newegg.com](http://www.newegg.com) ★★★★★

[Nikon Digital Camera](#) Save 5% And More On Select Digital Cameras, Plus Free Shipping. [www.bestbuy.com](http://www.bestbuy.com)

**Featured Brand**  
See Products from:

**SONY.**

Welcome [Sign In](#) or [Register](#)

Grid View List View ZIP Code: Sort By:

Facebook 0 Favorites 0 Recently Viewed 0 Price Alerts Feedback [What is this?](#) | [Hide](#)

**Price Range**

- Under \$109 (629)
- \$109 - \$199 (610)
- \$199 - \$353 (642)
- \$353+ (641)

**Condition**

- New (~2345)
- Used (~56)
- Refurbished (~94)

**Shipping**

- Free Shipping (~1104)

**Power Score**

- 7 to 10 (1877)
- 5 to 7 (271)
- 3 to 5 (33)
- Lower than 3 (4)

**Featured Product**  
**Sony Cyber-shot DSC-HX100V Black Digital Camera**  
As low as **\$428.95**  
COMPARE NOW  
Price Range **\$428.95 - 449.99**  
Sold By 10 Merchants

DSCHX100VB - 16.2 Megapixel - 4.80 mm-144 mm - 3" LCD - 30x Optical Zoom - Optical IS - 4608 x...  
Renew your passion for photography with the Cyber-shot DSC-HX100V digital camera. A Carl Zeiss... [More](#)  
[Add to Favorites](#) [Add to Compare](#)

**Canon PowerShot SX140 IS 14.1 Megapixel Compact Camera - Black**  
As low as **\$228.88**  
COMPARE NOW  
Price Range **\$228.88 - 274.99**  
Sold By 15 Merchants

5664B001 - 3" LCD - 12x Optical Zoom - Optical IS - 4320 x 3240 Image - 1280 x 720 Video -...  
Discover how simple and easy it is to be creative and shoot memorable photos and videos of your... [More](#)  
[Add to Favorites](#) [Add to Compare](#)

**Nikon COOLPIX S8200 Black Digital Camera**  
As low as **\$326.95**  
COMPARE NOW

26288 - 16.1 MP, 14x Opt. SD/SDHC/SDXC Card Slot  
Nikon's COOLPIX S8200 Digital Camera puts ease-of-use and a remarkable 16.1Mp CMOS sensor at... [More](#)



# Result: Selected Cluster

**Digital Cameras** 2,493 Matches | Home | Cameras | Digital Cameras | Cameras

---


[Panasonic GF3 Camera](#) Check Out The New LUMIX® GF3 With Intelligent Auto Plus. Learn More! [www.panasonic.com/LUMIX](http://www.panasonic.com/LUMIX)

[Canon Digital Camera Sale](#) Get Great Savings on Canon Cameras. Once you know, you Newegg. [www.newegg.com](http://www.newegg.com) ★★★★★

[Nikon Digital Camera](#) Save 5% And More On Select Digital Cameras, Plus Free Shipping. [www.bestbuy.com](http://www.bestbuy.com)

---

**Featured Brand**  
See Products from:



Welcome [Sign In](#) or [Register](#)

Grid View List View ZIP Code:  Sort By:

---

Facebook 0 Favorites 0 Recently Viewed 0 Price Alerts Feedback [What is this?](#) Hide

---

**Price Range**

- Under \$109 (629)
- \$109 - \$199 (610)
- \$199 - \$353 (642)
- \$353+ (641)

**Condition**

- New (~2345)
- Used (~56)
- Refurbished (~94)

**Shipping**


- Free Shipping (~1104)

**Power Score**

- 7 to 10 (1877)
- 5 to 7 (271)
- 3 to 5 (33)
- Lower than 3 (4)

---

**Sony Cyber-shot DSC-HX100V Black Digital Camera**




DSCHX100VB - 16.2 Megapixel - 4.80 mm-144 mm - 3" LCD - 30x Optical Zoom - Optical IS - 4608 x...  
Renew your passion for photography with the Cyber-shot DSC-HX100V digital camera.  
A Carl Zeiss... [More](#)

Add to Favorites Add to Compare

**As low as \$428.95**  
[COMPARE NOW](#)  
Price Range \$428.95 - 449.99  
Sold By 10 Merchants

---

**Canon PowerShot SX150 IS 14.1 Megapixel Compact Camera - Black**




5664B001 - 3" LCD - 12x Optical Zoom - Optical IS - 4320 x 3240 Image - 1280 x 720 Video -...  
Discover how simple and easy it is to be creative and shoot memorable photos and videos of your... [More](#)

Add to Favorites Add to Compare

**As low as \$228.88**  
[COMPARE NOW](#)  
Price Range \$228.88 - 274.99  
Sold By 15 Merchants

---

**Nikon COOLPIX S8200 Black Digital Camera**



26288 - 16.1 MP, 14x Opt, SD/SDHC/SDXC Card Slot  
Nikon's COOLPIX S8200 Digital Camera puts ease-of-use and a remarkable 16.1MP CMOS sensor at... [More](#)

**As low as \$326.95**



# Wrapper Generation

Input  
page

**ebay Classico** Benvenuto! Accedi o registrati

MOSTRA LE CATEGORIE EBAY CLASSICO EBAY ANNUNCI AUTO CASE

**SOLO fino al 30 settembre raddoppi i vantaggi!**

fotocamera Fotocamere digitali **Cerca** Avanzata

Ricerche correlate: Naviga in questa categoria

**10.395 risultati per fotocamera** Salva la ricerca

In Fotocamere Digitali

**Marca**

- ☐ Canon (1.788)
- ☐ Nikon (1.701)
- ☐ Samsung (1.454)
- ☐ Sony (1.232)
- [Scegli altro...](#)

**Modello**

- ☐ EOS 5D (103)
- ☐ D90 (96)
- ☐ CYBER-SHOT DSC-T100 (26)
- ☐ MICRO (23)
- [Scegli altro...](#)

**Condizione**

- ☐ Nuovo (9.787)
- ☐ Nuovo: altro (vedi dettagli) (45)
- ☐ Ricondizionato dal produttore (1)
- ☐ Ricondizionato dal venditore (3)
- [Scegli altro...](#)

**Prezzo**

€  a €

**Tipo di fotocamera digitale**

- ☐ Zoom ottico
- ☐ Formato scheda di memoria
- ☐ Venditore

**Tutti gli oggetti** Solo Aste online Compralo Subito








Vedi come: ☐ ☐ ☐

Mostra prima: Inserzioni più rilevanti

Pagina 1 di 434

Spedizione a ITA

Oggetto In Vetrina

 <b>Fotocamera Digitale V527 FUCSIA</b> 12 MP 8x Zoom LCD 2,7" <i>Compralo Subito</i> EUR 47,00	 <b>Fotocamera Digitale V527 NERA</b> 12 MP 8x Zoom LCD 2,7" <i>Compralo Subito</i> EUR 47,00	 <b>Fotocamera Digitale SUBACQUEA</b> fino a 3 MT 12 MegaPixel <i>Compralo Subito</i> EUR 52,90	 <b>FOTOCAMERA DIGITALE ★ KODAK M575 14 MPX ★ EASYSHARE ZOO...</b> <i>Compralo Subito</i> EUR 85,90 <i>Proposta d'acquisto</i>
 <b>Fotocamera Digitale SUBACQUEA x bambini</b> fino a 3 Metri <i>Compralo Subito</i> EUR 47,00	 <b>Fotocamera digitale Panasonic DMC fz30</b> <i>Compralo Subito</i> EUR 100,00	 <b>FOTOCAMERA DIGITALE x bambini</b> 8 Mega Pixel rosa V8690 <i>Compralo Subito</i> EUR 50,00	 <b>FOTOCAMERA DIGITALE BENQ DCE 300</b> <i>Compralo Subito</i> EUR 40,00



# Wrapper Generation

Recognized  
object

**ebay Classico** Benvenuto! Accedi o registrati

MOSTRA LE CATEGORIE EBAY CLASSICO EBAY ANNUNCI AUTO CASE OFFERTE

**SOLO fino al 30 settembre raddoppi i vantaggi!**

AdChoice

fotocamera Fotocamere digitali **Cerca** Avanzata ☐ Includi la descrizione

Ricerche correlate: Naviga in questa categoria

**10.395 risultati per fotocamera** Salva la ricerca

In Fotocamere Digitali

**Marca**

- ☐ Canon (1.788)
- ☐ Nikon (1.701)
- ☐ Samsung (1.454)
- ☐ Sony (1.232)
- [Scegli altro...](#)

**Modello**

- ☐ EOS 5D (103)
- ☐ D90 (96)
- ☐ CYBER-SHOT DSC-T100 (26)
- ☐ MICRO (23)
- [Scegli altro...](#)

**Condizione**

- ☐ Nuovo (9.787)
- ☐ Nuovo: altro (vedi dettagli) (45)
- ☐ Ricondizionato dal produttore (1)
- ☐ Ricondizionato dal venditore (3)
- [Scegli altro...](#)

**Prezzo**

€  a €

► Tipo di fotocamera digitale

► Zoom ottico

► Formato scheda di memoria

► Venditore

► Categoria

Tutti gli oggetti Solo Aste online Compralo Subito









Vedi come: ☐ ☐ ☐

Mostra prima: Inserzioni più rilevanti

Pagina 1 di 434

Spedizione a ITA

Oggetti in vetrina

 <p><b>Fotocamera Digitale V527 FUCSIA</b> 12 MP 8x Zoom LCD 2,7"</p> <p><b>Compralo Subito</b> EUR 47,00</p>	 <p><b>Fotocamera Digitale V527 NERA</b> 12 MP 8x Zoom LCD 2,7"</p> <p><b>Compralo Subito</b> EUR 47,00</p>	 <p><b>Fotocamera Digitale SUBACQUEA</b> fino a 3 MT 12 MegaPixel</p> <p><b>Compralo Subito</b> EUR 52,90</p>	 <p><b>FOTOCAMERA DIGITALE</b> ★ KODAK M575 14 MPX ★ EASYSHARE ZOO...</p> <p><b>Compralo Subito</b> EUR 85,90</p> <p><i>Proposta d'acquisto</i></p>
 <p><b>Fotocamera Digitale SUBACQUEA</b> x bambini fino a 3 Metri</p>	 <p><b>Fotocamera digitale Panasonic</b> DMC fz30</p>	 <p><b>FOTOCAMERA DIGITALE</b> x bambini 8 Mega Pixel rosa V8690</p>	 <p><b>FOTOCAMERA DIGITALE BENQ</b> DCE 300</p>



# Wrapper Generation

Recognized  
object

**ebay Classico** Benvenuto! Accedi o registrati

MOSTRA LE CATEGORIE EBAY CLASSICO EBAY ANNUNCI AUTO CASE

**SOLO fino al 30 settembre raddoppi i vantaggi!**

fotocamera Fotocamere digitali Cerca Avanzata

Ricerche correlate: Naviga in questa categoria

10.395 risultati per fotocamera Salva la ricerca

In Fotocamere Digitali

**Marca**

- ☐ Canon (1.788)
- ☐ Nikon (1.701)
- ☐ Samsung (1.454)
- ☐ Sony (1.232)
- [Scegli altro...](#)

**Modello**

- ☐ EOS 5D (103)
- ☐ D90 (96)
- ☐ CYBER-SHOT DSC-T100 (28)
- ☐ MICRO (23)
- [Scegli altro...](#)

**Condizione**

- ☐ Nuovo (9.787)
- ☐ Nuovo: altro (vedi dettagli) (45)
- ☐ Ricondizionato dal produttore (1)
- ☐ Ricondizionato dal venditore (3)
- [Scegli altro...](#)

**Prezzo**

€  a €

**Tipo di fotocamera digitale**

- ☐ Zoom ottico
- ☐ Formato scheda di memoria
- ☐ Venditore
- ☐ Categoria

Tutti gli oggetti Solo Aste online Compralo Subito

Vedi come: ☐ ☐ ☐

Mostra prima: Inserzioni più rilevanti

Pagina 1 di 434

**Oggetto in vetrina**

Image	Description	Price
	Fotocamera Digitale V527 FUCSIA 12 MP 8x Zoom LCD 2,7"	EUR 47,00
	Fotocamera Digitale V527 NERA 12 MP 8x Zoom LCD 2,7"	EUR 47,00
	Fotocamera Digitale SUBACQUEA fino a 3 MT 12 Megapixel	EUR 52,90
	FOTOCAMERA DIGITALE ★ KODAK M575 14 MPX ★ EASYSHARE ZOO...	EUR 85,90

Promuovi i tuoi oggetti

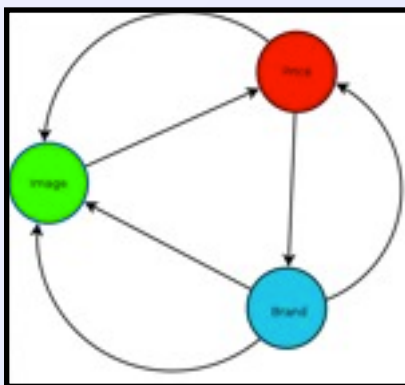
Image	Description
	Fotocamera Digitale SUBACQUEA x bambini fino a 3 Metri
	Fotocamera digitale Panasonic DMC fz30
	FOTOCAMERA DIGITALE x bambini 8 Mega Pixel rosa V8690
	FOTOCAMERA DIGITALE BENQ DCE 300

Some Brands  
are not  
recognized



# Wrapper Generation

Wrapper  
Generator  
by exploiting  
XPath



**ebay Classico** Benvenuto! Accedi o registrati

MOSTRA LE CATEGORIE EBAY CLASSICO EBAY ANNUNCI AUTO CASE

**SOLO fino al 30 settembre raddoppi i vantaggi!**

fotocamera Fotocamere digitali **Cerca** Avanzata

Ricerche correlate: Naviga in questa categoria

10.395 risultati per fotocamera Salva la ricerca

In Fotocamere Digitali

**Marca**

- ☐ Canon (1.788)
- ☐ Nikon (1.701)
- ☐ Samsung (1.454)
- ☐ Sony (1.232)
- [Scegli altro...](#)

**Modello**

- ☐ EOS 5D (103)
- ☐ D90 (96)
- ☐ CYBER-SHOT DSC-T100 (26)
- ☐ MICRO (23)
- [Scegli altro...](#)

**Condizione**

- ☐ Nuovo (9.787)
- ☐ Nuovo: altro (vedi dettagli) (45)
- ☐ Ricondizionato dal produttore (1)
- ☐ Ricondizionato dal venditore (3)
- [Scegli altro...](#)

**Prezzo**

€  a €

☒ Tipo di fotocamera digitale  
☒ Zoom ottico  
☒ Formato scheda di memoria  
☒ Venditore  
☒ Categoria

Tutti gli oggetti Solo Aste online Compralo Subito

Vedi come:

Mostra prima: Inserzioni più rilevanti Pagina 1 di 434

Oggetti in vetrina

 Fotocamera Digitale V527 FUCSIA 12 MP 8x Zoom LCD 2,7" <b>Compralo Subito</b> EUR 47,00	 Fotocamera Digitale V527 NERA 12 MP 8x Zoom LCD 2,7" <b>Compralo Subito</b> EUR 47,00	 Fotocamera Digitale SUBACQUEA fino a 3 MT 12 MegaPixel <b>Compralo Subito</b> EUR 52,90	 FOTOCAMERA DIGITALE ★ KODAK M575 14 MPX ★ EASYSHARE ZOO... <b>Compralo Subito</b> EUR 85,90 <i>Proposta d'acquisto</i>
 Fotocamera Digitale SUBACQUEA x bambini fino a 3 Metri Ingrandisci	 Fotocamera digitale Panasonic DMC fz30 Ingrandisci	 FOTOCAMERA DIGITALE x bambini 8 Mega Pixel rosa V8690	 FOTOCAMERA DIGITALE BENQ DCE 300

Promuovi i tuoi oggetti



# Experimental Evaluation

Web site	Object Descriptor	# Record for page	Record Type	# Founded Record for page	# Clusters
<a href="http://www.ebay.it">www.ebay.it</a>	1	54	Horizontal	54	3
<a href="http://www.unieuro.it">www.unieuro.it</a>	1	10	Horizontal	10	1
<a href="http://www.eplaza.it">www.eplaza.it</a>	1	10	Horizontal	10	1
<a href="http://www.eprice.it">www.eprice.it</a>	1	24	Horizontal	24	2
<a href="http://www.ritzcamera.com">www.ritzcamera.com</a>	1	25	Horizontal	25	1
<a href="http://www.buyacamera.co.uk">www.buyacamera.co.uk</a>	1	56	Horizontal	56	1
<a href="http://www.kelkoo.it">www.kelkoo.it</a>	2	24	Vertical	24	1
<a href="http://www.apmshop.it">www.apmshop.it</a>	1	10	Horizontal	10	1
<a href="http://www.gshopitalia.it">www.gshopitalia.it</a>	1	24	Grid	24	1
<a href="http://www.gcomm.it">www.gcomm.it</a>	1	20	Horizontal	20	1
<a href="http://www.photopix.it">www.photopix.it</a>	1	20	Horizontal	20	1
<a href="http://www.monimega.com">www.monimega.com</a>	2	20	Vertical	20	1
<a href="http://www.shop.it">www.shop.it</a>	1	20	Horizontal	20	1
<a href="http://www.creativestore.it">www.creativestore.it</a>	1	20	Horizontal	20	1
<a href="http://www.teds.com.au">www.teds.com.au</a>	1	7	Horizontal	7	1
<a href="http://www.visitek.ca">www.visitek.ca</a>	1	12	Horizontal	12	1
<a href="http://www.henrys.com">www.henrys.com</a>	1	12	Horizontal	12	1
<a href="http://www.dixons.co.uk">www.dixons.co.uk</a>	1	10	Horizontal	10	1
<a href="http://www.digitaldepot.com.uk">www.digitaldepot.com.uk</a>	1	10	Grid	10	2
<a href="http://www.pricegrabber.com">www.pricegrabber.com</a>	1	48	Horizontal	48	1
<a href="http://www.mediaworld.it">www.mediaworld.it</a>	1	13	Horizontal	13	1
<a href="http://www.fordigit.it">www.fordigit.it</a>	1	20	Horizontal	20	1
<a href="http://www.izideal.it">www.izideal.it</a>	2	26	Vertical	26	1
<a href="http://e-store.gecowebnet.eu">e-store.gecowebnet.eu</a>	1	20	Horizontal	20	1
<a href="http://www.shopz.it">www.shopz.it</a>	1	9	Horizontal	9	1

## Descriptor 1:

**Fotocamera ← #img, brand?, price;**

[#img IGE price]?,  
[#img IGE brand]?,  
[brand IGE price]?

## Descriptor 2:

**Fotocamera ← #img, brand?, price;**

[price N #img]?,  
[brand N #img]?



