Query Operations

Relevance Feedback

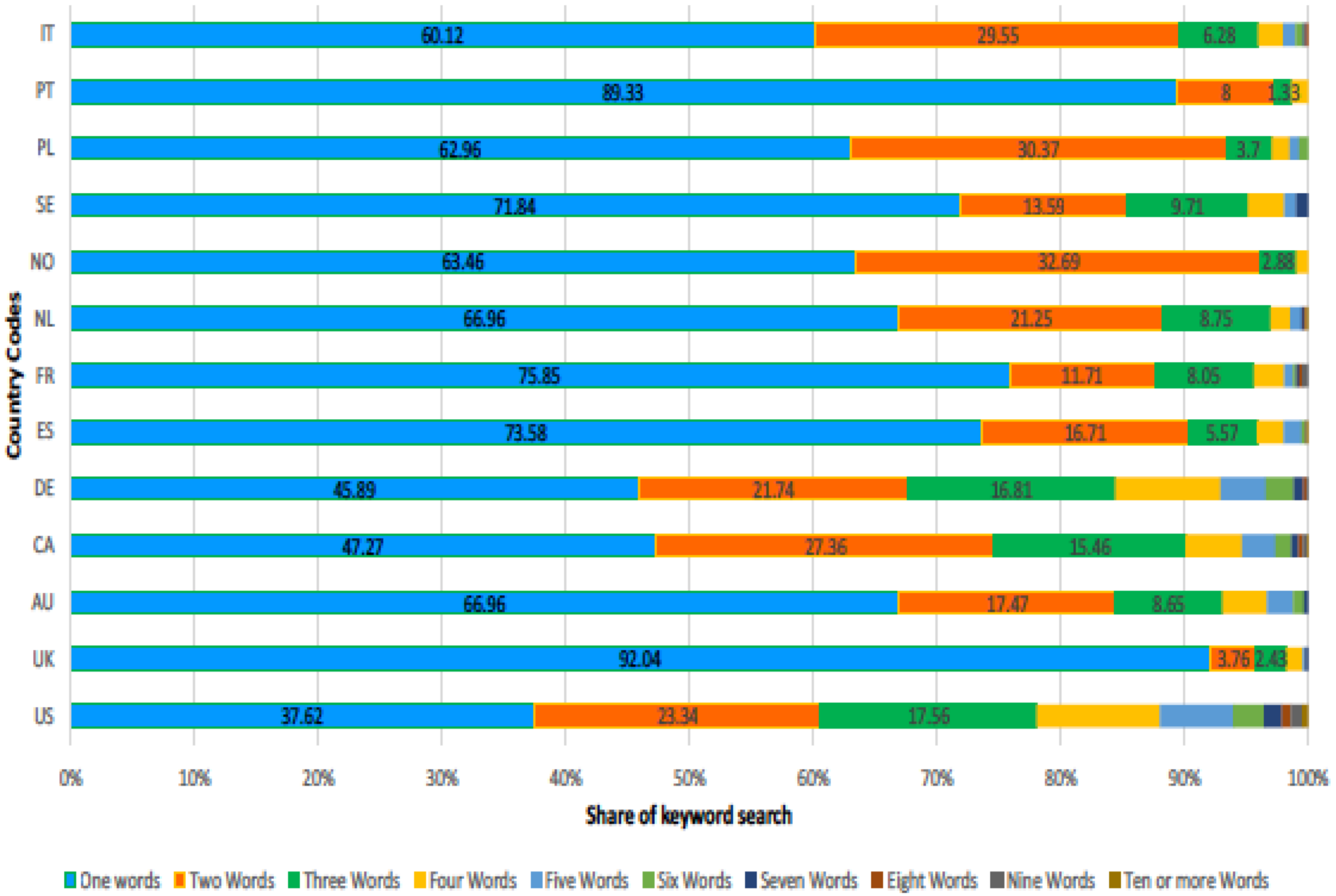
Query Expansion

Query interpretation

Relevance Feedback

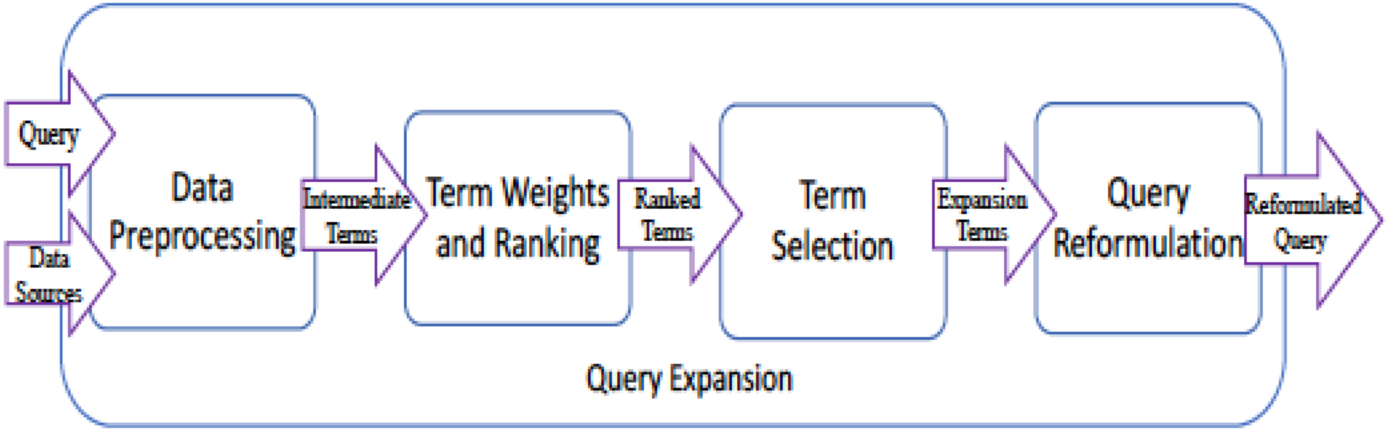
- After initial retrieval results are presented, allow the user to provide feedback on the relevance of one or more of the retrieved documents.
- Use this feedback information to reformulate the query.
- Produce new results based on reformulated query.
- Allows more interactive, multi-pass process.

