

A Possible Future for The Future of Web Search



Ricardo Baeza-Yates

**VP of Yahoo! Research
for EMEA & LatAm**

Yahoo! Research

- Mission: to develop the world-class science that will deliver the next generation of businesses to the company



- Established an open culture of collaboration with peers from academic and research institutions.

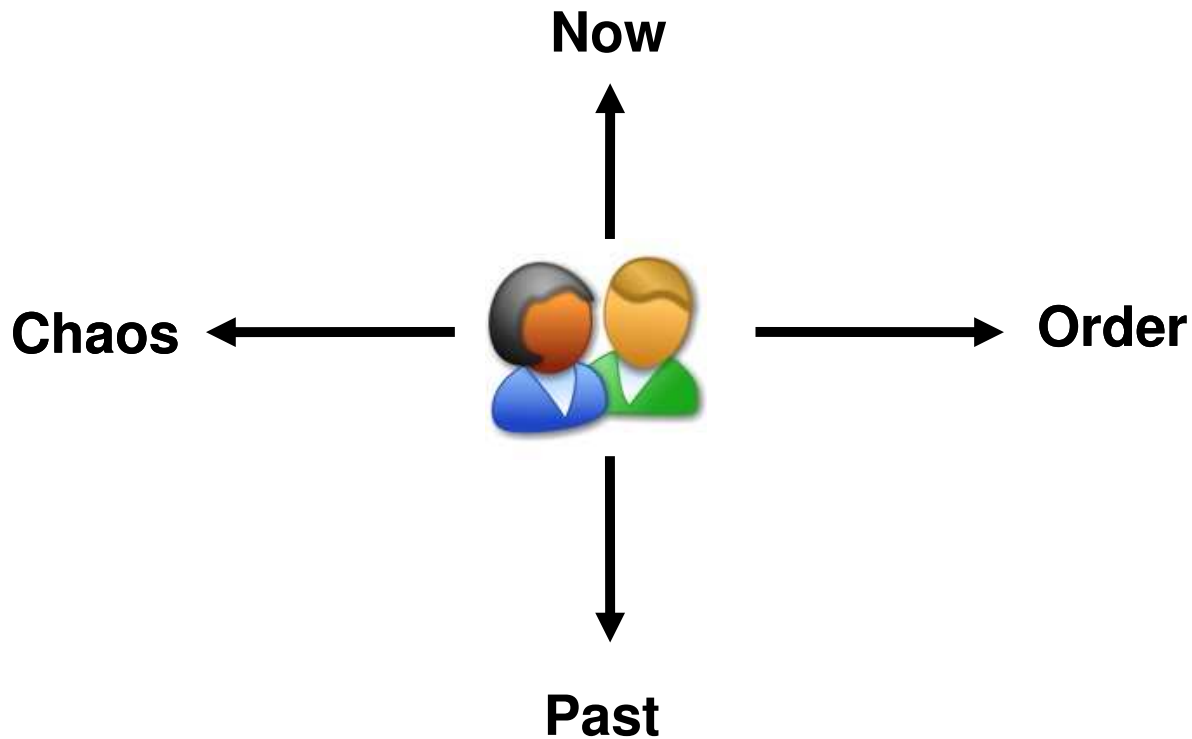
Conclusions

- Web search is no longer about document retrieval
 - Means for web-mediated goals
- New breed of search experiences
 - Demands search ecosystem combining content with intent
 - Exploring the Wisdom of Crowds behind the Web 2.0

Trends

- User Generated Content
 - Massive (quality vs. Quantity)
 - Social Networks
 - Real time (people + sensors)
- Impact
 - Fragmentation of ownership
 - Fragmentation of access (longer tail)
 - Fragmentation of right to access
- Viability
 - Business model based in advertising (?)

What we really want?