# Experimental Evaluation

Web Site	# Record for page	Record Type	# Founded Records for page	# Clusters
<a href="http://www.casaclick.it">www.casaclick.it</a>	15	Horizontal	15	2
<a href="http://www.habemuscasa.it">www.habemuscasa.it</a>	16	Horizontal	16	1
<a href="http://www.boglist.it">www.boglist.it</a>	8	Horizontal	8	1
<a href="http://www.shbarcelona.com">www.shbarcelona.com</a>	51	Horizontal	51	1
<a href="http://www.rpmrealty.com">www.rpmrealty.com</a>	8	Horizontal	8	1
<a href="http://www.nuroa.it">www.nuroa.it</a>	10	Horizontal	10	1
<a href="http://www.prendicasa.it">www.prendicasa.it</a>	20	Horizontal	20	1
<a href="http://www.trovocasa.corriere.it">www.trovocasa.corriere.it</a>	10	Horizontal	10	1
<a href="http://www.attico.it">www.attico.it</a>	20	Horizontal	20	1
<a href="http://www.compraticasa.it">www.compraticasa.it</a>	9	Horizontal	9	2
<a href="http://www.risorseimmobiliari.it">www.risorseimmobiliari.it</a>	35	Horizontal	35	1
<a href="http://www.apartmentsforsale.com">www.apartmentsforsale.com</a>	11	Horizontal	11	1
<a href="http://www.firstmallorca.com">www.firstmallorca.com</a>	10	Grid	10	1
<a href="http://www.lizsellslondon.com">www.lizsellslondon.com</a>	12	Horizontal	12	1
<a href="http://www.hiperprop.com">www.hiperprop.com</a>	10	Horizontal	10	1
<a href="http://www.idealista.it">www.idealista.it</a>	15	Horizontal	15	1
<a href="http://www.casa2home.it">www.casa2home.it</a>	8	Horizontal	8	1
<a href="http://www.intermediaimmobiliare.com">www.intermediaimmobiliare.com</a>	10	Horizontal	10	1
<a href="http://www.agenzia-puntocasa.it">www.agenzia-puntocasa.it</a>	10	Horizontal	10	1
<a href="http://www.immobiliariesole.net">www.immobiliariesole.net</a>	7	Horizontal	7	1

**casal** ← **#img, price?, area;**  
 [#img IGE price]?,  
 [area W #img]?,  
 [area E|W price]?



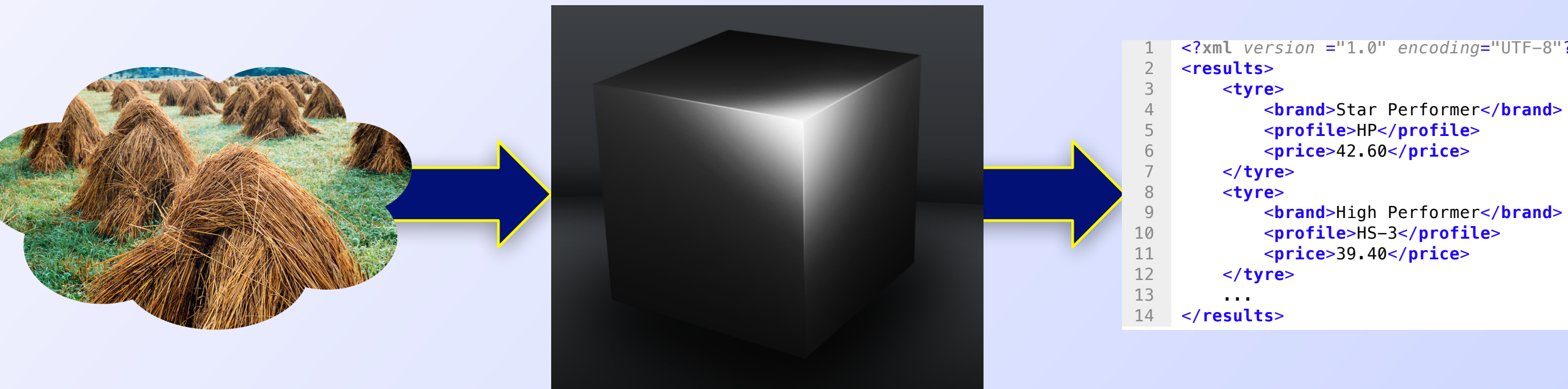
# DIADEM

## Fully-Automated Extraction



# Domain-Centric Data Extraction

- **Blackbox** that
  - turns **any** of the **thousands** of websites of a domain
  - into **structured data**





# Domain-Centric Data Extraction

- **Blackbox** that
  - turns **any** of the **thousands** of websites of a domain
  - into **structured data**





# DIADEM: Suffused by Knowledge

- Key insight **①**: all three types of knowledge
  - **every piece** of DIADEM is driven by knowledge
  - **exploration:** script/interaction knowledge
  - block/form/result page/description **analysis**
    - **all** combine **all three types**
  - **algorithms:**
    - search for “consistent” interpretation informed by domain knowledge
    - rather than uninformed as, e.g., in AutoWrappers



HTML page

Search

For saleTo rent

Research

Current values

Sold prices

e.g. Oxford or NW3

Min price

No min

Max price

No max

Type

Show all

Bedrooms

No min

Fewer options

Distance from location

This area only

Added

Any time

Sort by

Highest price

Keywords

Include

☒ New homes

☒ Shared ownership homes

☒ Retirement homes

☒ Under offer or sold STC

Search for sale

Browse by area

WHITE WALLS

people not just property

Home

About us

Sales

Lettings/Management

Free valuation

EPC

Careers

Links

Contact us

Property in Oxford search results

Save search

Go back to postcode area map

12345678910111213141516NEXT

1 - 10 of 153 properties

White Road, East Hendred

£1,499,950

New price

Located within East Hendred, this impressive fifteen bedroom property offers fantastic business and development potential. A five bedroom family house is accompanied by a purpose built ten bedroom guest house extension. The five [...] [Read more](#)

Bedrooms: 15 | Bathrooms: 13 | Area: East Hendred

View full details

360° virtual tour

Arrange a viewing

Map view

View all 10 photos

Bagley Wood Rd, Kennington OX1

£1,295,000

Sold

A substantial detached Edwardian family house with scope for updating and improvement, set within delightful gardens and wooded grounds extending to over five acres located on the edge of Bagley Wood. This fine [...] [Read more](#)

Bedrooms: 6 | Bathrooms: 3 | Area: Kennington

View full details

360° virtual tour

Arrange a viewing

Map view

View all 16 photos

White Road, East Hendred

£1,499,950

Arrange a viewing

Image gallery

360° virtual tour

Floorplan

Aerial view

Map view

Printable brochure

Email a friend

EPC

Free mortgage advice

Sales or Lettings: Sales

Property Type: House

Postcode Area: OX12

Bedrooms: 15

Bathrooms: 13

Furnishing: N/A

Area: East Hendred

Tenure: Freehold

Sq feet: 6648

Located within East Hendred, this impressive fifteen bedroom property offers fantastic business and development potential. A five bedroom family house is accompanied by a purpose built ten bedroom guest house extension.

The five bedroom family house benefits from generously proportioned living space with modern interiors, conservatory, dining room with feature bow window and stunning first floor views over substantial private gardens. This is further accompanied an attached recently purpose built AA 4\* rated ten bedroom guest house/bed and breakfast wing.

In addition, subject to planning, further development potential is possible.

The property is situated within its own grounds of approximately 2 acres. Local transport links are available nearby, providing access to Oxford and Wantage. Road users have easy access to the A34 (M40) towards London.

Coming soon to OX2

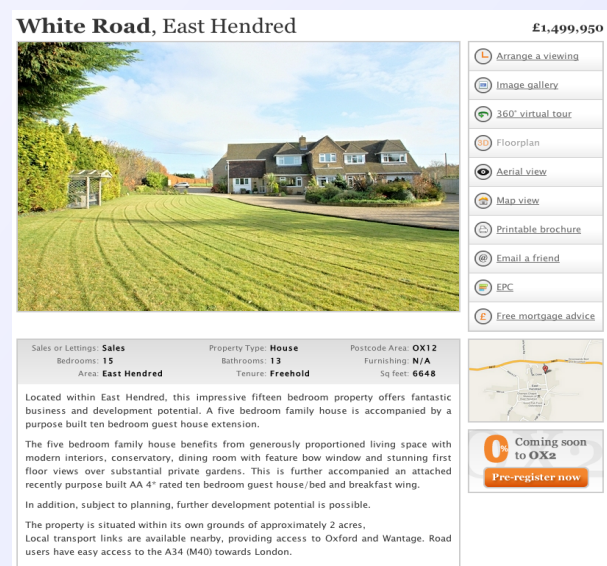
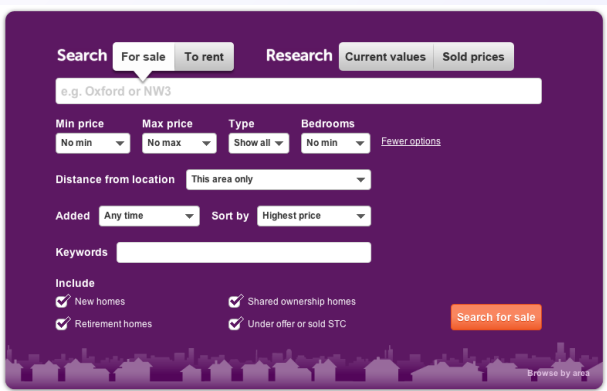
Pre-register now

149



# Detectors

HTML page



**Text**  
annotations

**HTML**  
rendering

**Document**  
structure

**Flat text**  
dependencies

**Image**  
features

**URL**  
analysis

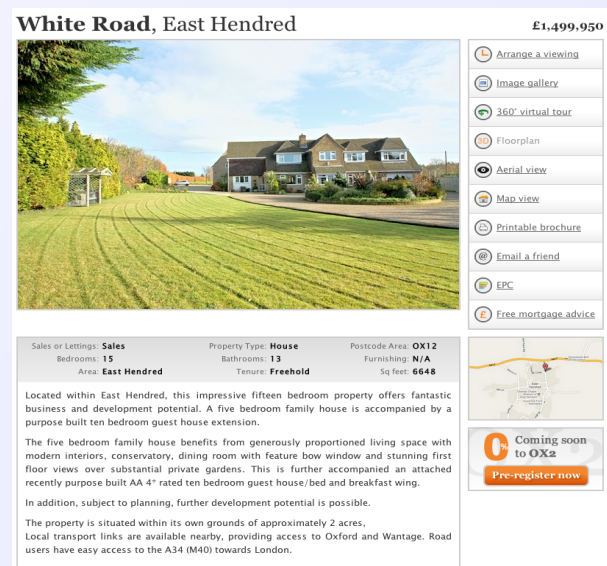
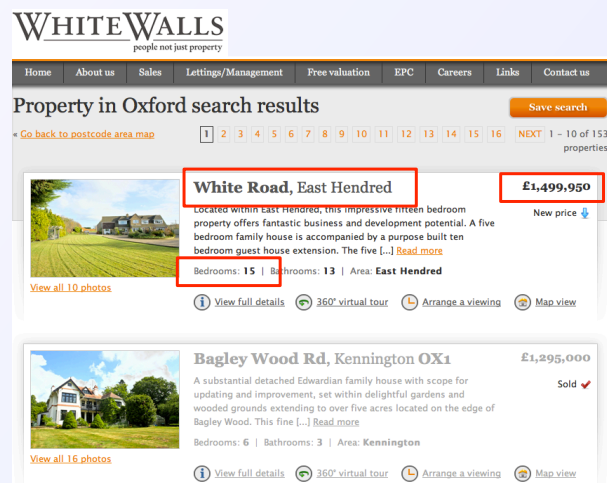
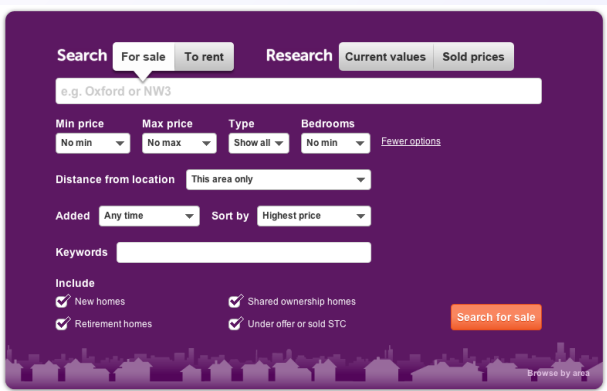
...

Observed phenomena



# Detectors

HTML page



**Text**  
annotations

**HTML**  
rendering

**Document**  
structure

**Flat text**  
dependencies

**Image**  
features

**URL**  
analysis

...

**Block**  
classification

**Table**  
understanding

**EPC chart**  
understanding

**Floor plan**  
understanding

Observed phenomena

Inferred phenomena



## Reasoners

## Detectors

**Text**  
annotations

**HTML**  
rendering

**Document**  
structure

**Flat text**  
dependencies

**Image**  
features

**URL**  
analysis

...

**Block**  
classification

**Table**  
understanding

**EPC chart**  
understanding

**Floor plan**  
understanding

**Form**  
Understanding

**Object**  
identification & alignment

Context-driven  
**block** analysis

Domain  
knowledge



Observed phenomena

Inferred phenomena

Phenomenology

Concepts



## Reasoners

## Detectors

### Text

annotations

### HTML

rendering

### Document

structure

### Flat text

dependencies

### Image

features

### URL

analysis

...

### Block

classification

### Table

understanding

### EPC chart

understanding

### Floor plan

understanding

### Form

Understanding

### Object

identification & alignment

Context-driven  
**block** analysis

Domain  
knowledge



**Domain  
specific**

Observed phenomena

Inferred phenomena

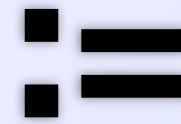
Phenomenology

Concepts

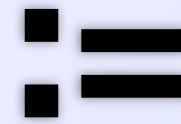






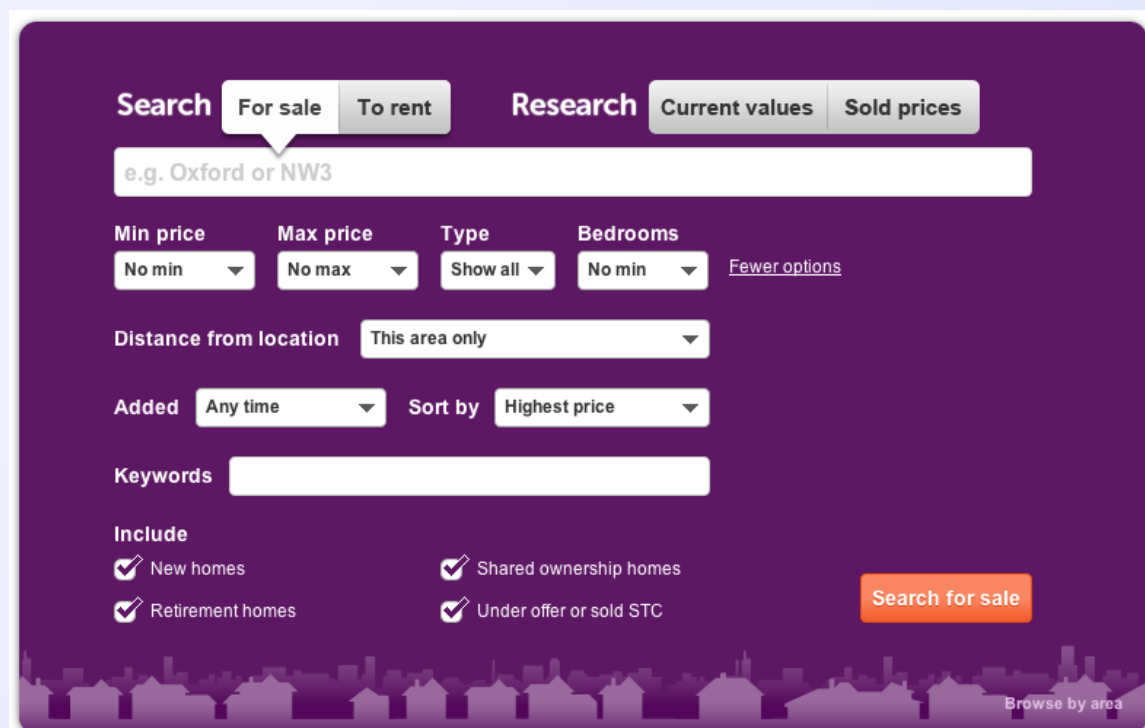
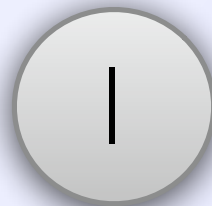






# Form

## Understanding & Filling



**Search** **For sale** **To rent** **Research** **Current values** **Sold prices**

e.g. Oxford or NW3

**Min price** **Max price** **Type** **Bedrooms**

No min ▼ No max ▼ Show all ▼ No min ▼ [Fewer options](#)

**Distance from location** This area only ▼

**Added** Any time ▼ **Sort by** Highest price ▼

**Keywords**

**Include**

☒ New homes ☒ Shared ownership homes

☒ Retirement homes ☒ Under offer or sold STC

**Search for sale**

Browse by area



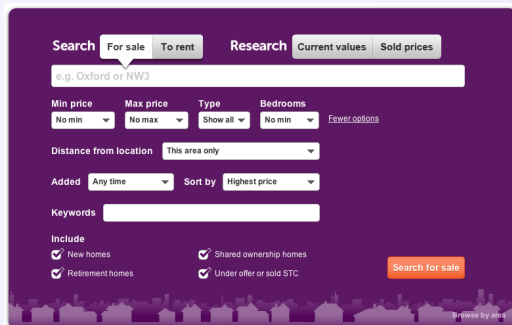


**Form**  
Understanding & Filling

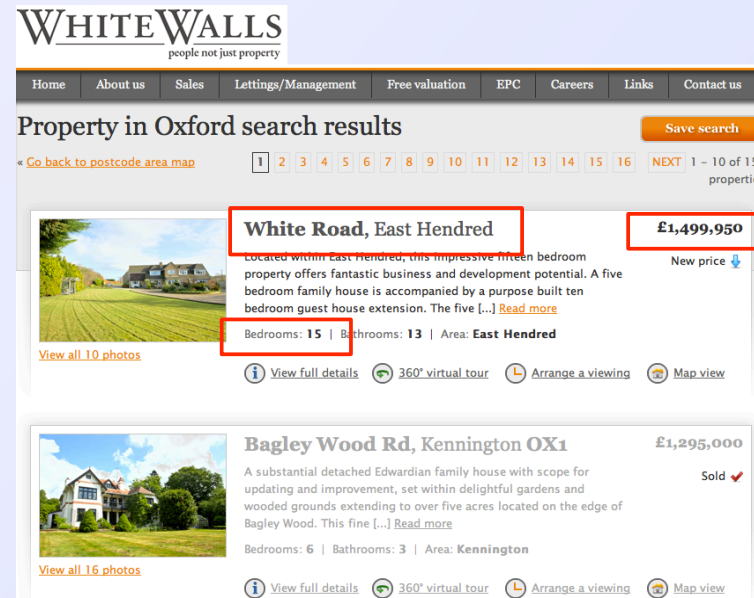
**Object**  
identification & alignment

1

2



## Result pages



**White Road, East Hendred** £1,499,950

Located within East Hendred, this impressive fifteen bedroom property offers fantastic business and development potential. A five bedroom family house is accompanied by a purpose built ten bedroom guest house extension. The five [...] [Read more](#)

