

# Big Data on the Web

## Web Information Sources

- Wikipedia
- Newspapers
- Social networks
- Scientific data
- Online shops
- ...

## Information Extraction?

# Keyword Search



mozart



All

Videos

Images

News

Maps

More ▾

Search tools

About 83,000,000 results (0.65 seconds)

## Wolfgang Amadeus Mozart - Wikipedia, the free encyclopedia

[https://en.wikipedia.org/wiki/Wolfgang\\_Amadeus\\_Mozart](https://en.wikipedia.org/wiki/Wolfgang_Amadeus_Mozart) ▾

Wolfgang Amadeus **Mozart** baptised as Johannes Chrysostomus Wolfgangus Theophilus **Mozart**, was a prolific and influential composer of the Classical era, ...

[Salzburg](#) - [Death of Wolfgang Amadeus ...](#) - [Constanze Mozart](#) - [List of compositions](#)

## Wolfgang Amadeus Mozart — Wikipédia

[https://fr.wikipedia.org/wiki/Wolfgang\\_Amadeus\\_Mozart](https://fr.wikipedia.org/wiki/Wolfgang_Amadeus_Mozart) ▾ [Translate this page](#)

Joannes Chrysostomus Wolfgangus Theophilus **Mozart**, ou Wolfgang Amadeus **Mozart** (en allemand : ['vɔlfɡaŋ ama'deus 'mo:tsaʁt]) né à Salzbourg, ...

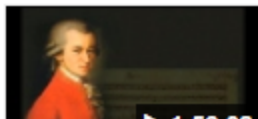
[Mémoire eidétique](#) - [Constance Mozart](#) - [Requiem de Mozart](#) - [Leopold Mozart](#)

## Wolfgang Mozart - Pianist, Composer - Biography.com

[www.biography.com/people/wolfgang-mozart-9417115](http://www.biography.com/people/wolfgang-mozart-9417115) ▾

A prolific artist, Austrian composer Wolfgang **Mozart** created a string of operas, concertos, symphonies and sonatas that profoundly shaped classical music. ... [Wolfgang Amadeus Mozart - Mini Biography](#) (TV-14; 3:05) Wolfgang Amadeus **Mozart** learned the piano at the age of three, and soon ...

## The Best of Mozart - YouTube



<https://www.youtube.com/watch?v=Rb0UmrCXxVA> ▾

Jan 8, 2013 - Uploaded by HALIDONMUSIC

[Wolfgang Amadeus Mozart](#) - The Best of **Mozart**. ... The Best of Classical Music -

# Keyword Search



mozart



All

Videos

Images

News

Maps

More ▾

Search tools

About 83,000,000 results (0.65 seconds)

## Wolfgang Amadeus Mozart - Wikipedia, the free encyclopedia

[https://en.wikipedia.org/wiki/Wolfgang\\_Amadeus\\_Mozart](https://en.wikipedia.org/wiki/Wolfgang_Amadeus_Mozart) ▾

Wolfgang Amadeus Mozart baptised as Johannes Chrysostomus Wolfgangus Theophilus Mozart, was a prolific and influential composer of the Classical era, ...

[Salzburg](#) - [Death of Wolfgang Amadeus ...](#) - [Constanze Mozart](#) - [List of compositions](#)

## Wolfgang Amadeus Mozart — Wikipédia

[https://fr.wikipedia.org/wiki/Wolfgang\\_Amadeus\\_Mozart](https://fr.wikipedia.org/wiki/Wolfgang_Amadeus_Mozart) ▾ [Translate this page](#)

Joannes Chrysostomus Wolfgangus Theophilus Mozart, ou Wolfgang Amadeus Mozart (en allemand : ['vɔlfɡaŋ ama'deus 'mo:tsaʁt]) né à Salzbourg, ...