Search is Evolving

- Already, more than a list of docs
- Moving towards identifying a user's task
- Enabling means for task completion
- New experiences based on the Web 2.0
- Challenges: on-line, scalability

More complete information in one search

This screenshot shows a Yahoo! search result for "legal sea foods boston ma". The search bar at the top contains the query. Below the search bar, there are navigation links for "Web", "Images", "Video", "Local", and "Shopping". The main content area is divided into several sections:

- Shortcuts:** A red box highlights a "Legal Sea Foods near Boston" section. It includes a map of Boston with a red pin and a list of three nearby locations with their addresses and phone numbers.
- Deep Links:** A red box highlights a section with multiple search results for "Legal Sea Foods". Each result includes a snippet of text and a link to the full page.
- Enhanced Results:** A red box highlights a section with detailed information for two specific locations. Each entry includes a star rating, address, phone number, and a small thumbnail image of the restaurant's interior or food.

On the right side of the page, there are "SPONSOR RESULTS" for "Legal Sea Food" and "Legal Sea Foods Boston".

Shortcuts

Deep Links

Enhanced Results

This screenshot shows a Yahoo! Local search result for "legal sea foods" in Boston, MA. The search bar at the top contains the query. Below the search bar, there are navigation links for "City Guide", "My Local", and "Directory". The main content area is divided into several sections:

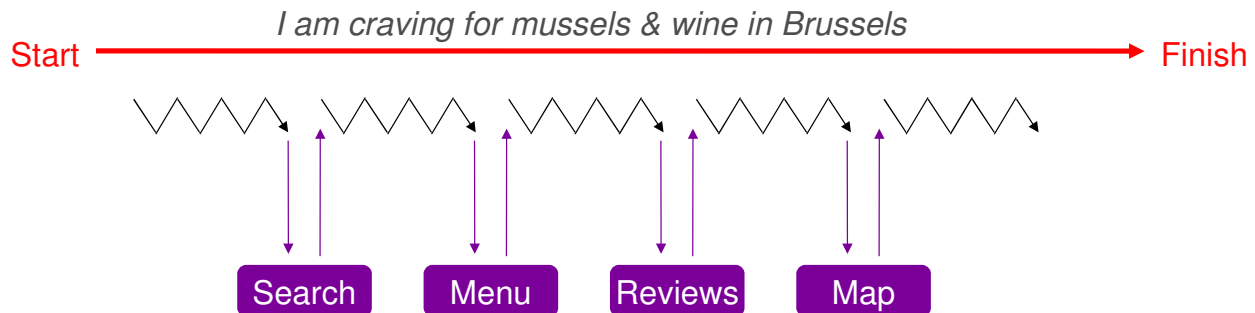
- Search Filters:** A section with "Narrow your search by:" and "Sorted By:" options.
- Search Results:** A list of search results for "Legal Sea Foods Restaurants". Each result includes a star rating, address, phone number, and a small thumbnail image of the restaurant's interior or food.
- Map View:** A section with a map of Boston showing the location of the search results. The map includes a scale bar and a "See Map View" button.

On the right side of the page, there are "SPONSOR RESULTS" for "Legal Sea Food" and "Legal Sea Foods Boston".

Search: Content vs. Intent

Premise:

- People don't want to search
- People want to get tasks done and get straight to their answers



11

Net

- We move from a web of pages to a **web of objects**
- Objects are **people, places, businesses, restaurants ...**
- Objects have attributes
 - Missing, noisy, etc.
- Intents are satisfied by presenting **objects and attributes**
- Attributes define faceted search