Bedrooms: 15 | Bathrooms: 13 | Area: East Hendred

[View all 10 photos](#) [View full details](#) [360° virtual tour](#) [Arrange a viewing](#) [Map view](#)

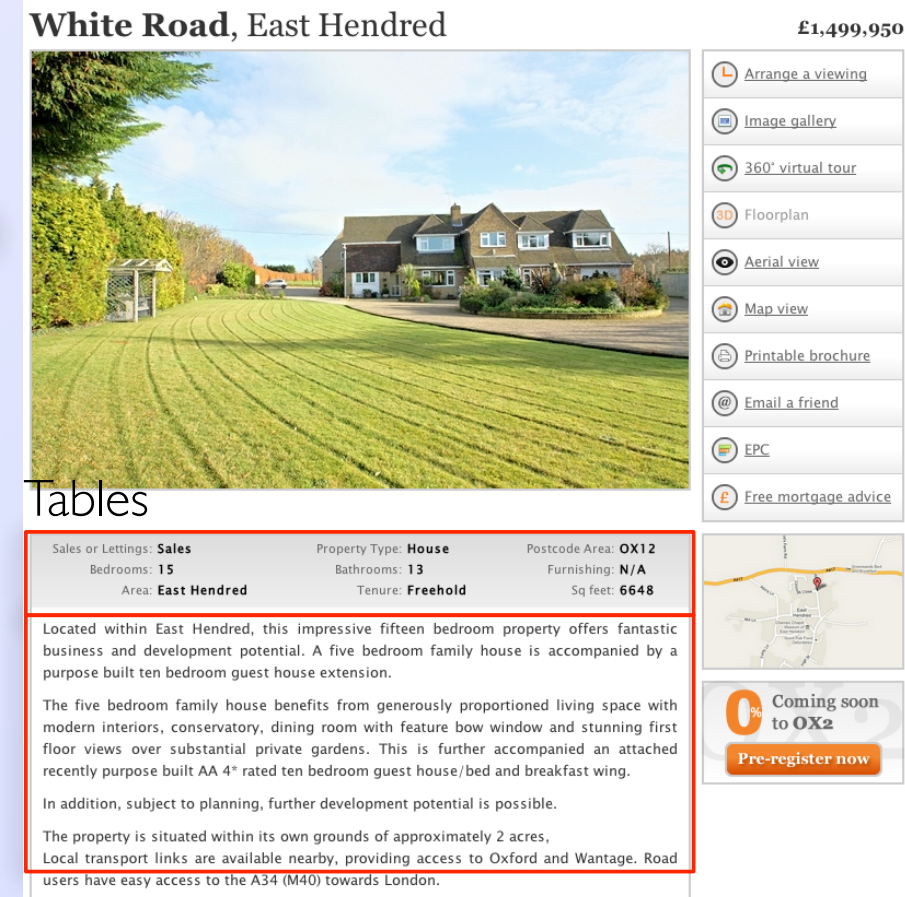
**Bagley Wood Rd, Kennington OX1** £1,295,000

A substantial detached Edwardian family house with scope for updating and improvement, set within delightful gardens and wooded grounds extending to over five acres located on the edge of Bagley Wood. This fine [...] [Read more](#)

Bedrooms: 6 | Bathrooms: 3 | Area: Kennington

[View all 16 photos](#) [View full details](#) [360° virtual tour](#) [Arrange a viewing](#) [Map view](#)

## Single entity (details) pages



**White Road, East Hendred** £1,499,950

[Arrange a viewing](#) [Image gallery](#) [360° virtual tour](#) [Floorplan](#) [Aerial view](#) [Map view](#) [Printable brochure](#) [Email a friend](#) [EPC](#) [Free mortgage advice](#)

**Tables**

Sales or Lettings:	Property Type:	Postcode Area:
<b>Sales</b>	<b>House</b>	<b>OX12</b>
Bedrooms: <b>15</b>	Bathrooms: <b>13</b>	Furnishing: <b>N/A</b>
Area: <b>East Hendred</b>	Tenure: <b>Freehold</b>	Sq feet: <b>6648</b>

Located within East Hendred, this impressive fifteen bedroom property offers fantastic business and development potential. A five bedroom family house is accompanied by a purpose built ten bedroom guest house extension.

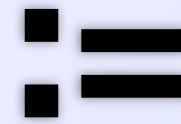
The five bedroom family house benefits from generously proportioned living space with modern interiors, conservatory, dining room with feature bow window and stunning first floor views over substantial private gardens. This is further accompanied an attached recently purpose built AA 4\* rated ten bedroom guest house/bed and breakfast wing.

In addition, subject to planning, further development potential is possible.

The property is situated within its own grounds of approximately 2 acres, Local transport links are available nearby, providing access to Oxford and Wantage. Road users have easy access to the A34 (M40) towards London.

**Coming soon to OX2**  
[Pre-register now](#)





**Object**  
identification & alignment

Context-driven  
**block** analysis

2

3


## Result pages

**WHITE WALLS**  
people not just property

Home About us Sales Lettings/Management Free valuation EPC Careers Links Contact us

Property in Oxford search results [Save search](#)

[Go back to postcode area map](#) 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 NEXT 1 - 10 of 153 properties




**White Road, East Hendred** **£1,499,950**

Located within East Hendred, this impressive fifteen bedroom property offers fantastic business and development potential. A five bedroom family house is accompanied by a purpose built ten bedroom guest house extension. The five [...] [Read more](#)

Bedrooms: 15 | Bathrooms: 13 | Area: East Hendred

[View all 10 photos](#) [View full details](#) [360° virtual tour](#) [Arrange a viewing](#) [Map view](#)



**Bagley Wood Rd, Kennington OX1** **£1,295,000**


A substantial detached Edwardian family house with scope for updating and improvement, set within delightful gardens and wooded grounds extending to over five acres located on the edge of Bagley Wood. This fine [...] [Read more](#)

Bedrooms: 6 | Bathrooms: 3 | Area: Kennington

[View all 16 photos](#) [View full details](#) [360° virtual tour](#) [Arrange a viewing](#) [Map view](#)

## Single entity (details) pages

**White Road, East Hendred** **£1,499,950**



- [Arrange a viewing](#)
- [Image gallery](#)
- [360° virtual tour](#)
- [Floorplan](#)
- [Aerial view](#)
- [Map view](#)
- [Printable brochure](#)
- [Email a friend](#)
- [EPC](#)
- [Free mortgage advice](#)

**Tables**

Sales or Lettings: <b>Sales</b>	Property Type: <b>House</b>	Postcode Area: <b>OX12</b>
Bedrooms: <b>15</b>	Bathrooms: <b>13</b>	Furnishing: <b>N/A</b>
Area: <b>East Hendred</b>	Tenure: <b>Freehold</b>	Sq feet: <b>6648</b>

Located within East Hendred, this impressive fifteen bedroom property offers fantastic business and development potential. A five bedroom family house is accompanied by a purpose built ten bedroom guest house extension.

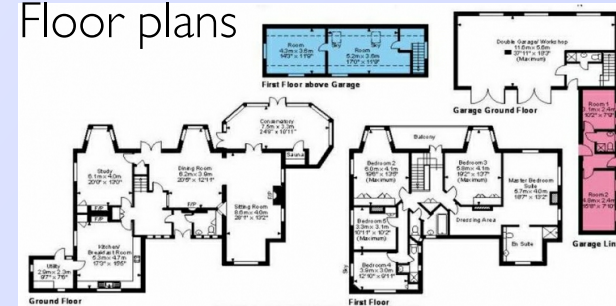
The five bedroom family house benefits from generously proportioned living space with modern interiors, conservatory, dining room with feature bow window and stunning first floor views over substantial private gardens. This is further accompanied an attached recently purpose built AA 4\* rated ten bedroom guest house/bed and breakfast wing.

In addition, subject to planning, further development potential is possible.

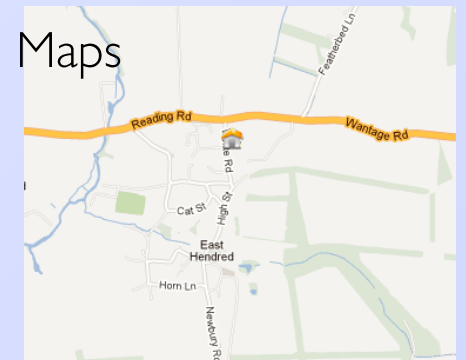
The property is situated within its own grounds of approximately 2 acres, Local transport links are available nearby, providing access to Oxford and Wantage. Road users have easy access to the A34 (M40) towards London.

**Coming soon to OX2**  
[Pre-register now](#)

## Floor plans

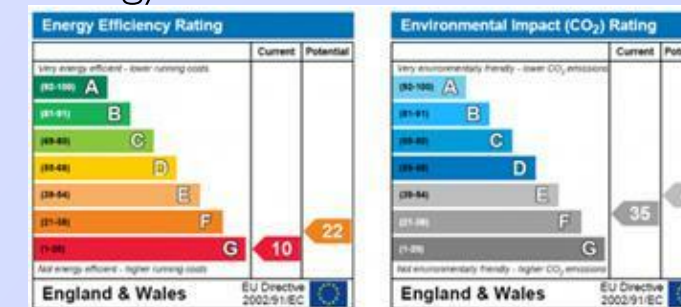


## Maps

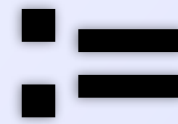


Flat Text

## Energy Performance Chart







**Object**  
Identification & alignment

2

Context-driven  
**block** analysis

3

**OXPath**  
**Wrapper**

4

Cloud extraction

Data integration

es


Results  
6 7 8 9 10 11 12 13 14 15 16 NEXT 1 - 10 of 153 properties

White Road, East Hendred £1,499,950  
New price  
This property offers an impressive fifteen bedroom family house with scope for development. A five bedroom family house is accompanied by a purpose built ten bedroom guest house extension. The five [-] Read more  
Bedrooms: 13 | Area: East Hendred

White Road, Kennington OX1 £1,895,000  
Sold  
This Edwardian family house with scope for improvement, set within delightful gardens and extending to over five acres located on the edge of the village [-] Read more  
Bedrooms: 3 | Area: Kennington

## Single entity (details) pages

White Road, East Hendred £1,499,950



Tables

Sales or Lettings:	Sales	Property Type:	House	Postcode Area:	OX12
Bedrooms:	15	Bathrooms:	13	Furnishing:	N/A
Area:	East Hendred	Tenure:	Freehold	Sq feet:	6648

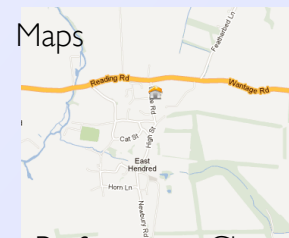
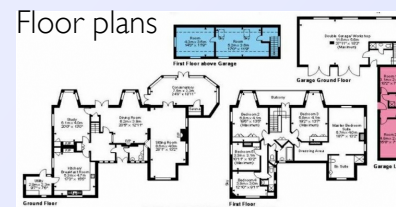
Located within East Hendred, this impressive fifteen bedroom property offers fantastic business and development potential. A five bedroom family house is accompanied by a purpose built ten bedroom guest house extension.

The five bedroom family house benefits from generously proportioned living space with modern interiors, conservatory, dining room with feature bow window and stunning first floor views over substantial private gardens. This is further accompanied an attached recently purpose built AA 4\* rated ten bedroom guest house/bed and breakfast wing.

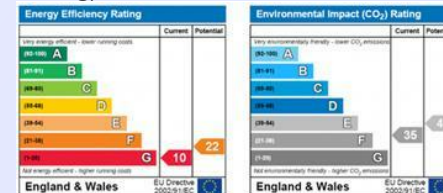
In addition, subject to planning, further development potential is possible.

The property is situated within its own grounds of approximately 2 acres. Local transport links are available nearby, providing access to Oxford and Wantage. Road users have easy access to the A34 (M40) towards London.

Coming soon to OX2  
Pre-register now

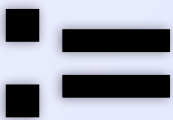


## Energy Performance Chart



Flat Text





**Form**  
Understanding & Filling

1

**Object**  
identification & alignment

2

Context-driven  
**block** analysis

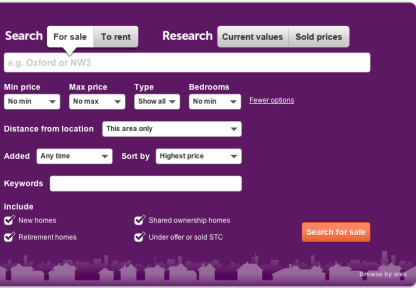
3

**OXPath**  
Wrapper

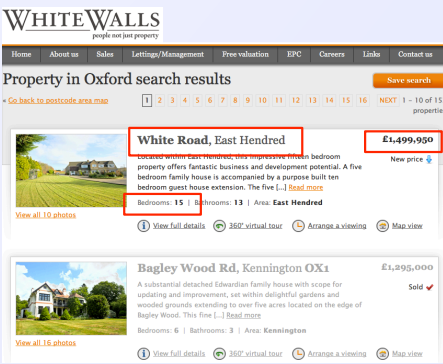
4

Cloud extraction

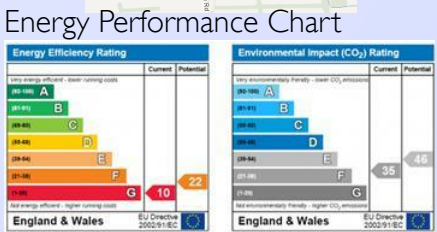
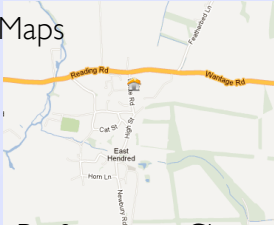
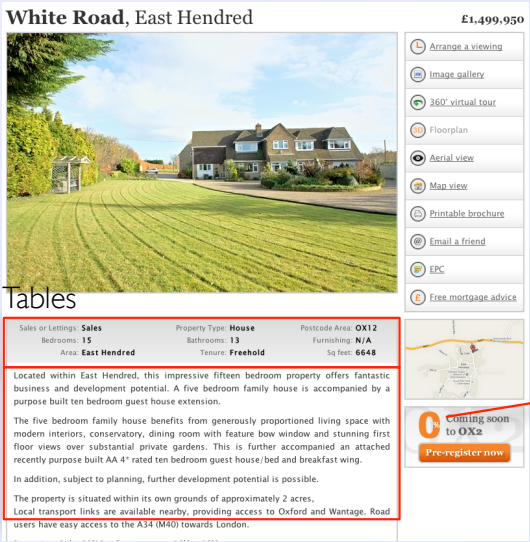
Data integration