Statistics on query dimension on web search engines by country



From: Azad Deepak 2017

Workflow of query expansion



From: Azad Deepak 2017

Query expansion methods

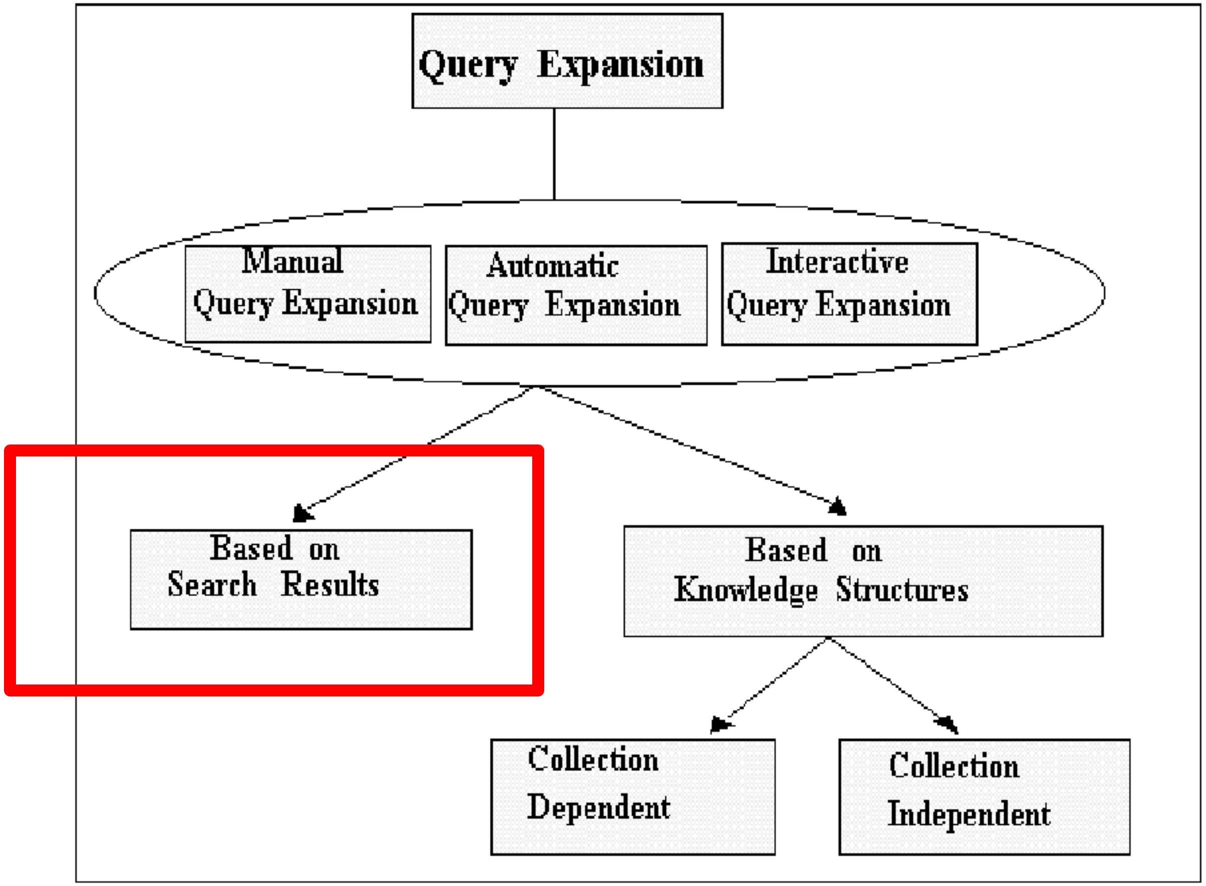
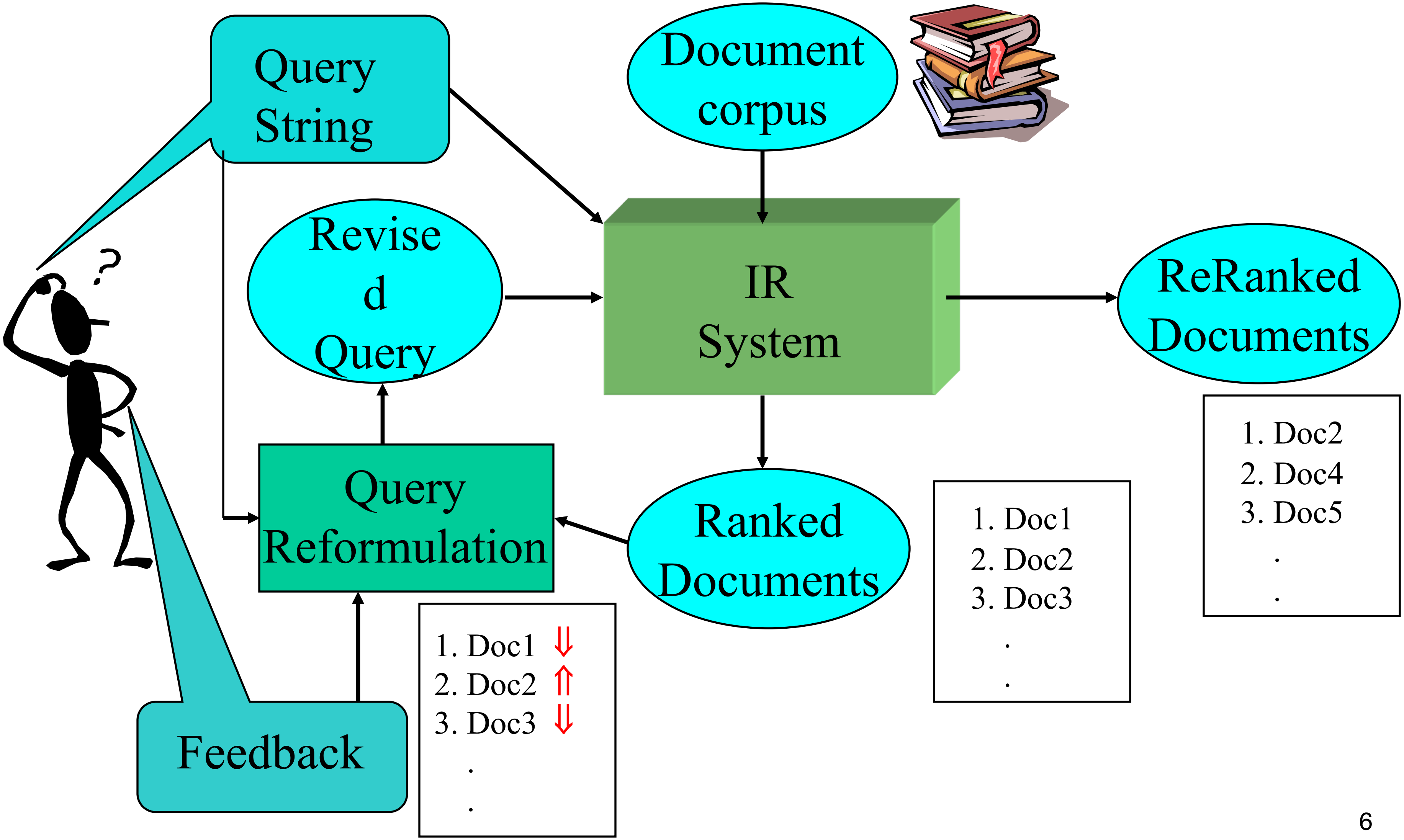