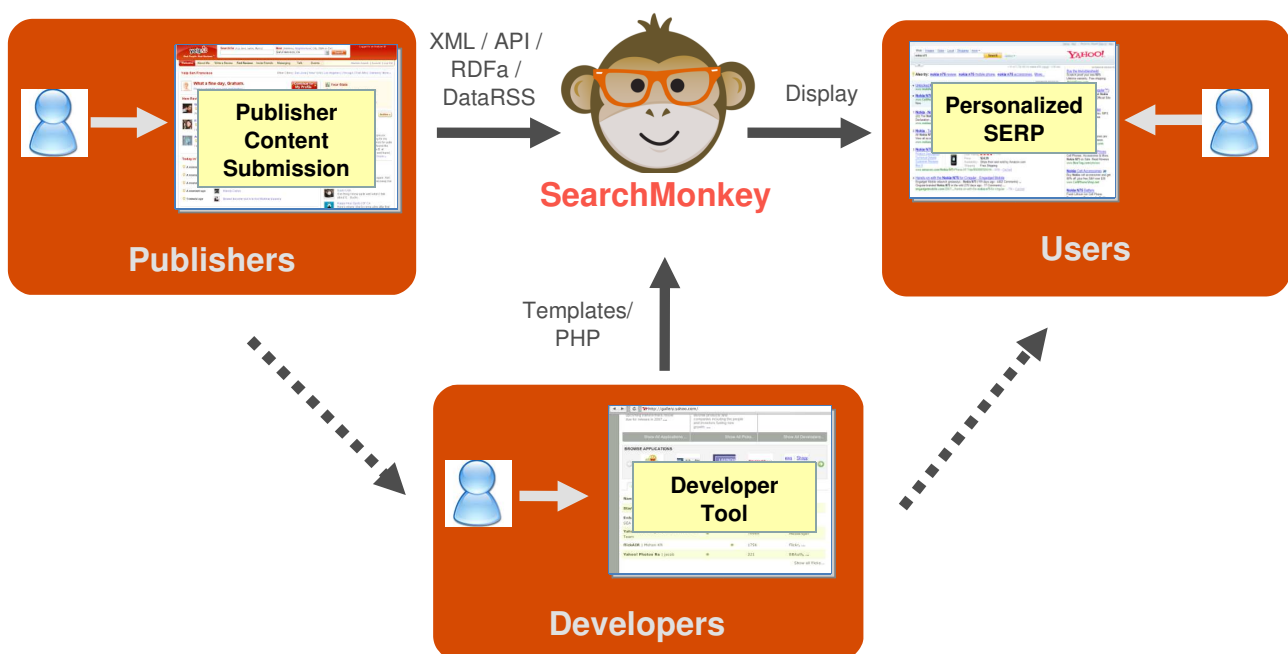
12

How do we get structured objects/attributes?

- Web Content
 - Metadata/Taxonomies/Folksonomies
 - Machine learning techniques
 - Classification/Extraction/Semantic Web
- Web Usage
 - Implicit relations
- Building out an open ecosystem
 - Publishers have incentives to contribute

13

The SearchMonkey Ecosystem



Opening search - what does it mean?

The screenshot shows a Yahoo! search results page for the query "art of pizza chicago". The search bar at the top contains the text "art of pizza chicago" and a "Search" button. Below the search bar, there are several search results. The first result is "The Art of Pizza - Chicago, IL - Citysearch.com", which includes a brief description and a link to the website. Other results include "The Art of Pizza - Lakeview - Chicago, IL 60657" and "The Art of Pizza - Metromix Chicago". The page also features a "SPONSOR RESULTS" section on the right side, which includes "Chicago Pizza Chicago" and "Nick & Bruno's Pizzeria". At the bottom of the page, there is a "YAHOO!" logo and a "Search" button.

- Clear win for: developers, site owners, users and Yahoo!

- Go from "to-do" to "done"

BEFORE

[The Art of Pizza - Lakeview - Chicago, IL 60657](#)

... that say they don't like Chicago deep dish pizza have never had Art of Pizza. ... Top 10

Chicago meals, I would include an Art of Pizza Deep Stuffed ...

www.acmerreviews.com/biz/the-art-of-pizza-chicago - 145k - Cached

AFTER

[The Art of Pizza - Lakeview - Chicago, IL 60657](#)



Reviews

Photos

Send to a friend

Send to Phone

Ratings: ★★★★★

Address: 3033 N Ashland Ave, Chicago, IL

Phone: (773) 327-5600

Price Range: \$

www.acmerreviews.com/biz/the-art-of-pizza-chicago - 145k - Cached



Site Owners

The screenshot shows a Yahoo! search results page for the query "raspberry chocolate mousse recipe". The search bar at the top contains the text "raspberry chocolate mousse recipe" and a "Search" button. Below the search bar, there are several search results. The first result is "Raspberly White Chocolate Mousse Recipe - French Raspberly White...", which includes a brief description and a link to the website. Other results include "Raspberly Chocolate Mousse Cake Recipe - AcmeRecipe" and "Lemon Raspberly White Chocolate Mousse Cake - Allrecipes". The page also features a "SPONSOR RESULTS" section on the right side, which includes "Chocolate Recipe" and "Delicious Cookie Recipe". At the bottom of the page, there is a "YAHOO!" logo and a "Search" button.