**Result pages**



**Single entity (details) pages**

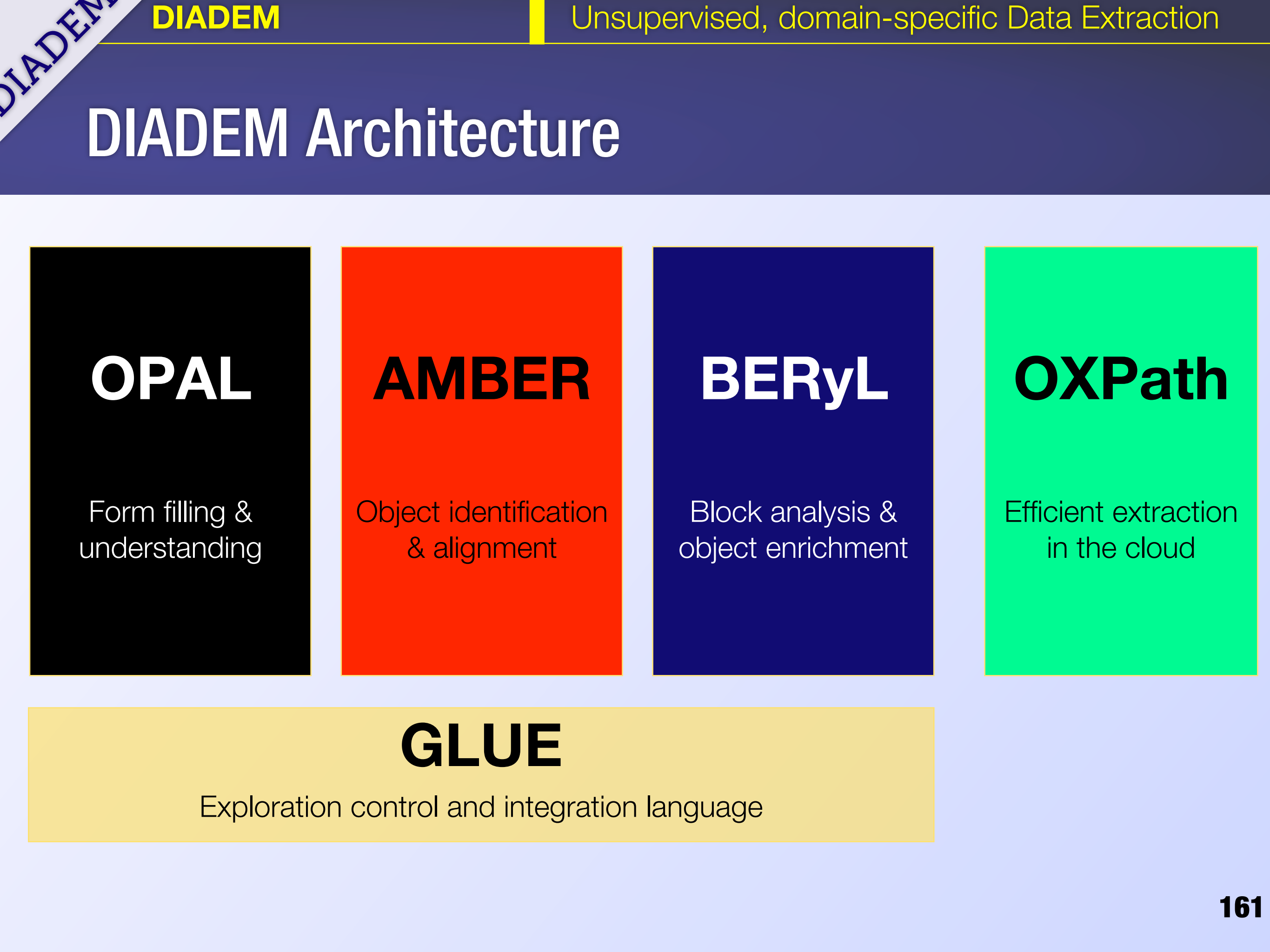


Flat Text

5

**Exploration** Control  
and Integration Language





# DIADEM Architecture

**OPAL**

Form filling & understanding

**AMBER**

Object identification & alignment

**BERyL**

Block analysis & object enrichment

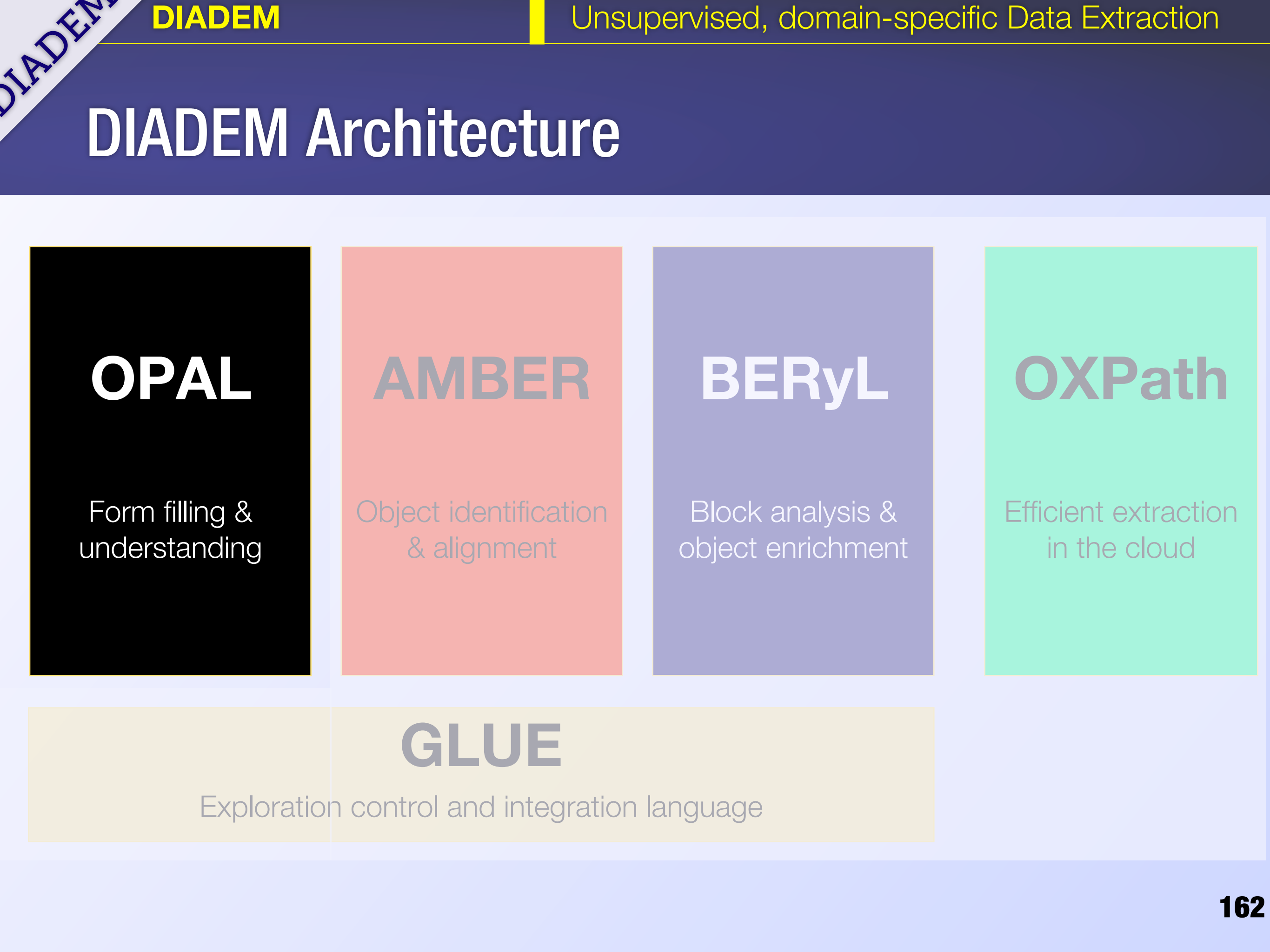
**OXPath**

Efficient extraction in the cloud

**GLUE**

Exploration control and integration language





# DIADEM Architecture

**OPAL**

Form filling & understanding

**AMBER**

Object identification & alignment

**BERyL**

Block analysis & object enrichment

**OXPath**

Efficient extraction in the cloud

**GLUE**

Exploration control and integration language



# Navigation in DIADEM: OPAL

- OPAL is DIADEM's novel framework for
  - form and interface **understanding** and
  - form and interface **navigation**
- previously navigation mostly
  - **crawler-like:** navigate all facets of an interface
  - **probing-based:** attempts many “blind” submissions
- wide applicability beyond data extraction
  - meta search; automation; assisted/mobile interfaces



# Navigation in DIADEM: OPAL

- OPAL is DIADEM's novel framework for
  - form and interface **understanding** and
  - form and interface **navigation**
- previously navigation mostly
  - crawler-like**: Furche, Gottlob, Grasso, Guo, Orsi, Schallhart, **OPAL**
  - probing-based**: **Automated form understanding for the deep web**  
*WWW 2012*
- wide applicability
  - meta search; automation; assisted/mobile interfaces



# Navigation in DIADEM: OPAL

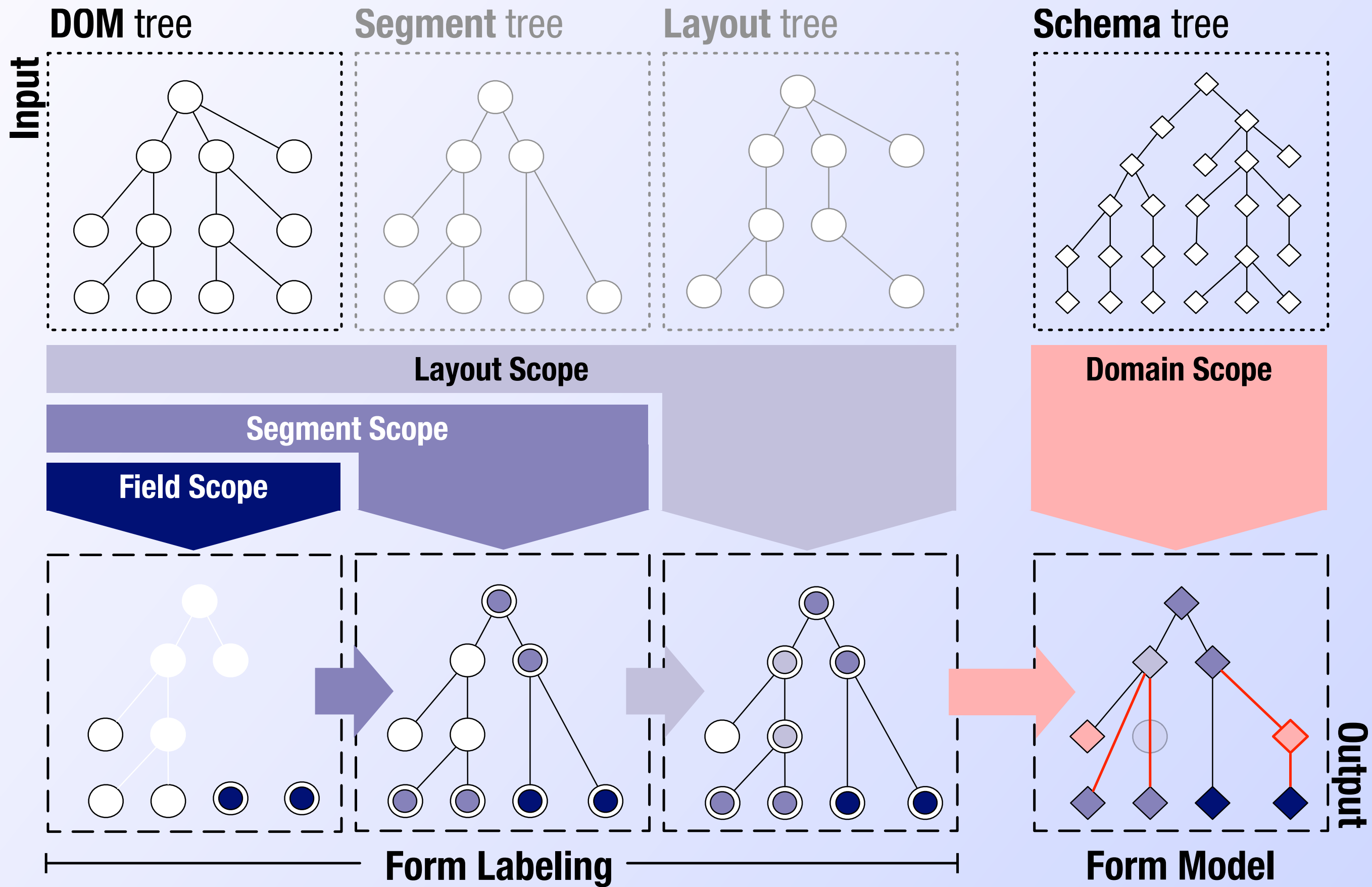
- OPAL is DIADEM's novel framework for
  - form and interface **understanding** and
  - form and interface **navigation**
- previously navigation mostly
  - **crawler-like:** navigate all facets of an interface
  - **probing-based:** attempts many "blind" submissions
- wide applicability beyond data extraction
  - meta search; automation; assisted/mobile interfaces



# Navigation in DIADEM: OPAL

- OPAL like DIADEM
  - based integrally on **knowledge** about typical interfaces in a domain
- form understanding  $\approx$  form labeling + form classification
  - **form labeling:** sophisticated domain-independent pattern analysis
    - based on knowledge about textual, visual, and structural **patterns** of forms
    - knowledge-based approach makes changing and adding patterns easier
    - previously: mostly fixed patterns, only of one kind (only visual, only textual, ...)
  - **form classification:**
    - fairly simple, Datalog-based ontology, but with strong templates
    - to reduce the cost of adding new domains







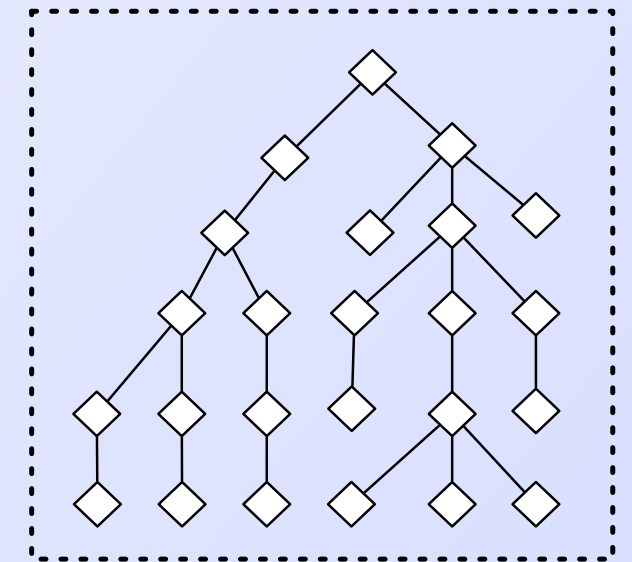
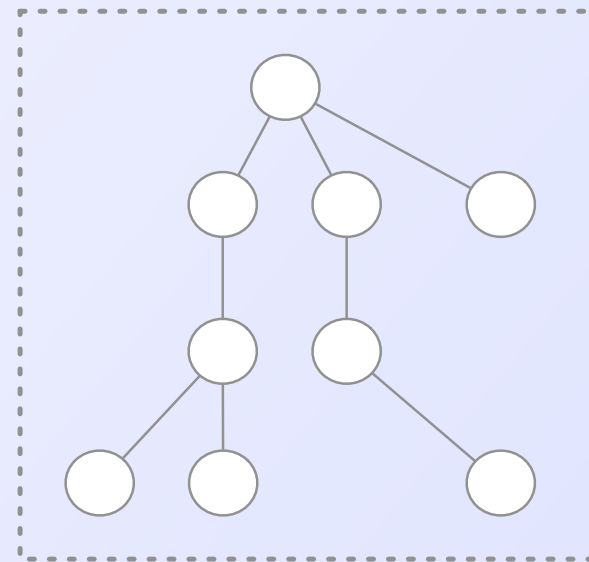
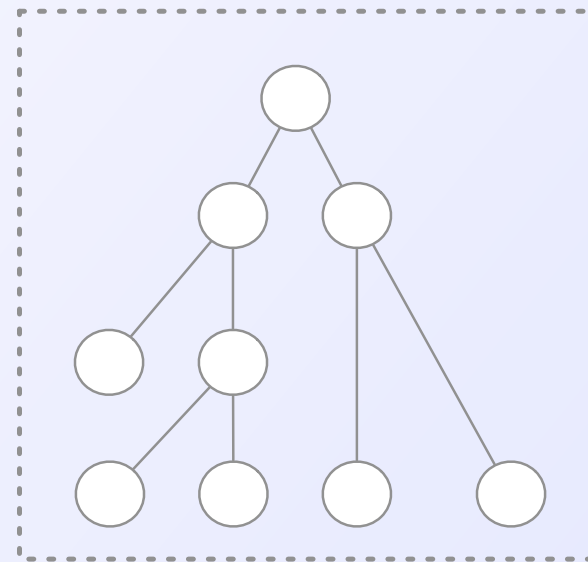
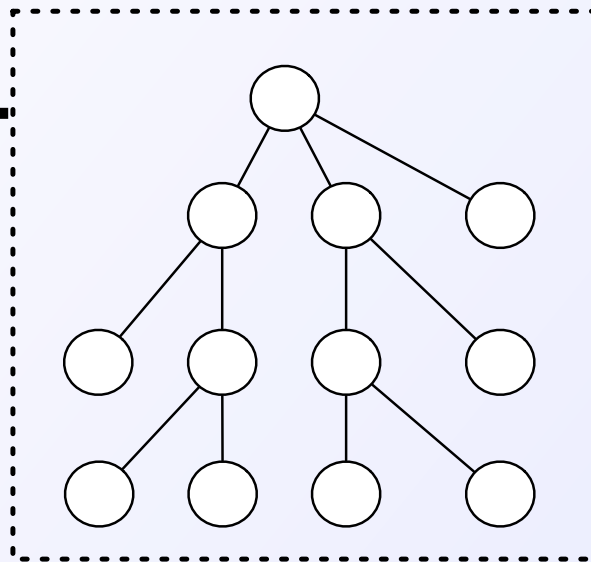
**DOM tree**

**Segment tree**

**Layout tree**

**Schema tree**

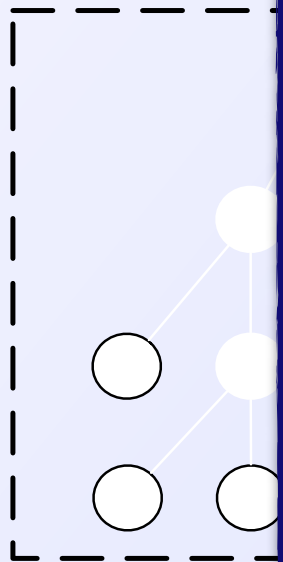
Input



**Layout Scope**

**Domain Scope**

Field

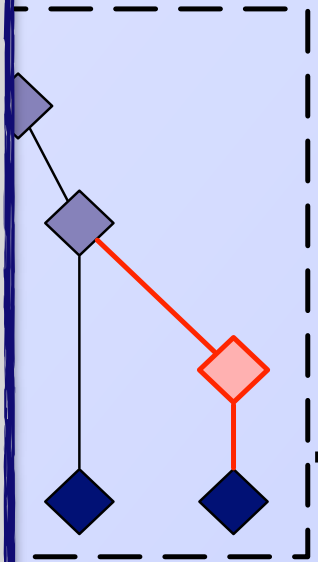


■ **form labeling** in three scopes

- field scope: each field by itself
- segment scope: groups of similar fields
- layout scope: visual layout of entire form

■ **form classification** using Datalog ontology for

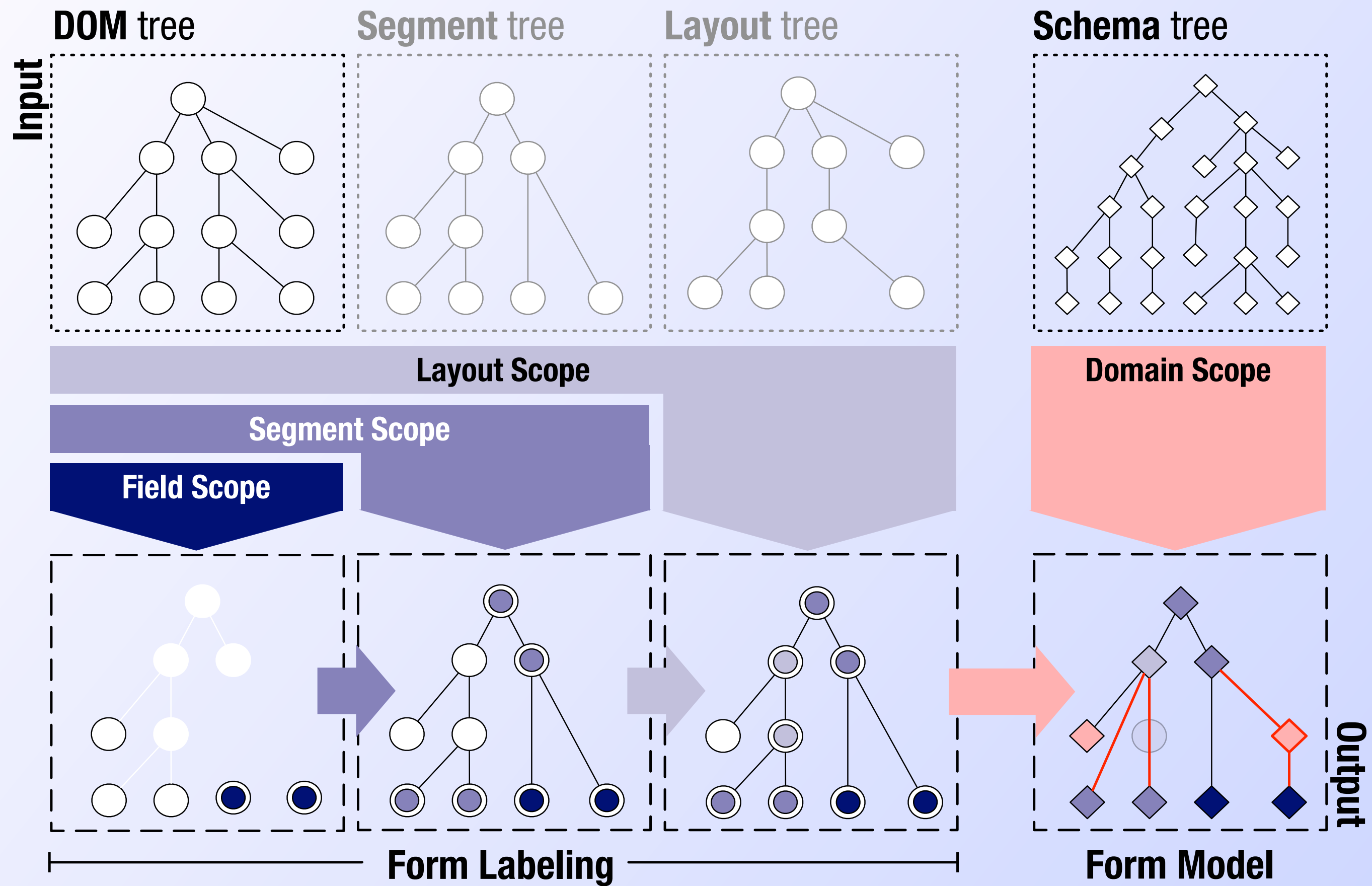
- **annotation** constraints: between label and field
- **structural** constraints: in field groups