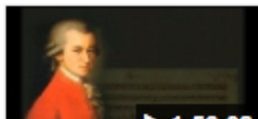
[Mémoire eidétique](#) - [Constance Mozart](#) - [Requiem de Mozart](#) - [Leopold Mozart](#)

## Wolfgang Mozart - Pianist, Composer - Biography.com

[www.biography.com/people/wolfgang-mozart-9417115](http://www.biography.com/people/wolfgang-mozart-9417115) ▾

A prolific artist, Austrian composer Wolfgang Mozart created a string of operas, concertos, symphonies and sonatas that profoundly shaped classical music. ... [Wolfgang Amadeus Mozart - Mini Biography](#) (TV-14; 3:05) [Wolfgang Amadeus Mozart](#) learned the piano at the age of three, and soon ...

## The Best of Mozart - YouTube



<https://www.youtube.com/watch?v=Rb0UmrCXxVA> ▾

Jan 8, 2013 - Uploaded by HALIDONMUSIC

[Wolfgang Amadeus Mozart](#) - The Best of Mozart. ... The Best of Classical Music -

# Knowledge



## Wolfgang Amadeus Mozart



Composer

Wolfgang Amadeus Mozart, baptised as Johannes Chrysostomus Wolfgangus Theophilus Mozart, was a prolific and influential composer of the Classical era, born in Salzburg. Mozart showed prodigious ability from his earliest childhood. [Wikipedia](#)

**Died:** December 5, 1791, [Vienna, Austria](#)

**Influenced:** [Ludwig van Beethoven](#)

**Influenced by:** [Johann Sebastian Bach](#), [Joseph Haydn](#), [George Frideric Handel](#), [Carl Philipp Emanuel Bach](#), [Michael Haydn](#)

### Compositions

[Requiem](#)

[The Magic Flute](#)

# Knowledge Database



Browse using ▾

Formats ▾

Faceted Browse

## About: Wolfgang Amadeus Mozart

An Entity of Type : [agent](#), from Named Graph : <http://dbpedia.org>, within Data Space : [dbpedia.org](#)

Wolfgang Amadeus Mozart (German: [ˈvɔlfɡaŋ amaˈdeːʊs ˈmoːt͡saʁt], English see fn.; 27 January 1756 – 5 December 1791), baptised as Johannes Chrysostomus Wolfgangus Theophilus Mozart, was a prolific and influential composer of the Classical era. Born in Salzburg, Mozart showed prodigious ability from his earliest childhood. Already competent on keyboard and violin, he composed from the age of five and performed before European royalty.

Property	Value
<a href="#">dbo:abstract</a>	<div><ul style="list-style-type: none"><li>Wolfgang Amadeus Mozart (German: [ˈvɔlfɡaŋ amaˈdeːʊs ˈmoːt͡saʁt], English see fn.; 27 January 1756 – 5 December 1791), baptised as Johannes Chrysostomus Wolfgangus Theophilus Mozart, was a prolific and influential composer of the Classical era. Born in Salzburg, Mozart showed prodigious ability from his earliest childhood. Already competent on keyboard and violin, he composed from the age of five and performed before European royalty. At 17, he was engaged as a musician at the Salzburg court, and later he travelled in search of a better position. While visiting Vienna in 1781, he was dismissed from his Salzburg position. He then became a freeman of the capital, where he achieved fame but little financial security. During his final years in Vienna, he composed over 600 works, many acknowledged as pinnacles of symphonic, concertante, chamber, operatic, and choral music. Mozart is one of the most enduringly popular of classical composers, and his influence on subsequent Western art music is profound. Beethoven composed his own early works in the shadow of Mozart, and Joseph Haydn wrote that "posterity will not see again in 100 years". <sup>(en)</sup></li></ul></div>
<a href="#">dbo:alias</a>	<div><ul style="list-style-type: none"><li>Mozart, Johann Chrysostom Wolfgang Amadeus (full name) <sup>(en)</sup></li></ul></div>
<a href="#">dbo:birthDate</a>	<div><ul style="list-style-type: none"><li>1756-01-27 <sup>(xsd:date)</sup></li></ul></div>

# Information Extraction

## Keyword Search

- Ranked Answers – statistics
- Logical explanation of the first answer – database query