- Straight to the answers
- Increase in quality of site traffic
- Fosters loyalty and engagement

BEFORE

[Raspberly Chocolate Mousse Cake Recipe - AcmeRecipe](#)

This award winning Recipe is simple to make with fresh ingredients. This cake can be served topped with chocolate shavings, raspberries and mint...

www.acmerrecipe.com/desserts/01223 - 49k - Cached



AFTER

[Raspberly Chocolate Mousse Cake Recipe - AcmeRecipe](#)



Print Recipe

Shopping List

Nutritional Info

Related recipes

User Rating: ★★★★★ (18)

This award winning recipe is simple to make with fresh ingredients. This cake can be served topped with chocolate shavings, raspberries and mint. ...

www.acmerrecipe.com/desserts/01223 - 49k - Cached

Enhanced Results and Infobars:
Ratings and reviews, images, deep links, and other name-value pairs.

The Wisdom of Crowds

- James Surowiecki, a *New Yorker* columnist, published this book in 2004
 - “Under the **right** circumstances, groups are remarkably intelligent”
- Importance of diversity, independence and decentralization

Aggregating data

“large groups of people are smarter than an elite few, no matter how brilliant—they are better at solving problems, fostering innovation, coming to wise decisions, even predicting the future”.

18

The Wisdom of Crowds

- Crucial for Search Ranking
- Text: Web Writers & Editors
 - not only for the Web!
- Links: Web Publishers
- Tags: Web Taggers
- Queries: All Web Users!
 - Queries and actions (or no action!)

19

Query: paris

- locations: europe, france, ile de france, london, montmartre
- subjects: arc de triomphe, art, eiffel tower, louvre, museum, notre dame, seine
- activities: travel, trip
- time: 2005, 2006, 2007, vacation

Help

You can refine your query using the tag-cloud on the left

- Use **tag** to post new query using tag
- Use **+** to add terms to query
- Use **x** to remove terms from query

Photo Results



Photo Details



Paris - Gare du Pont Cardinet - 28-07-2007 - 9h03
Taken by: [Panoramas](#)
[View photo on Flickr](#)
Tags: bridge sky panorama paris france reflection rain de puente pod gare perspective pluie ponte reflet most ciel zebra pont brug angkor passage brücke chemin petite 1925 1920 fer passerelle lopburi rer panneaux köprü narai ceinture piétons angkorian most cardinet batignolles үѳүрү smartblend indicateurs étiennecazin angkorien

Endless image browsing

Faceted Image Search

Web | Images | Video | Local | Shopping | more

brussels Search Options

YAHOO!

Show only: Wallpaper size Black & White [More filters](#)

1 - 20 of 2,315,140 for brussels - 0.08 s | SafeSearch is ON [turn off](#)

Explore **Bruxelles, Belgium**

- Statue of the Pissing Boy**
62 images
- Central Square**
71 images
- Royal Palace**
403 images
- Place du Jeu de Balle**
456 images
- Autoworld**
41,247 images
- Europalia**
210 images
- Belgian Royal Museums of Fine**

brussels.jpg 700 x 497 | 175k trekearth.com

brussels grandplace.jpg 800 x 600 | 117k trekearth.com

brussels grand place Jason's Travels flickr.com

Brussels.jpg 1536 x 1024 | 424k isavetravel.com

great market brussels belgium gbatistini flickr.com

couple of days in brussels Rui M Leal flickr.com

xmas in brussels ii.jpg 785 x 572 | 199k trekearth.com

great market brussels belgium gbatistini flickr.com

man piss in brussels belgium mambol935 flickr.com

a rainy day in brussels kimicon flickr.com

Done

Visual Diversity

- **Topical diversity**

Query: "Jaguar"