Output

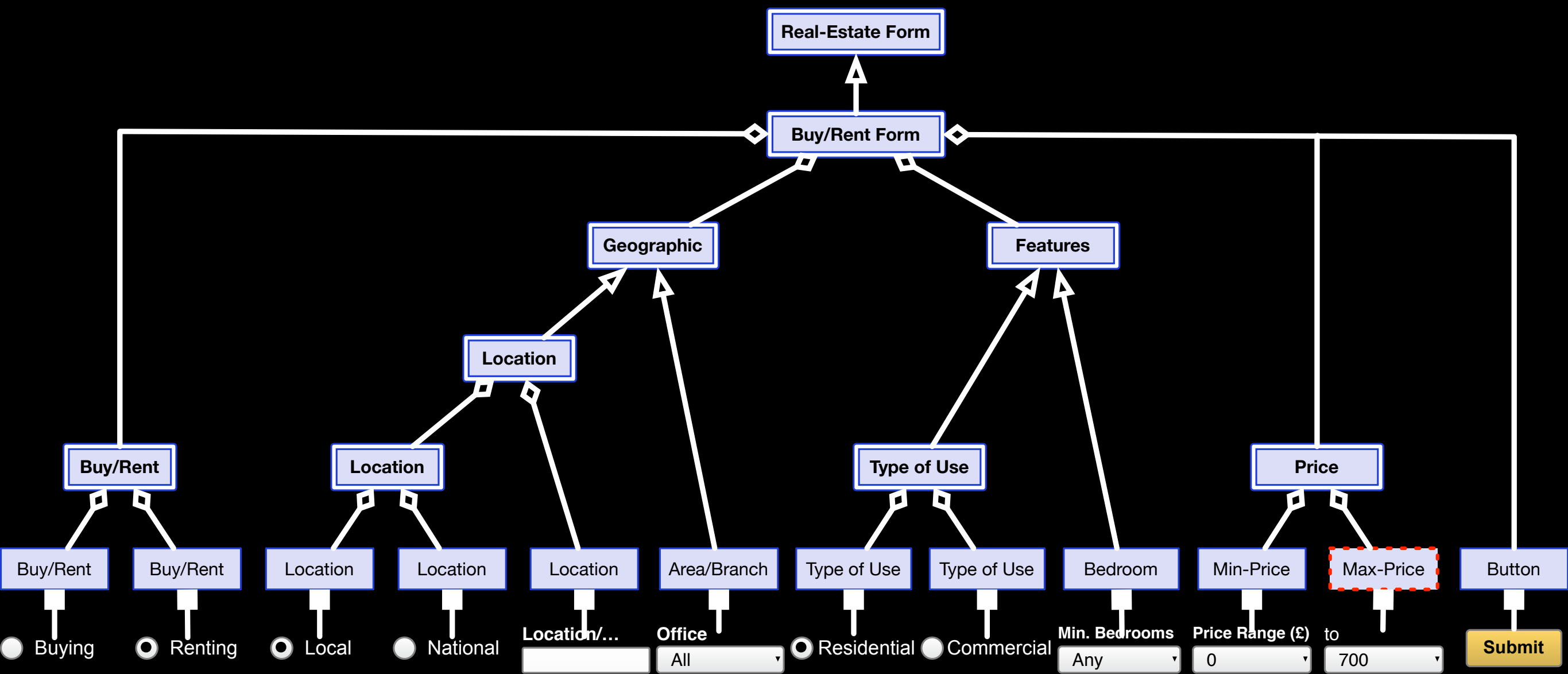
**Model**



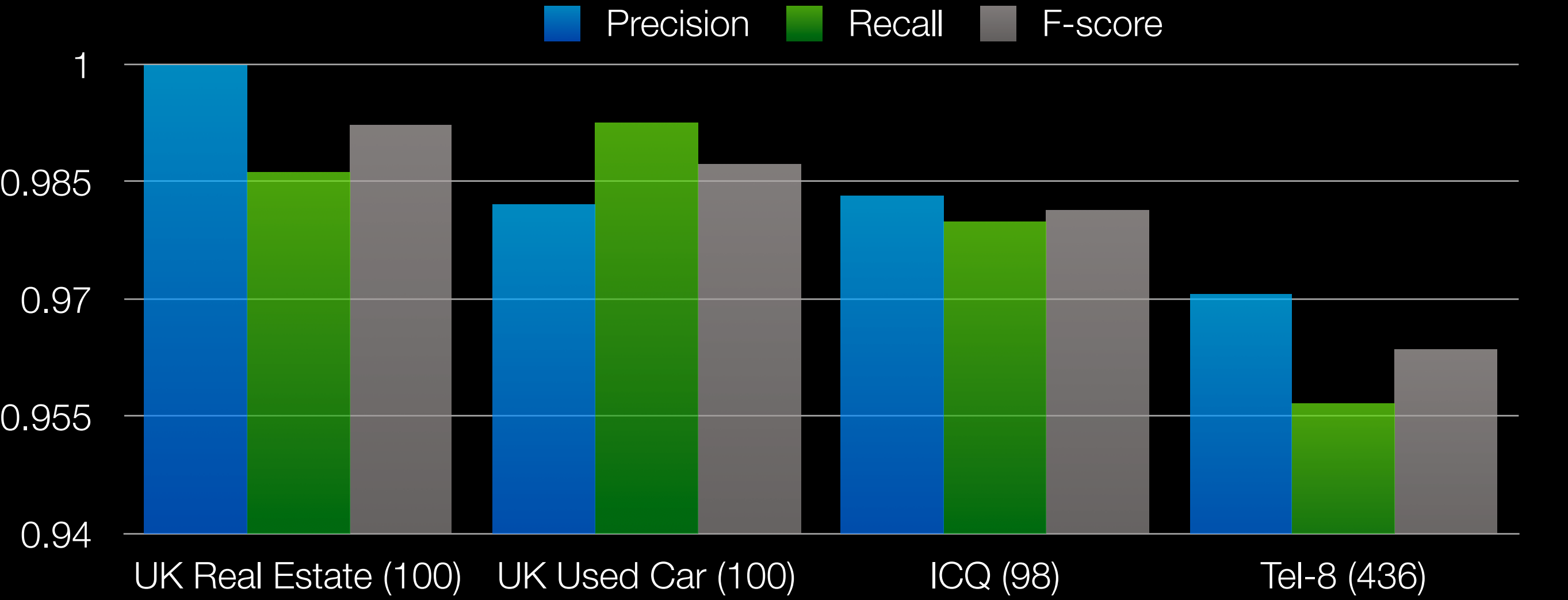




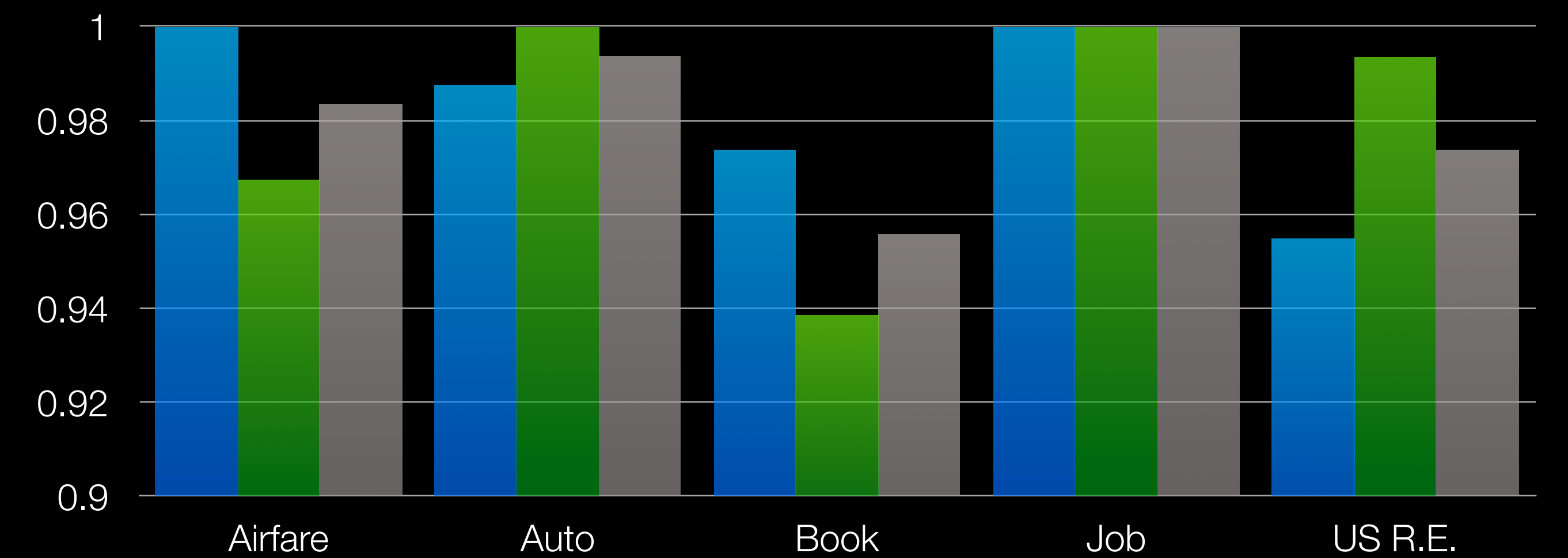
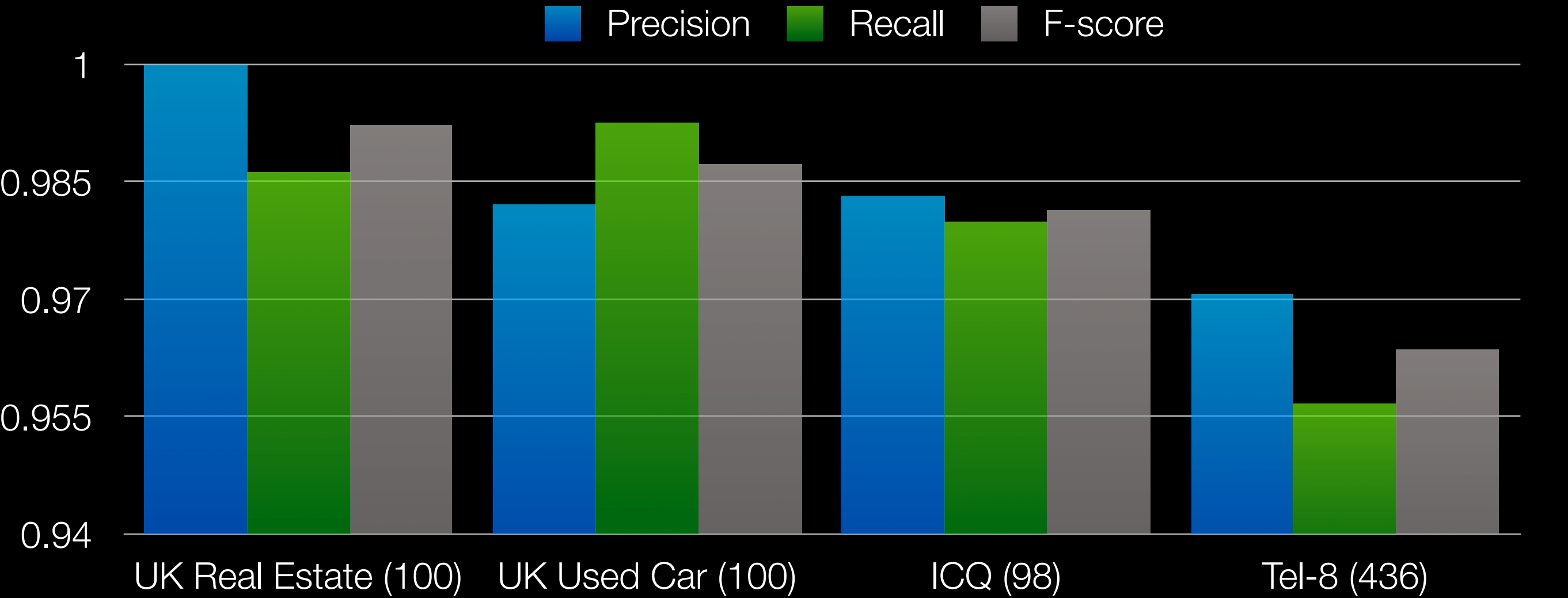
# OPAL Classification over Sample Form



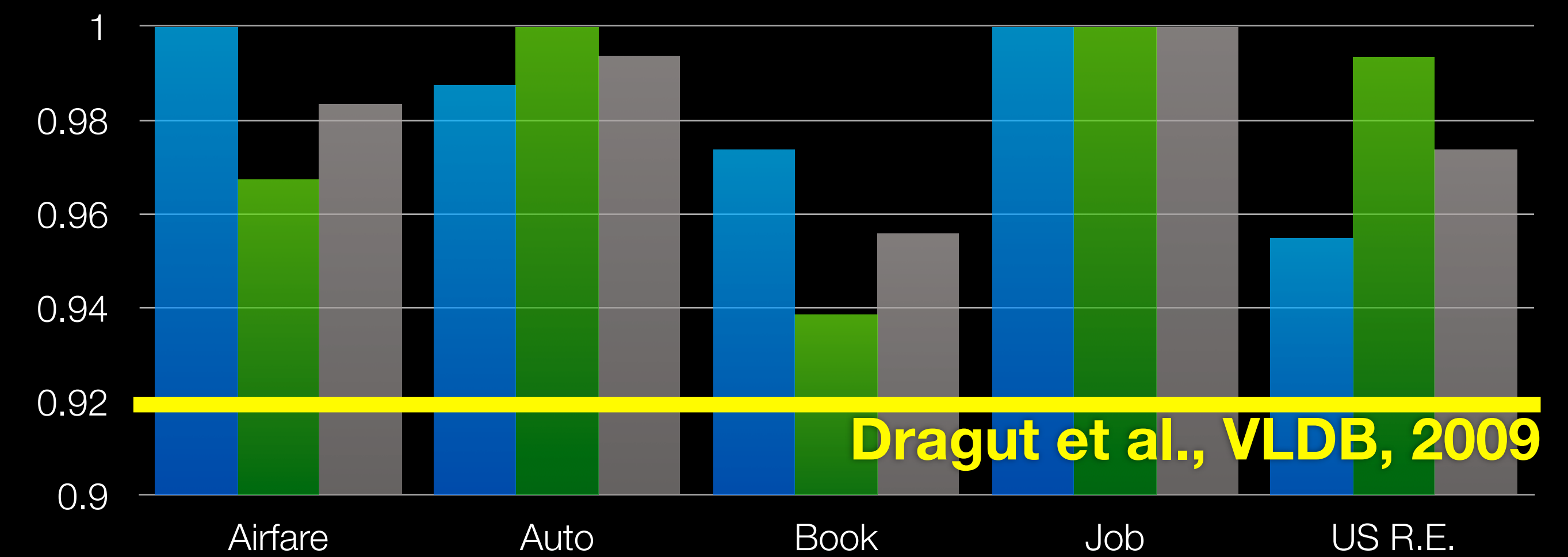
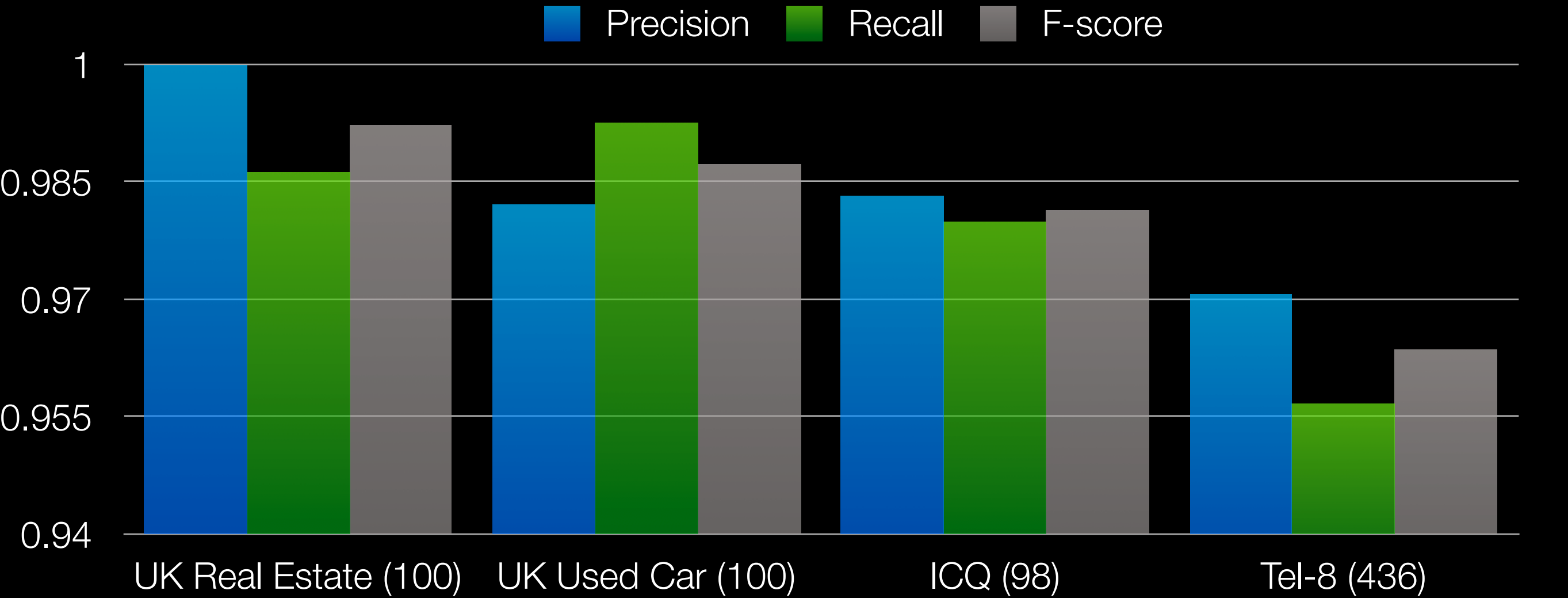