The Web has Become Semantic

# Knowledge Representation

## Wikipedia-like Databases

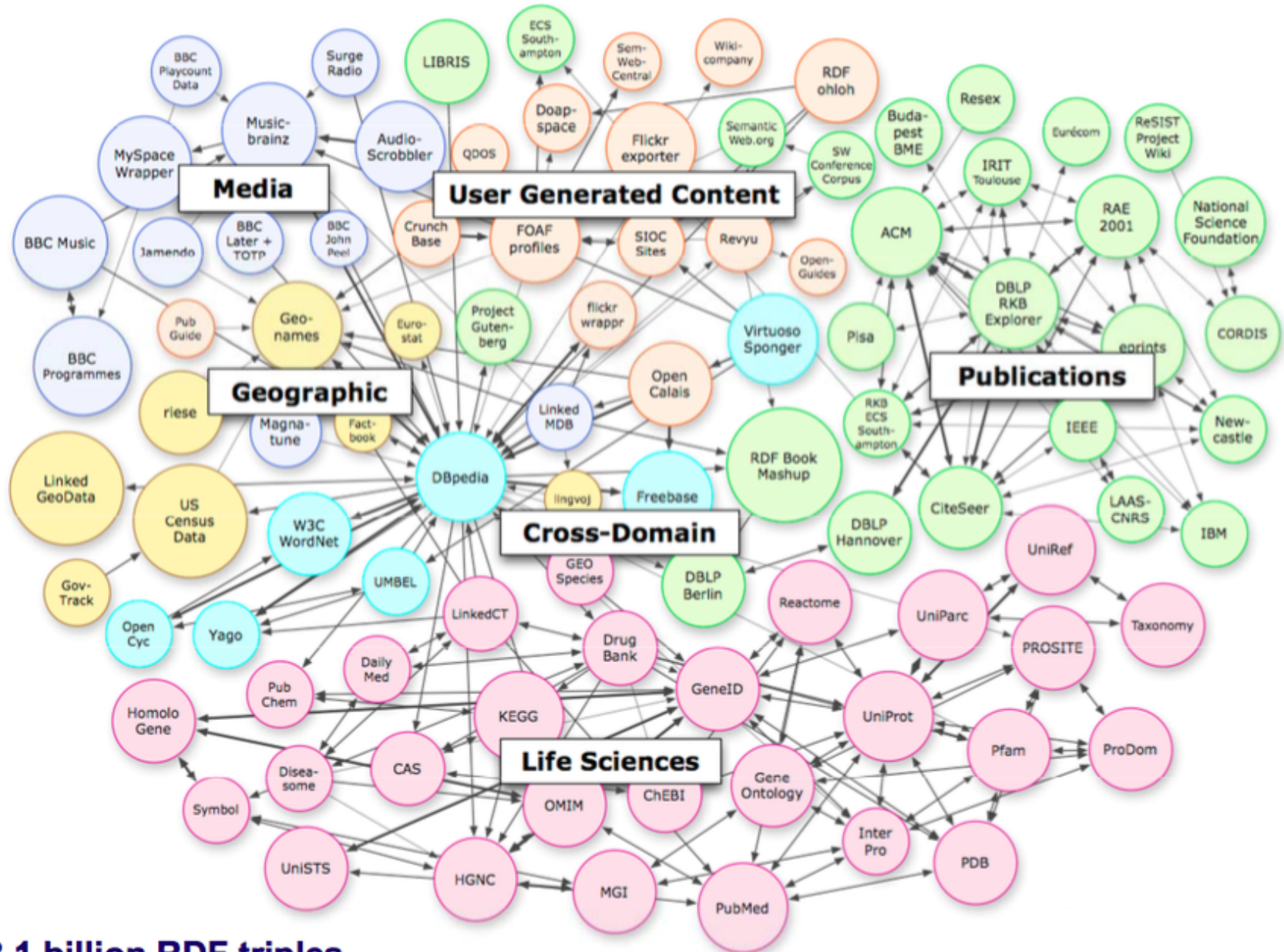
- DBpedia
- Google's knowledge graph
- ...