- **Visual diversity**

Query: "Jaguar X-type"



- **Other dimensions: spatial, temporal, social**

Exploit Visual Annotations

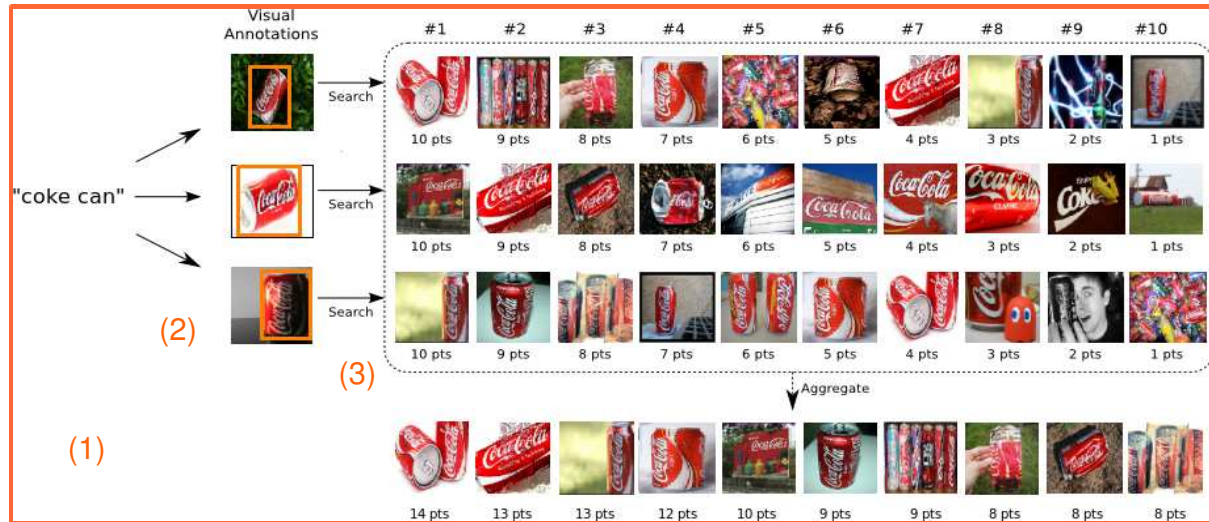
Flickr allows another kind of annotations (notes)

- Associate **text** with **visual area**
- Highly relevant to content
→ **Visual Annotation**
- Valuable to learn different visual representations of an object



Use visual annotations (text & image) and rank aggregation to improve retrieval

Better Image Search



Tags are better than image features

But together they are much better as they are complementary

SEMEDIA

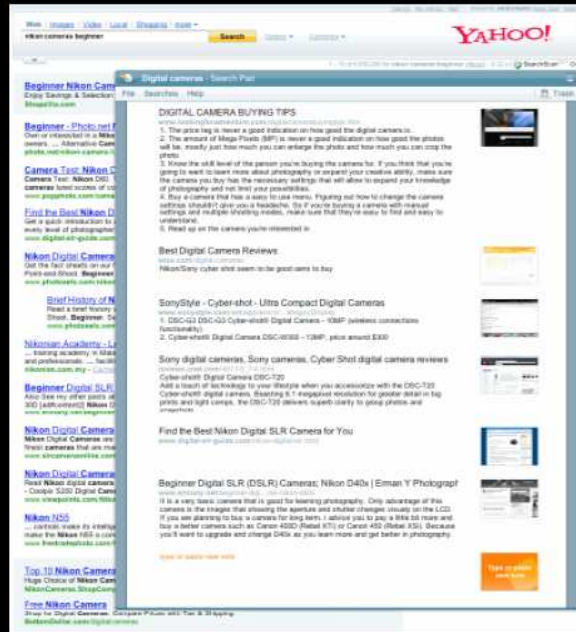
- Search Environments for Media
 - Focused on image & video search
 - www.semedia.org