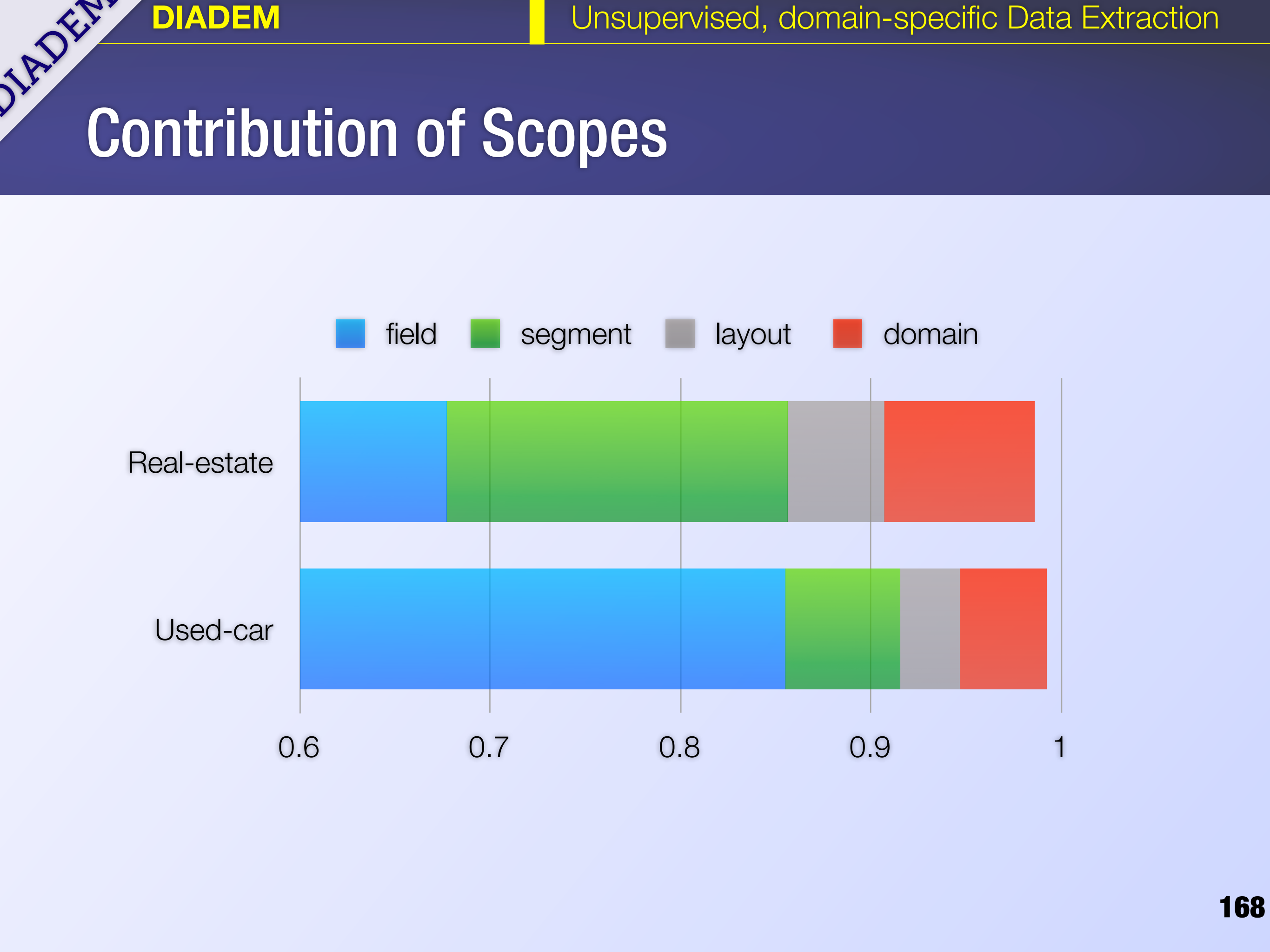




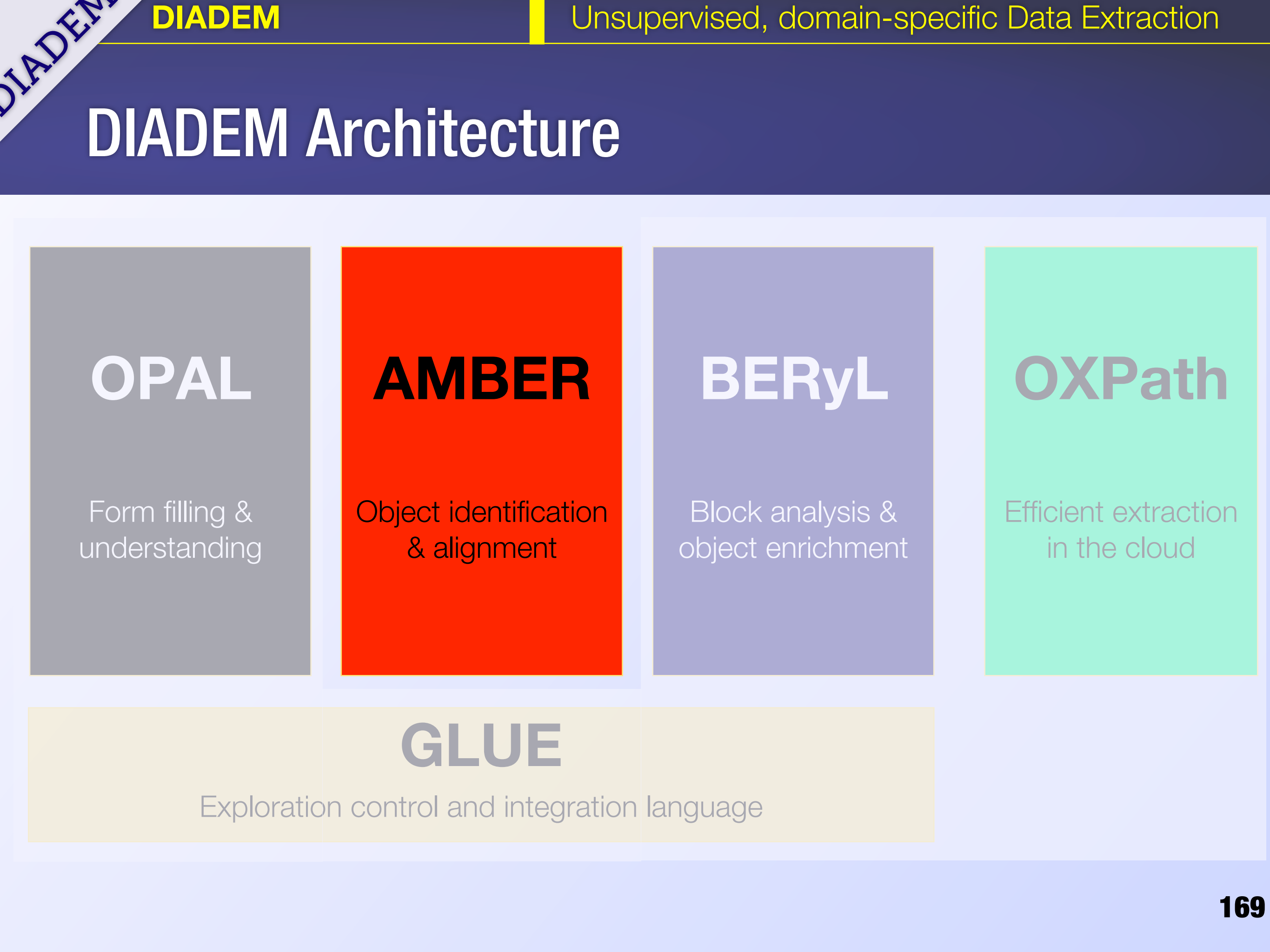












# DIADEM Architecture

**OPAL**

Form filling & understanding

**AMBER**

Object identification & alignment

**BERyL**

Block analysis & object enrichment

**OXPath**

Efficient extraction in the cloud

**GLUE**

Exploration control and integration language





**£1,950 pw (£8,473 pcm)**

5 bedroom property to rent

[Burgess Mead, Oxford](#)

Short Let Accommodation a large family house ideally situated on this popular development within walking distance to Port Meadows and Jericho. This well presented townhouse has recently been decorated throughout and offers spacious ...

[Save to favourites](#) | [Contact agent](#) | [Full details and 3 photos](#)

Marketed by Penny & Sinclair, OX2 - 01865 360094



**£1,800 pw (£7,821 pcm)**

4 bedroom property to rent

[Blandford Avenue, Oxford](#)

Short Let Accommodation a fabulous and contemporary new house that is furnished to a very high standard. Located in North Oxford it has excellent transport links in and out of the city centre. Entrance hall, living room with bay window, ...

[Save to favourites](#) | [Contact agent](#) | [Full details and 6 photos](#)

Marketed by Penny & Sinclair, OX2 - 01865 360094



**£1,610 pw (£6,995 pcm)**

6 bedroom semi detached house to rent

[Norham Road, Oxford, Oxfordshire](#)

Impressive, well presented Victorian semi detached house in sought after location

[Save to favourites](#) | [Contact agent](#) | [Full details and 5 photos](#)

Marketed by Savills, RG9 - 01491 818539



**£1,599 pw (£6,950 pcm)**

6 bedroom detached house to rent

[Wootton Village, Boars Hill, Oxford](#)

A 3 storey 6 bedroom home with a pool, tennis court and separate detached barn with double bedroom and mezzanine sleeping area, c 3 miles from Oxford City centre. Many character features include flagstone floors, exposed beams, open ...

[Save to favourites](#) | [Contact agent](#) | [Full details and 10 photos](#)

Marketed by Finders Keepers, OX14 - 01235 330060







**£1,950 pw (£8,473 pcm)**

**5 bedroom property to rent**

[Burgess Mead, Oxford](#)

Short Let Accommodation a large family house ideally situated on this popular development within walking distance to Port Meadows and Jericho. This well presented townhouse has recently been decorated throughout and offers spacious ...

[Save to favourites](#) | [Contact agent](#) | [Full details and 3 photos](#)

Marketed by Penny & Sinclair, OX2 - 01865 360094



**£1,800 pw (£7,821 pcm)**

**4 bedroom property to rent**

[Blandford Avenue, Oxford](#)

Short Let Accommodation a fabulous and contemporary new house that is furnished to a very high standard. Located in North Oxford it has excellent transport links in and out of the city centre. Entrance hall, living room with bay window, ...

[Save to favourites](#) | [Contact agent](#) | [Full details and 6 photos](#)

Marketed by Penny & Sinclair, OX2 - 01865 360094



**£1,610 pw (£6,995 pcm)**

**6 bedroom semi detached house to rent**

[Norham Road, Oxford, Oxfordshire](#)

Impressive, well presented Victorian semi detached house in sought after location

[Save to favourites](#) | [Contact agent](#) | [Full details and 5 photos](#)

Marketed by Savills, RG9 - 01491 818539



**£1,599 pw (£6,950 pcm)**

**6 bedroom detached house to rent**

[Wootton Village, Boars Hill, Oxford](#)

A 3 storey 6 bedroom home with a pool, tennis court and separate detached barn with double bedroom and mezzanine sleeping area, c 3 miles from Oxford City centre. Many character features include flagstone floors, exposed beams, open ...

[Save to favourites](#) | [Contact agent](#) | [Full details and 10 photos](#)

Marketed by Finders Keepers, OX14 - 01235 330060







**Price**

**£1,950 pw (£8,473 pcm)**

5 bedroom property to rent

[Burgess Mead, Oxford](#)

Short Let Accommodation a large family house ideally situated on this popular development within walking distance to Port Meadows and Jericho. This well presented townhouse has recently been decorated throughout and offers spacious ...

[Save to favourites](#) | [Contact agent](#) | [Full details and 3 photos](#)

Marketed by Penny & Sinclair, OX2 - 01865 360094

**PENNY & SINCLAIR**



**Price**

**£1,800 pw (£7,821 pcm)**

4 bedroom property to rent

[Blandford Avenue, Oxford](#)

Short Let Accommodation a fabulous and contemporary new house that is furnished to a very high standard. Located in North Oxford it has excellent transport links in and out of the city centre. Entrance hall, living room with bay window, ...

[Save to favourites](#) | [Contact agent](#) | [Full details and 6 photos](#)

Marketed by Penny & Sinclair, OX2 - 01865 360094

**PENNY & SINCLAIR**



**Price**

**£1,610 pw (£6,995 pcm)**

6 bedroom semi detached house to rent

[Norham Road, Oxford, Oxfordshire](#)

Impressive, well presented Victorian semi detached house in sought after location

[Save to favourites](#) | [Contact agent](#) | [Full details and 5 photos](#)

Marketed by Savills, RG9 - 01491 818539

**savills**



**Price**

**£1,599 pw (£6,950 pcm)**

6 bedroom detached house to rent

[Wootton Village, Boars Hill, Oxford](#)

A 3 storey 6 bedroom home with a pool, tennis court and separate detached barn with double bedroom and mezzanine sleeping area, c 3 miles from Oxford City centre. Many character features include flagstone floors, exposed beams, open ...

[Save to favourites](#) | [Contact agent](#) | [Full details and 10 photos](#)

Marketed by Finders Keepers, OX14 - 01235 330060

**Finders Keepers**  
[www.finders.co.uk](http://www.finders.co.uk)





**Price**  
**£1,950 pw (£8,473 pcm)**

**5 bedroom property to rent**

**Location**  
**Burgess Mead, Oxford**

Short Let Accommodation a large family house ideally situated on this popular development within walking distance to Port Meadows and Jericho. This well presented townhouse has recently been decorated throughout and offers spacious ...

[Save to favourites](#) | [Contact agent](#) | [Full details and 3 photos](#)

Marketed by Penny & Sinclair, OX2 - 01865 360094

**PENNY &  
SINCLAIR**



**Price**  
**£1,800 pw (£7,821 pcm)**

**4 bedroom property to rent**

**Location**  
**Wood Avenue, Oxford**

Short Let Accommodation a fabulous and contemporary new house that is furnished to a very high standard. Located in North Oxford it has excellent transport links in and out of the city centre. Entrance hall, living room with bay window, ...

[Save to favourites](#) | [Contact agent](#) | [Full details and 6 photos](#)

Marketed by Penny & Sinclair, OX2 - 01865 360094

**PENNY &  
SINCLAIR**



**Price**  
**£1,610 pw (£6,995 pcm)**

**6 bedroom semi detached house to rent**

**Location**  
**Henley Road, Oxford, Oxfordshire**

Impressive, well presented Victorian semi detached house in sought after location

[Save to favourites](#) | [Contact agent](#) | [Full details and 5 photos](#)

Marketed by Savills, RG9 - 01491 818539

**savills**



**Price**  
**£1,599 pw (£6,950 pcm)**

**6 bedroom detached house to rent**

**Location**  
**Wotton Village, Boars Hill, Oxford**

A 3 storey 6 bedroom home with a pool, tennis court and separate detached barn with double bedroom and mezzanine sleeping area, c 3 miles from Oxford City centre. Many character features include flagstone floors, exposed beams, open ...

[Save to favourites](#) | [Contact agent](#) | [Full details and 10 photos](#)

Marketed by Finders Keepers, OX14 - 01235 330060

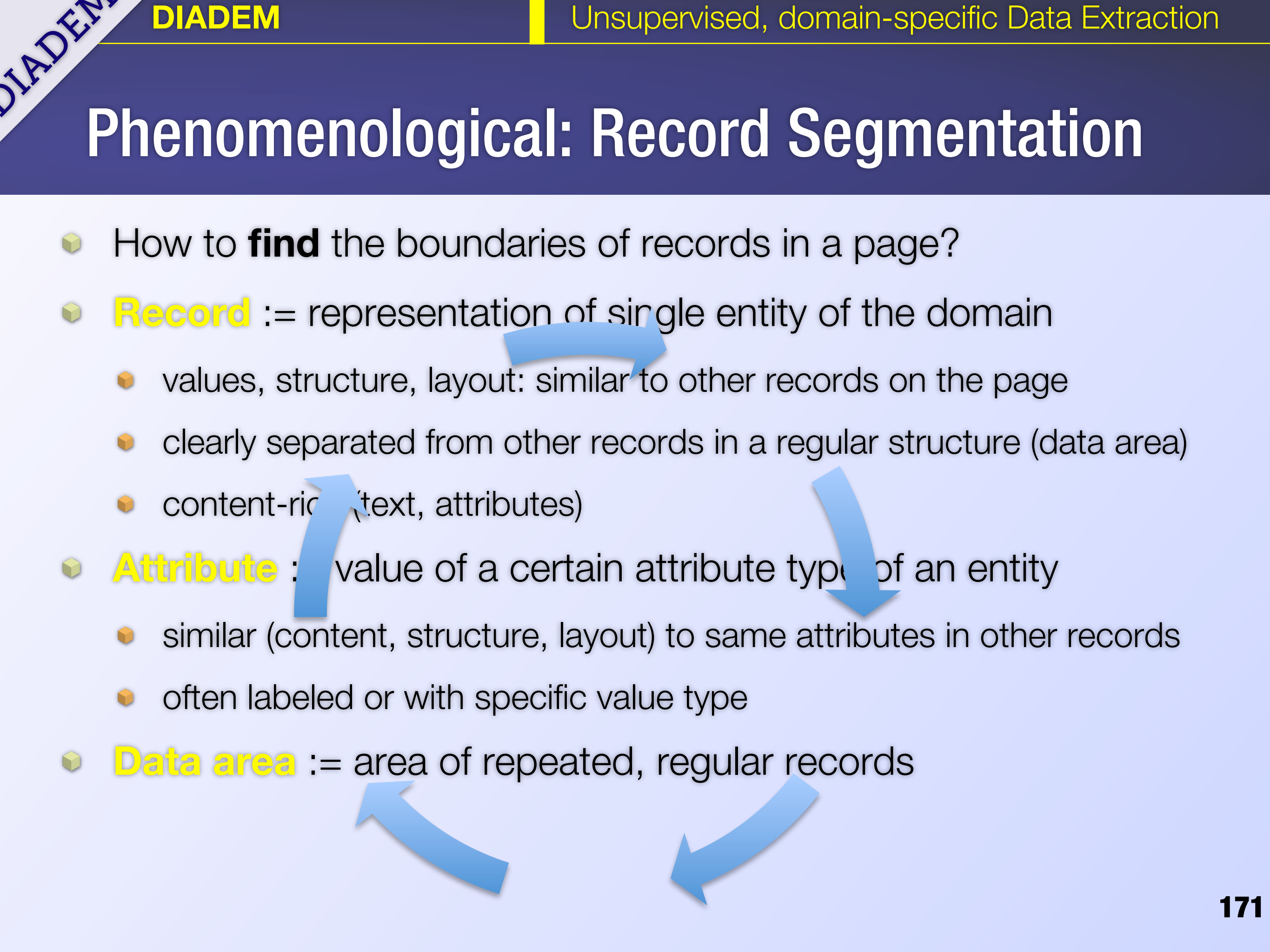
**Finders  
Keepers**  
[www.finders.co.uk](http://www.finders.co.uk)