## RDF Databases

- Open linked data
- Links to DBpedia
- ...



# Linked Open Data



**Over 13.1 billion RDF triples**

**Over 142 million RDF links between data sources**



# Database Queries

## Logic of knowledge

Knowledge graph	=	Database
Logic formula	=	Database query
Logical reasoning	=	Query answering

## First-order logic (Frege 1911)

# Data Mining

## Problem

Many interesting relations are not definable from existing relations by logical queries

## Example

Recommendations

## Discovery of New Relations

From statistics on frequent patterns and existing relations by machine learning techniques



# Bêtise de Cambrai



Bêtises de Cambrai are a French boiled sweet made in the town of Cambrai. "Bêtise" is French for "nonsense" or "stupid mistake" and the sweets are said to have been invented by accident by the son of a confectioner named Afchain. [Wikipedia](#)

**Region or state:** [Cambrai](#)

People also search for

[View 5+ more](#)



[Calisson](#)



[Anise of Flavigny](#)



[Babelutte](#)



[Maroilles cheese](#)



[Coucoug...](#)

# Inria Project-Teams

## Representation and processing of data and knowledge

Knowledge Representation	Graphic Exmo Orpailleur
Databases	<b>Links</b> Dahu Cedar Smis
Data Mining	<b>Magnet</b> LacoDam

**Lille 1: most of the 17 projects of Inria Lille**

# Linked Data

## Collections of Databases

- Links between elements
- Example: open linked data

## Heterogeneous Formats

- Relational databases (Oracle, SAP, ...)
- RDF databases
- Graph databases (Neo4U, ...)
- Key-value stores (QuasardDB, ...)
- NoSQL databases (IBM JAQL, MongoDB, ...)