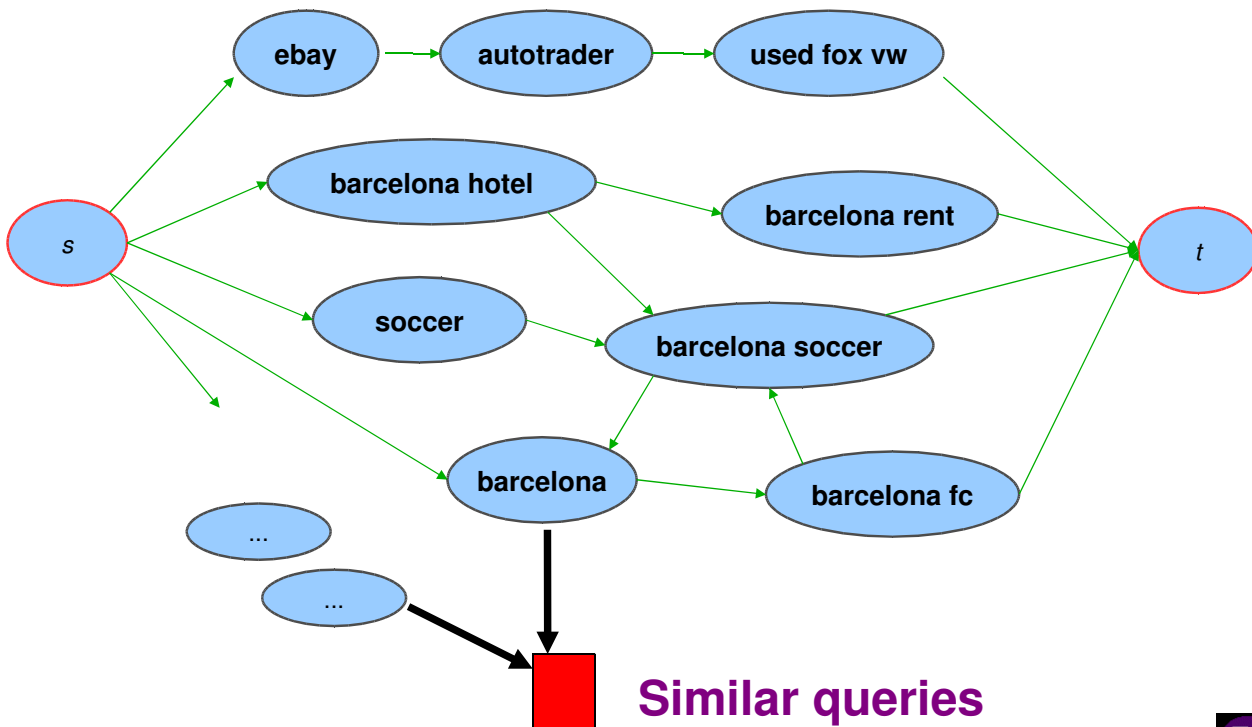
SearchPad

“... keeps track of search query terms ... when it **detects a trend**, offers to save the result in an online document.”

CNET



Query-Flow Graph



Exploiting Microformats

The screenshot shows a Yahoo! search results page for the query "ivan herman". The browser's address bar shows the URL: `http://ybcn-svr6.barcelona.corp.yahoo.com:8080/inf/search.do?ip=ivan+herman&y=Search&f=lo`. The search results list several items, with annotations pointing to specific features:

- Geolocation:** A red box highlights the "Geolocation" label in the top right corner of the browser window.
- Rich abstract:** A red box highlights the "Rich abstract" label on the left side, pointing to the first search result, "Ivan's blog".
- Related pages based on metadata:** A red box highlights the "Related pages based on metadata" label on the left side, pointing to the "is primary topic of" section for the first result.
- Events from personal calendar, Conferences, and bio from LinkedIn:** A red box highlights this label on the right side, pointing to a map view of search results showing various locations and events like "Ivan's BD", "Robo's BD", "My BD", and "David's BD".