# Phenomenological: Record Segmentation

- How to **find** the boundaries of records in a page?
- Record** := representation of single entity of the domain
  - values, structure, layout: similar to other records on the page
  - clearly separated from other records in a regular structure (data area)
  - content-rich (text, attributes)
- Attribute** := value of a certain attribute type of an entity
  - similar (content, structure, layout) to same attributes in other records
  - often labeled or with specific value type
- Data area** := area of repeated, regular records





- How to **find** the boundaries of records in a page?
- Record** := representation of single entity of the domain
  - values, structure, layout: similar to other records on the page
  - clearly separated from other records in a regular structure (data area)
  - content-rich (text, attributes)
- Attribute** := value of a certain attribute type of an entity
  - similar (content, structure, layout) to same attributes in other records
  - often labeled or with specific value type
- Data area** := area of repeated, regular records



# Phenomenological: Record Segmentation

- Finding optimal “record model”
  - in presence of strong and weak constraints
  - model generation or probabilistic reasoning approaches infeasible
    - > 200k input facts, > 400 rules
    - does not address low precision, but **low recall** is at least as much of an issue
- therefore: **AMBER** search informed by domain knowledge
  - use domain knowledge to guess **data area & record segmentation**
  - support **alignment** with domain knowledge





# Phenomenological: Record Segmentation

- Finding optimal “record model”
  - in presence of strong and weak constraints
  - model generation or probabilistic reasoning approaches infeasible
    - > 200k input facts, > 400 rules
    - does not address low precision, but **low recall** is at least as much of an issue
- therefore: **AMBER** search informed by domain knowledge
  - use domain knowledge to guess **data area & record segmentation**
  - support **alignment** with domain knowledge

Furche, Gottlob, Grasso, Orsi, Schallhart and Wang. **Little Knowledge Rules The Web: Domain-Centric Result Page Extraction**. In *Web Reasoning and Rule Systems (RR)*. 2011



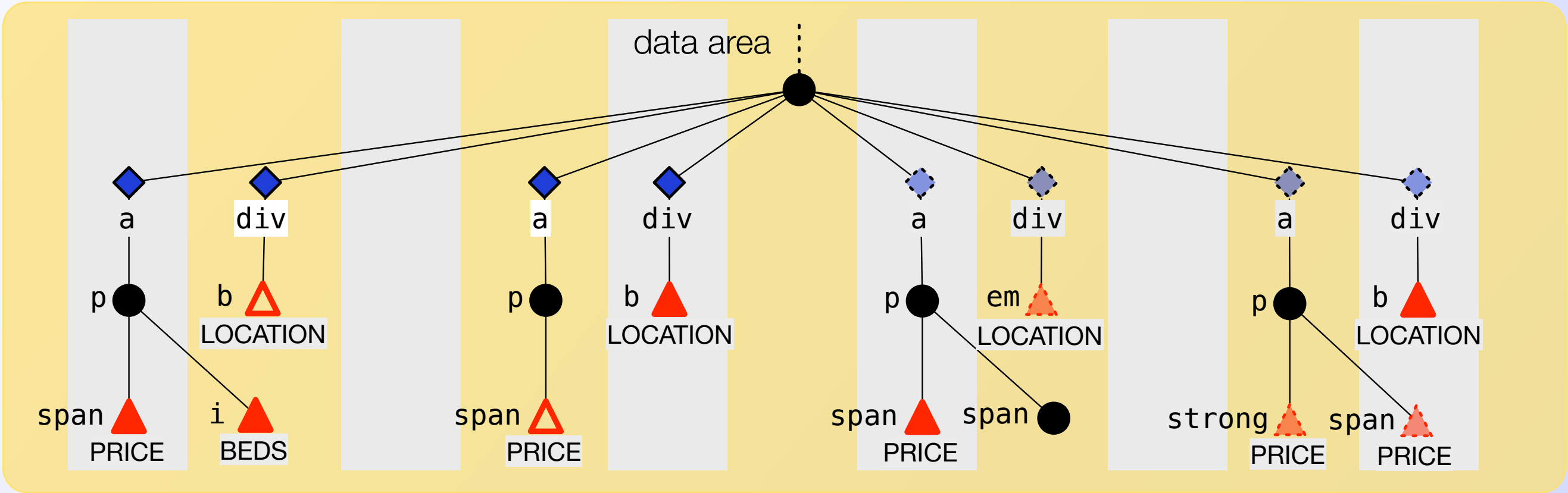
# Phenomenological: Record Segmentation

- Finding optimal “record model”
  - in presence of strong and weak constraints
  - model generation or probabilistic reasoning approaches infeasible
    - > 200k input facts, > 400 rules
    - does not address low precision, but **low recall** is at least as much of an issue
- therefore: **AMBER** search informed by domain knowledge
  - use domain knowledge to guess **data area & record segmentation**
  - support **alignment** with domain knowledge



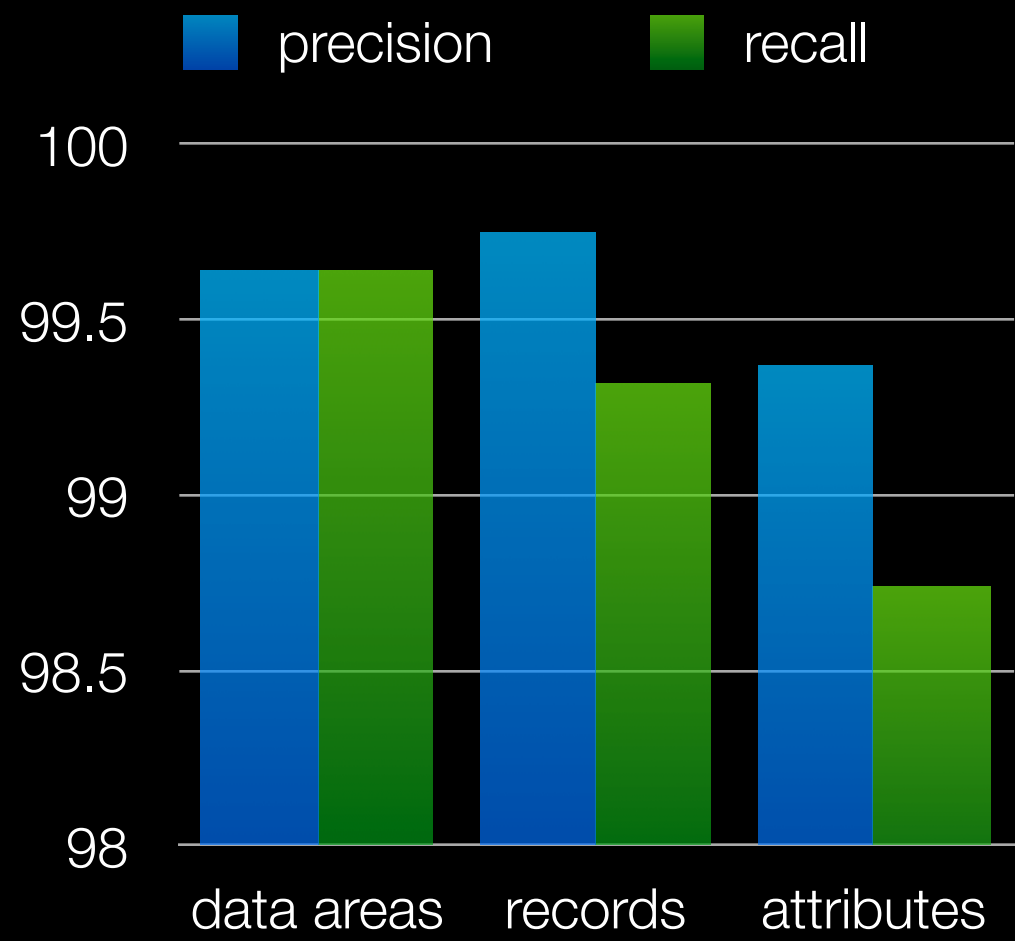


# Phenomenological: Attribute Alignment

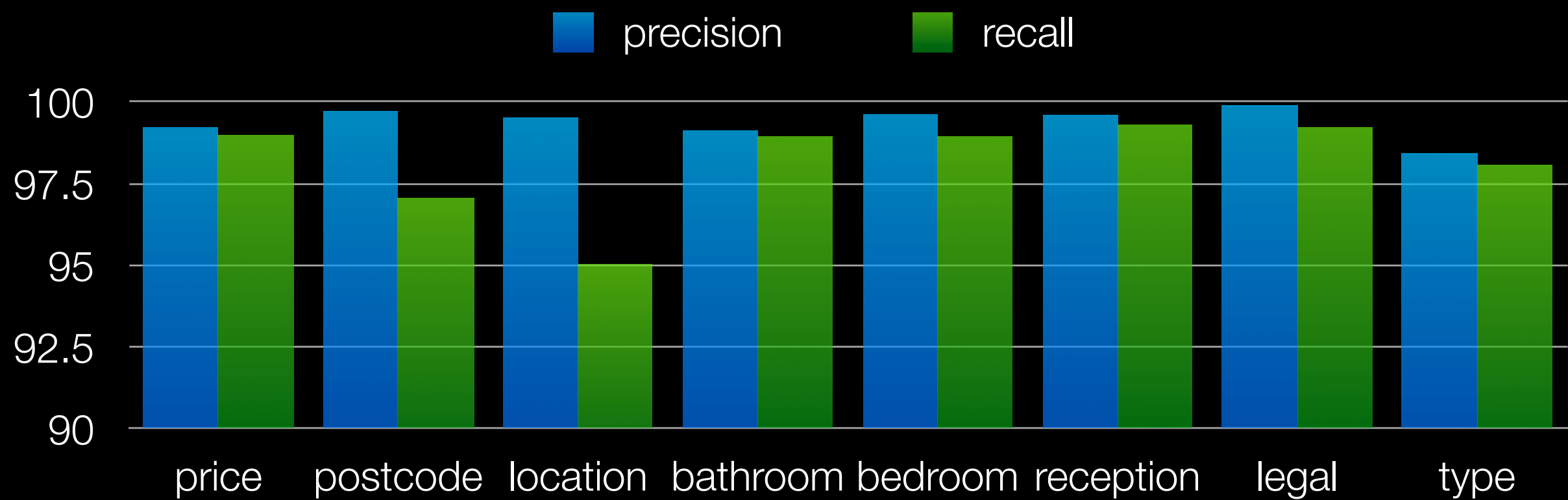
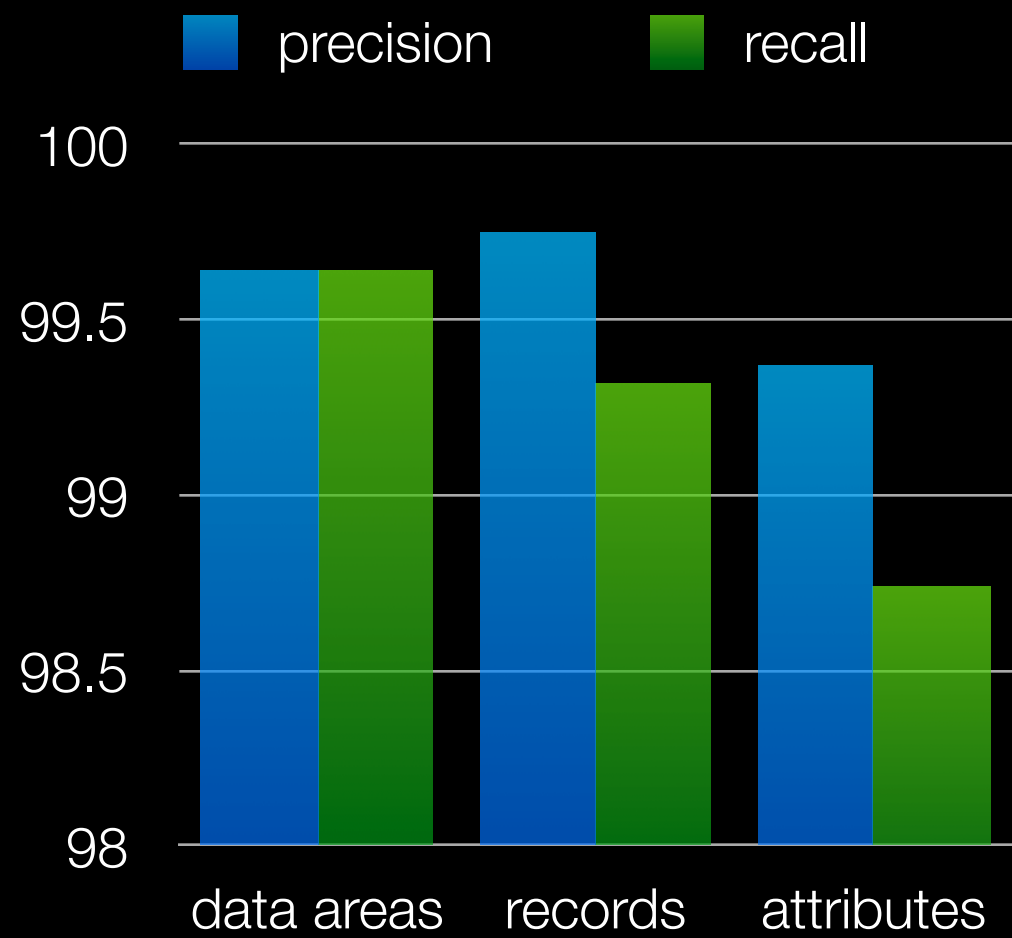


- optimal attribute alignment
  - well-supported and consistent for  $\Sigma$
  - maximal in the number of mandatory attributes

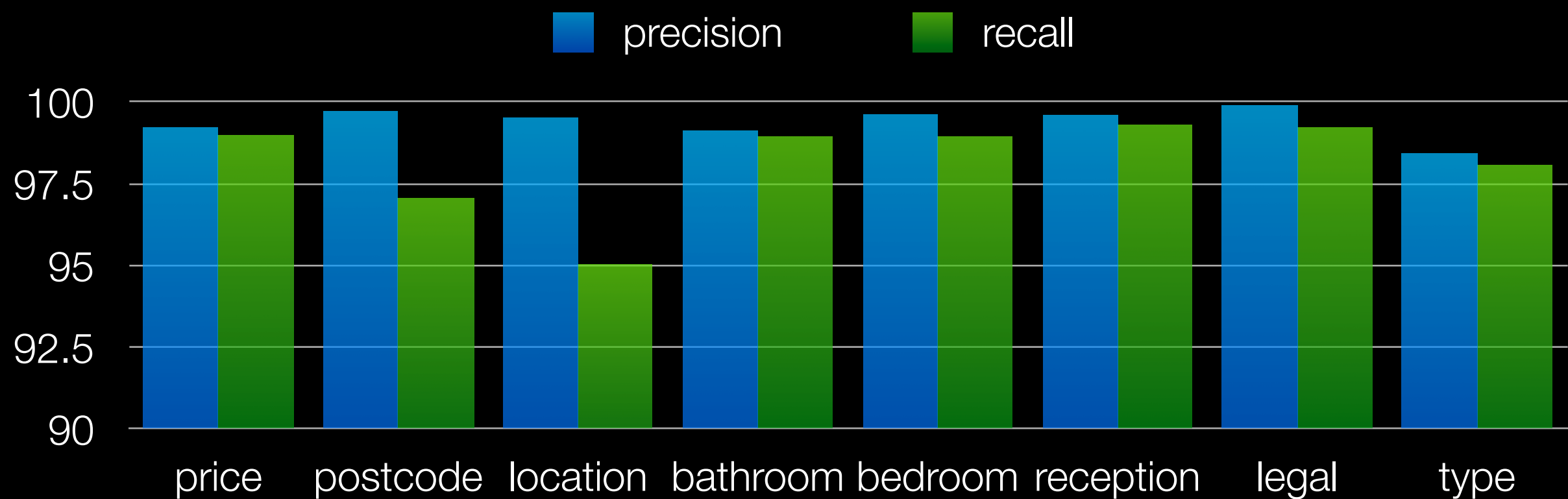
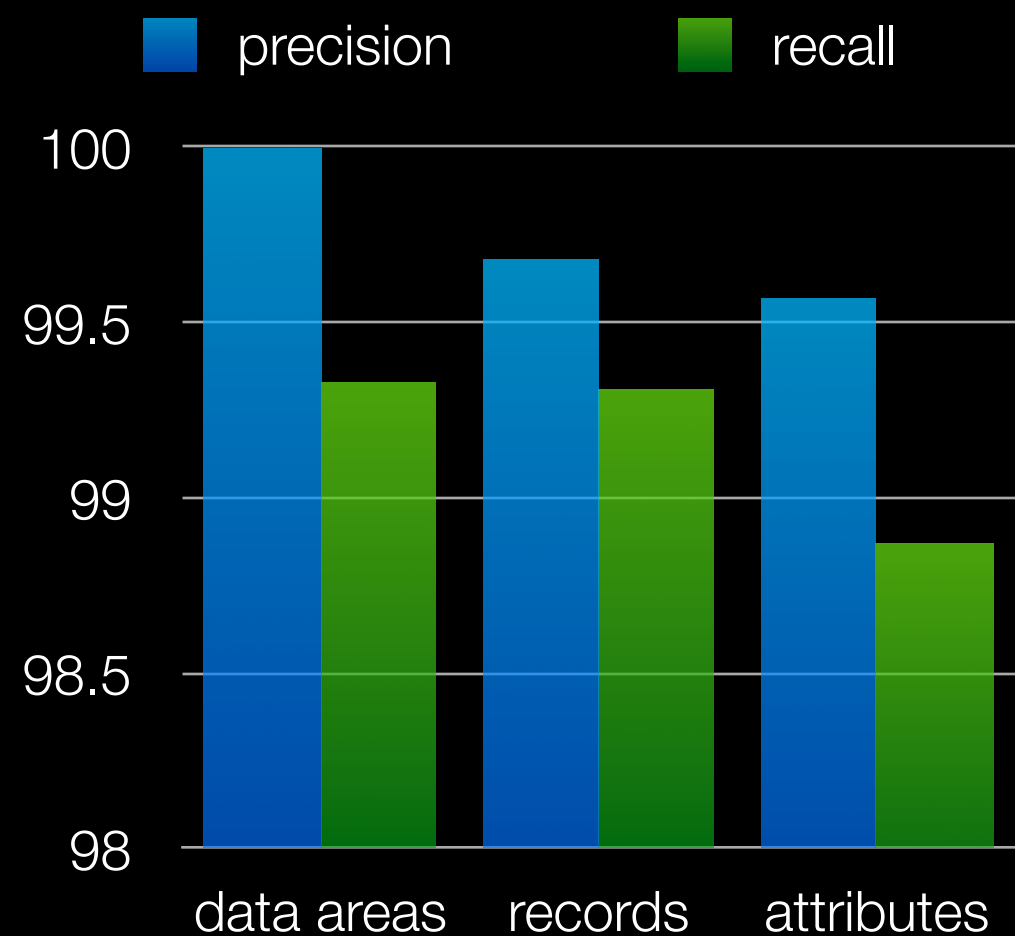
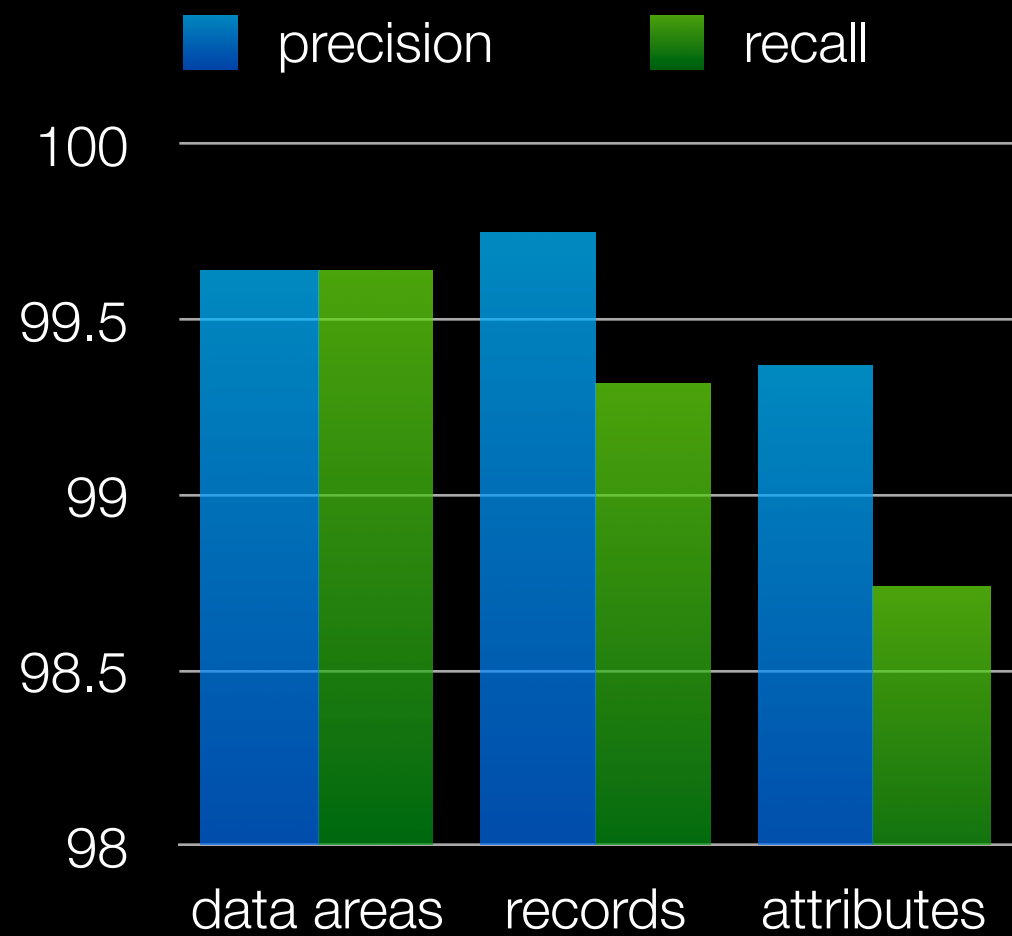