## Heterogeneous Open Vocabularies

- Ontologies
- Schema mappings

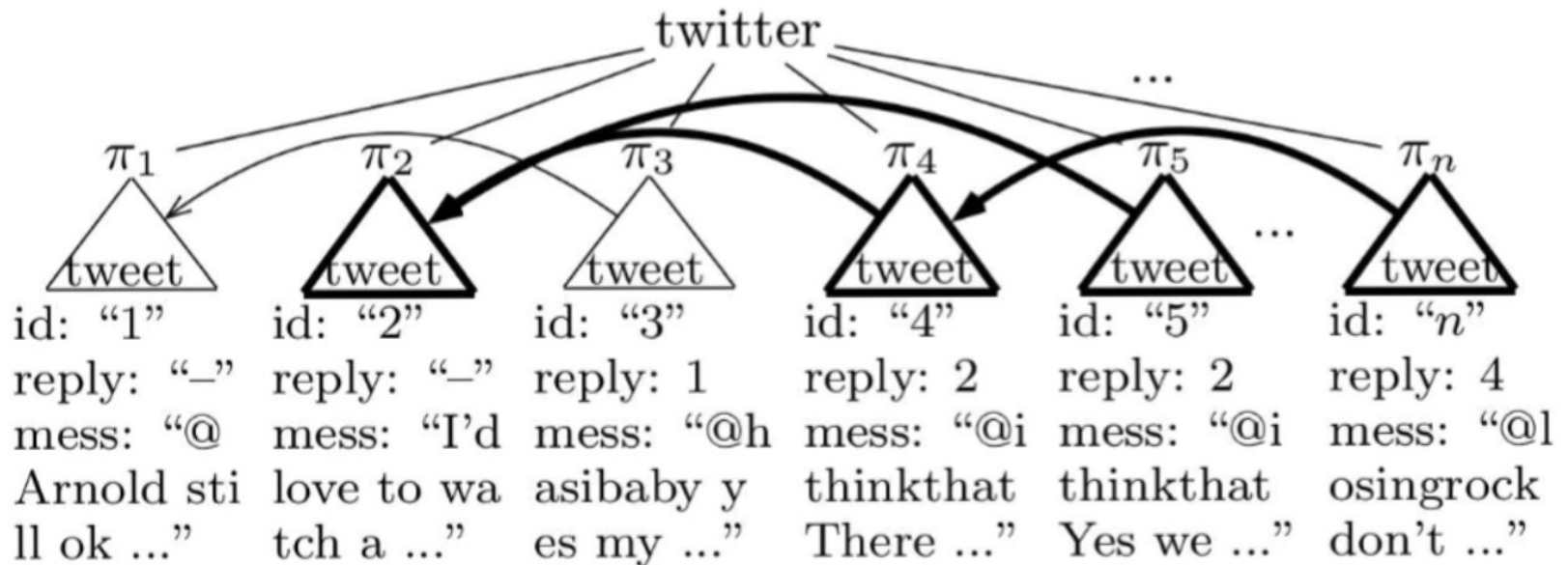
# Links

**Objective: Improve the querying power for linked data**

- Heterogeinity :           Data Integration
- Inconsistency:           Provenance information
- Missing information: Open world assumption
- Dynamicity:           Real time query answering



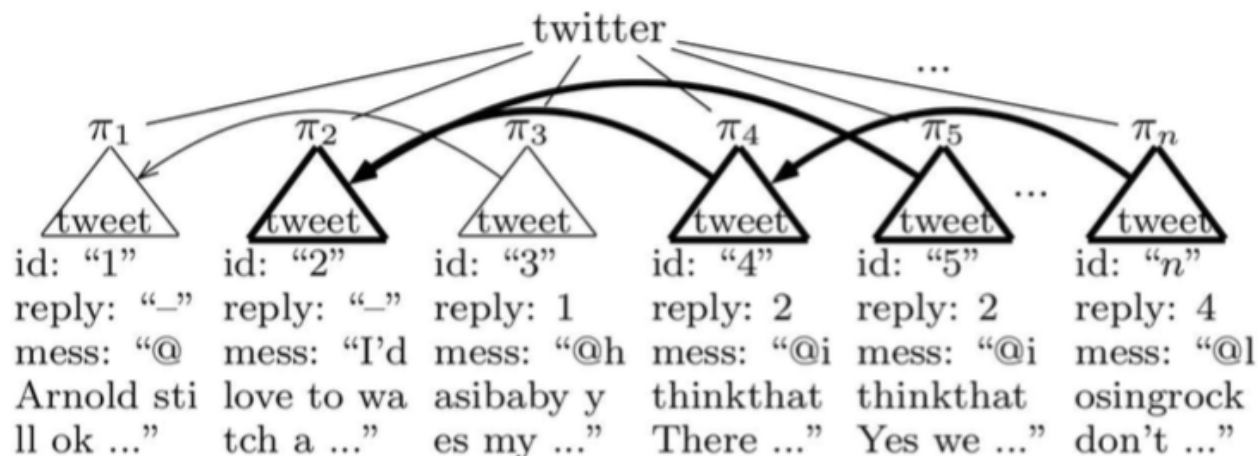
# Example: Twitter Streams



**A Twitter Streams is a dynamic XML database**

**Problem: find and update all dialogs starting with love or hate  
in real time**

# XPath 3.0 Query



$Q_{start}$ : `//tweet[mess[contains(.,'love') or contains(.,'hate')]]`

$Q_{reply}$ : `let $y := . return following-sibling::tweet[reply = $y/id]`

$Q_{nodes}$ :  `$Q_{start} / (Q_{reply})^*$`

$Q_{dialog}$ : `for $y in  $Q_{nodes}$  return ($y/id/text(), $y/mess/text(), $y/reply/text())`

# Full XPath 3.0 Query Evaluation on XML Streams

## Needed

- Low latency
- High time efficiency
- Large coverage

## Difficulties

- How to cover all of XPath 3.0 in streaming mode: data joins, first-order logic, recursive functions, ...
- Optimal latency is impossible in general (PhD Gauwin 2006-09), so approximate optimal latency (PhD Sebastian 2011-16)
- 

## Our Solution

- QuiXPath: Transfer project Inria - Innovimax

# Open Question

**What will be the next killer application  
of big data on the Web?**