Bridging implicit and explicit metadata

The screenshot shows the left sidebar of a Wikipedia article for Pablo Ruiz Picasso. It includes a search box, a "Go" button, and a "Search" button. Below that is a "toolbox" with links for "What links here", "Related changes", "Upload file", "Special pages", "Printable version", "Permanent link", and "Cite this article". At the bottom is a "languages" section with a list of language options including Afrikaans, العربية, বাংলা, Bân-lâm-gú, Bosanski, Brezhoneg, Български, Català, Český, Cymraeg, and Dansk.

Pablo Ruiz Picasso (October 25, 1881 – April 8, 1973), often referred to simply as **Picasso**, was a Spanish painter and sculptor. His full name is **Pablo Diego José Francisco de Paula Juan Nepomuceno María de los Remedios Cipriano de la Santísima Trinidad Clito Ruiz y Picasso**.^[1] One of the most recognized figures in 20th century art, he is best known as the co-founder, along with Georges Braque, of cubism.

Contents [show]

Biography [edit]

Pablo Picasso was born in **Málaga, Spain**, the first child of José Ruiz y Blasco and María Picasso y López. He was christened with the names Pablo, Diego, José, Francisco de Paula, Juan Nepomuceno, Maria de los Remedios, and Cipriano de la Santísima Trinidad.^[2] Picasso's father was a painter whose specialty was the naturalistic depiction of birds and who for most of his life was also a professor of art at the School of Crafts and a curator of a local museum. The young Picasso showed a passion and a skill for drawing from an early age; according to his mother,^[3] his first word was "piz," a shortening of *lápiz*, the Spanish word for pencil.^[4] It was from his father that Picasso had his first formal academic art training, such as figure drawing and painting in oil. Although Picasso attended art schools throughout his childhood, often those where his father taught, he never finished his college-level course of study at the Academy of Arts

The screenshot shows the infobox for Pablo Picasso. It features a portrait of Picasso wearing a flat cap, with the caption "Picasso (January 1962)". Below the portrait is a table of personal information:

Birth name	Pablo Diego José Francisco de Paula Juan Nepomuceno María de los Remedios Cipriano de la Santísima Trinidad Martyr Patricio Clito Ruiz y Picasso
Born	October 25, 1881 Málaga, Spain
Died	April 8, 1973 (aged 91) Mougins, France

Correlator: Relations in Wikipedia

Yahoo | My Yahoo! | Mail | More

Report it

Correlator

religious war

1552: A small Protestant victory in 1552 forced Charles , weary from three decades of war , to sign the Pe

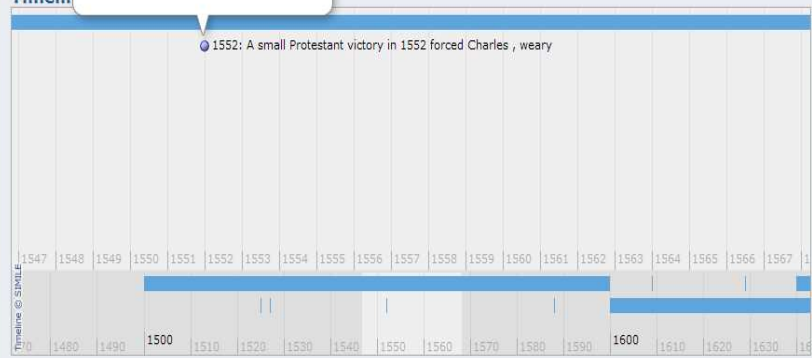
A small Protestant victory in 1552 forced Charles , weary from three decades of war , to sign the Peace of Passau , which granted some freedoms to Protestants and ended all of Charles hopes at religious unity within his empire .