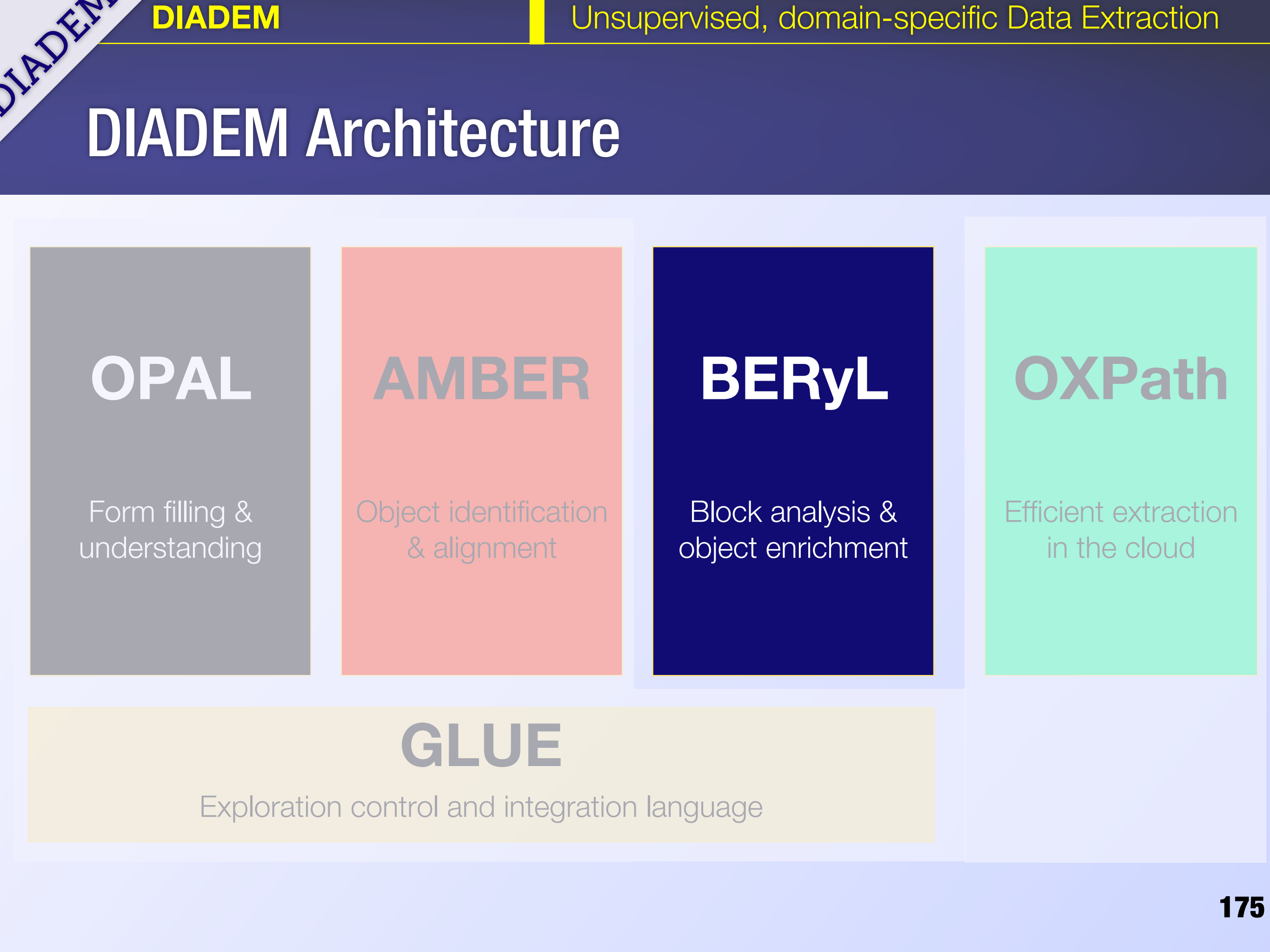












# DIADEM Architecture

**OPAL**

Form filling & understanding

**AMBER**

Object identification & alignment

**BERyL**

Block analysis & object enrichment

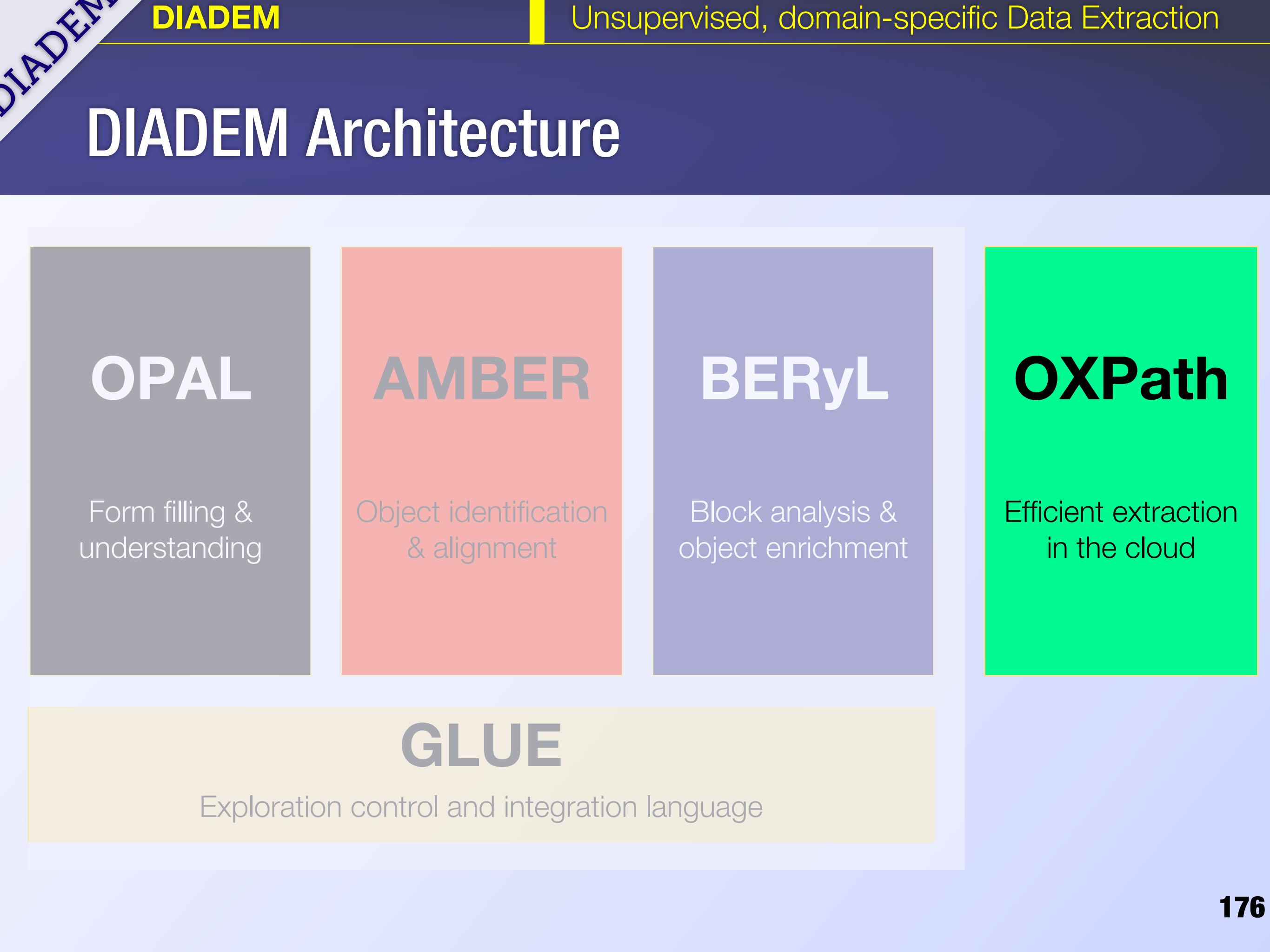
**OXPath**

Efficient extraction in the cloud

**GLUE**

Exploration control and integration language





# DIADEM Architecture

**OPAL**

Form filling &  
understanding

**AMBER**

Object identification  
& alignment

**BERyL**

Block analysis &  
object enrichment

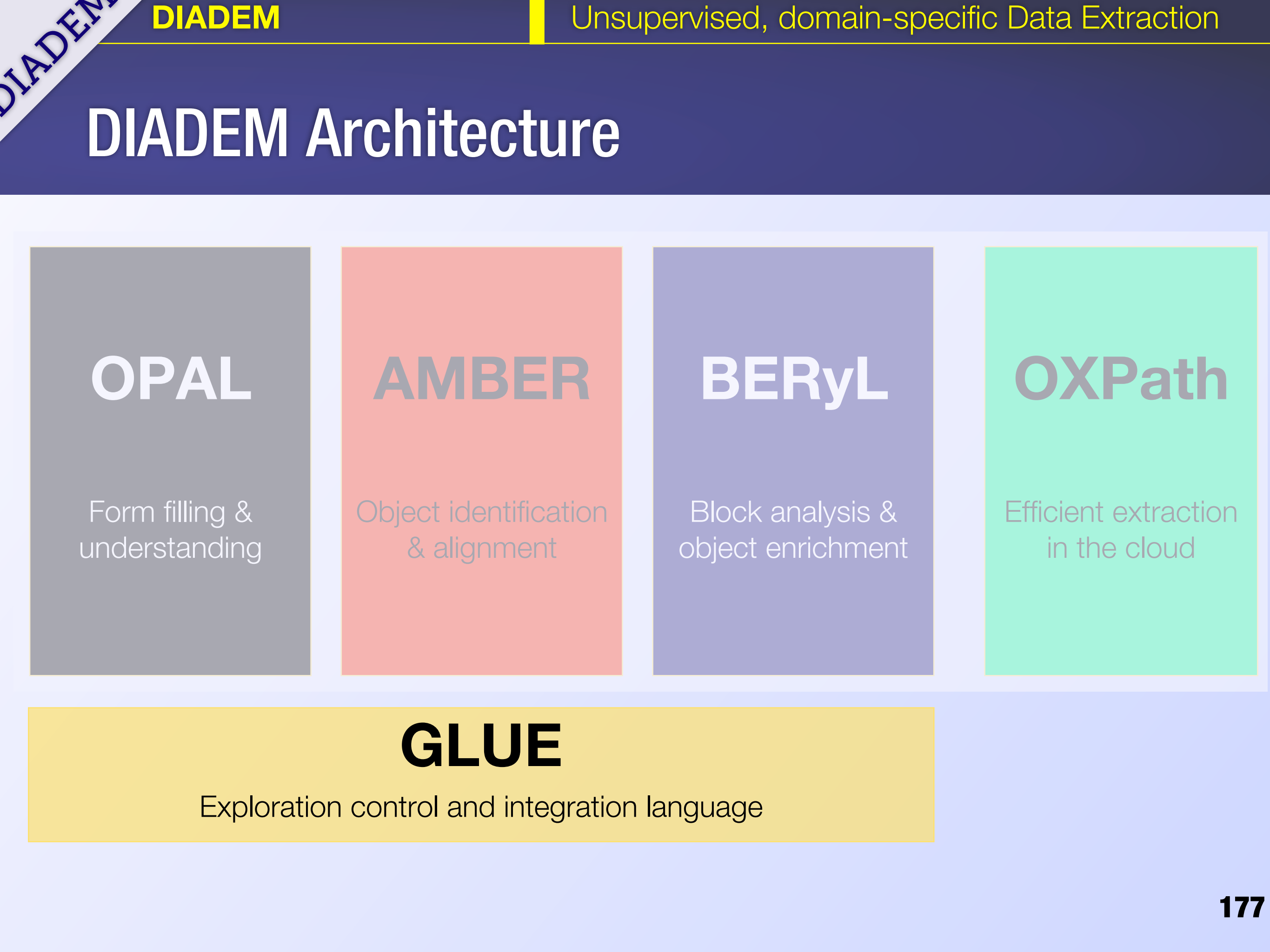
**GLUE**

Exploration control and integration language

**OXPath**

Efficient extraction  
in the cloud





# DIADEM Architecture

**OPAL**

Form filling & understanding

**AMBER**

Object identification & alignment

**BERyL**

Block analysis & object enrichment

**OXPath**

Efficient extraction in the cloud

**GLUE**

Exploration control and integration language



# Summary

- Examples of knowledge (and its representation) in DIADEM
  - observational:** clues for price (“looks like a price”) and location
    - representation: **Gazetteers, JAPE** rules, WEKA classifiers & **Datalog<sup>¬,Agg</sup> rules**
  - phenomenological:** a real estate record and its attributes
    - representation: **Datalog<sup>¬,Agg,±</sup> rules**
  - ontological:** constraints for real estate form
    - representation: template language on top of **Datalog<sup>¬,Agg,±</sup> rules**
  - script:** strategy for exploring post-form pages
    - representation: modularised **Datalog<sup>¬,Agg</sup> rules**



	Knowledge			
	– (site-specific)	Schema	Instances	Patterns
Visual	<div>SXPath</div> <div>XPath + Visual Relations</div>	<div>SCRAP</div> <div>Robust wrappers by schema</div>	<div>WRAPPO</div> <div>Visual patterns and annotations</div>	
	<div>OXPath</div> <div>XPath + Actions</div>	<div>Lixto</div> <div>Visual designer + expressive extraction language Elog</div>	<div>DIADEM</div> <div>Form understanding, object identification, and exploration driven by domain knowledge</div>	



# Data Extraction in the Big Data Age

- Data extraction is crucial to make “good” data
  - from heterogeneous, legacy web sources
- But: not only about legacy
  - who really thinks that everyone will be publishing “good” RDF soon
- **Data extraction complements the Semantic Web**
  - by providing access to vast stores of information
    - focused on “transient” data (e.g., product offers) rather than
    - “static”, common-sense, Wikipedia-like data where human curation is feasible
  - and also profits from it through the
    - availability of high quality **instance** collections
    - availability of **scalable ontology** reasoners



# Where are we?

- **Known knowns:** we know **what** and **how**
  - site-specific or supervised data extraction
- **Known unknowns:** we know **what**
  - templates need to be discovered
  - but: what we are interested in is known
- **Unknown unknowns:**
  - where we don't even know what we are looking for
  - never-ending learning of domain concepts
  - semi-supervised



# Where are we?

- **Known knowns:** we know **what** and **how**
  - site-specific or supervised data extraction
- **Known unknowns:** we know **what**
  - templates need to be discovered
  - but: what we are interested in is known
- **Unknown unknowns:**
  - where we don't even know what we are looking for
  - never-ending learning of domain concepts
  - semi-supervised