Tue, 01 Jan 1552 00:00:00 GMT



Events

Timeline



Events in the timeline

16th century

(From **W**Monastic state of the Teutonic Knights) "During the Protestant Reformation , endemic religious upheavals and wars occurred , and , in 1525 , during the aftermath of the Polish-Teutonic War (1519-1521) , the last Grand Master of the Teutonic Knights , Albert of Brandenburg-Ansbach , a member of a cadet branch of the House of Hohenzollern , resigned his position , adopted the Lutheran faith and assumed the title of " Duke of Prussia . " In a deal partially brokered by Martin Luther , the Duchy of Prussia became the first Protestant state . "

Correlator: Relations in Wikipedia

Search

Wikipedia Names Places Events Concepts News Answers



Locations related to 'religious war'

Town

[Jerusalem](#)

Country

[England](#)

Country

[France](#) [Iraq](#) [Germany](#) [United States](#) [Poland](#)

[Pakistan](#) [Spain](#) [Netherlands](#) [Israel](#) [Russia](#)

(From **W**Ulster-Scots) "There was a generation of calm in **Ireland** thereafter until another civil war broke out , again on ethnic and religious lines , in 1689 . "

(From **W**Williamite) "However , for Williamites in England , Scotland and **Ireland** , William was seen as the guarantor of civil and religious liberty and the Protestant monarchy against Catholic absolutism . "

(From **W**Kingdom of Scotland) "In 1603 James VI King of Scots , became King James of England thus Scotland entered into a personal union with England and **Ireland** . "

Correlator: Relations in Wikipedia

The screenshot shows the Correlator interface with the search query "protestant england". The interface includes a search bar, navigation icons for Wikipedia, Names, Places, Events, Concepts, News, and Answers, and a "Report it" link. The main content area is divided into two columns. The left column, titled "Names related to 'protestant england'", lists related terms such as "Protestant Reformation", "Protestant England - England", "Queen Mary I - Mary - Mary I", "Protestant Church", "Roman Catholicism - Catholicism", "James I", "William", "Parliament", "Bible", "Queen Elizabeth I - Elizabeth I - Elizabeth", "Roman Catholic Church - Catholic Church", "Cromwell", and "Evangelical Church". Below the list, there are three entries with brief descriptions: "Catholicism" is probably related to protestant england; Henry Morse, born a Protestant in 1595, converted to Catholicism; and the Family Compact, which supported the Church of England. The right column displays a network graph with "protestant england" at the center, connected to various related terms like "Roman Bible", "Roman Catholicism", "Catholic Church", "Elizabeth I", "Protestant Reformation", "James I", "William", "Evangelical Church", "Queen Mary I", and "Parliament".

New Pages

- For topics without a Wikipedia page, Correlator creates a “synthetic page” with an overview of the topic
- Query:
 - art deco chicao
- Synthetic page:
 - Defines Art Deco
 - Defines Chicago
 - Shows relations between Art Deco and Chicago

The screenshot shows the Correlator interface with the search query "art deco chicao". The interface includes a search bar, navigation icons for Wikipedia, Names, Places, Events, Concepts, News, and Answers, and a "Report it" link. The main content area is divided into two columns. The left column, titled "Art Deco", provides a detailed overview of the Art Deco movement, including its history from 1932 to 1939, its influence on architecture, interior design, and industrial design, and its popularity in the 1930s. It also mentions the California Art Deco Society and Chicago landmarks. The right column, titled "Chicago", provides a detailed overview of the city, including its location in Illinois, its population, and its status as a major transportation hub. It also mentions the city's incorporation in 1837 and its location at the site of a portage between the Great Lakes and the Mississippi River watershed. Below the main content, there are two sections: "Category: 1930 architecture" and "Category: Skyscrapers in Chicago". The "1930 architecture" section lists several buildings, including the Merchandise Mart, the Chicago Board of Trade Building, and the Grace Building (Sydney). The "Skyscrapers in Chicago" section lists the Chicago Board of Trade Building, the LaSalle National Bank Building, and the Four Seasons Hotel Chicago. Each section includes a brief description of the building and a link to view more entries from the category.

Conclusions

- Web search is no longer about document retrieval
 - Means for web-mediated goals
- New breed of search experiences
 - Ecosystem: SearchMonkey
 - Usage: SearchPad (beta)
 - Usage: TagExplorer (demo)
 - Usage & tagging: Faceted Image Search (beta)
 - Content: Correlator (demo)
- **Virtuous cycle from people to people**



Questions?

rbaeza@acm.org

<http://search.yahoo.com>

<http://research.yahoo.com>

<http://sandbox.yahoo.com>