Figure 1: Query Expansion: Methods and Sources

Relevance Feedback Architecture



Query String

Document corpus

Revised Query

IR System

ReRanked Documents

1. Doc2
2. Doc4
3. Doc5

.

.

Query Reformulation

Ranked Documents

1. Doc1
2. Doc2
3. Doc3

.

.

Feedback

1. Doc1 ⇓
2. Doc2 ⇑
3. Doc3 ⇓

.

.

Query Reformulation

- Revise query to account for feedback:
  - Query Expansion: Add new terms to query extracted from relevant documents.
  - Term Re-weighting: Increase weight of terms in relevant documents and decrease weight of terms in irrelevant documents.
- Several algorithms for query reformulation.

Query Reformulation for VSR (vector space retrieval)

- General idea: change query vector using vector algebra:
  - **Add** the vectors for the **relevant** documents to the query vector.
  - **Subtract** the vectors for the **irrelevant** docs from the query vector.
- This adds both positively and negatively weighted terms to the query as well as reweighting the initial terms.

Optimal Query

- Assume that the relevant (to the user's query) set of documents C_r is known.
- Then the best query that ranks <u>all and only</u> the relevant documents at the top is:

$$\vec{q}_{opt} = \frac{1}{|C_r|} \sum_{\forall \vec{d}_j \in C_r} \vec{d}_j - \frac{1}{N - |C_r|} \sum_{\forall \vec{d}_j \notin C_r} \vec{d}_j$$

Where N is the total number of documents. The query vector **sum** the weights w_{ij} for all dj in C_r and **subtracts** all the weights w'_{ik} for all d_k not in C_r .

Example (query is «information retrieval»)

- Vocabulary:*(information, method, performance, retrieval, system)*
- **D1: (1,0,1,1,0)** “*information retrieval performances*”
- **D2: (1,0,1,1,1)** “*performance of information retrieval systems*”
- **D3: (0,1,0,0,1)** “*system's method*”
- C_r: **D1, D2**; N-C_r =**D3**

$$q_{opt} = \frac{1}{2}\{(1,0,1,1,0)+(1,0,1,1,1)\} - \frac{1}{3-2}(0,1,0,0,1) =$$

$$\frac{1}{2}(2,0,2,2,1)-(0,1,0,0,1) = (1,0,1,1,0.5)-(0,1,0,0,1) =$$

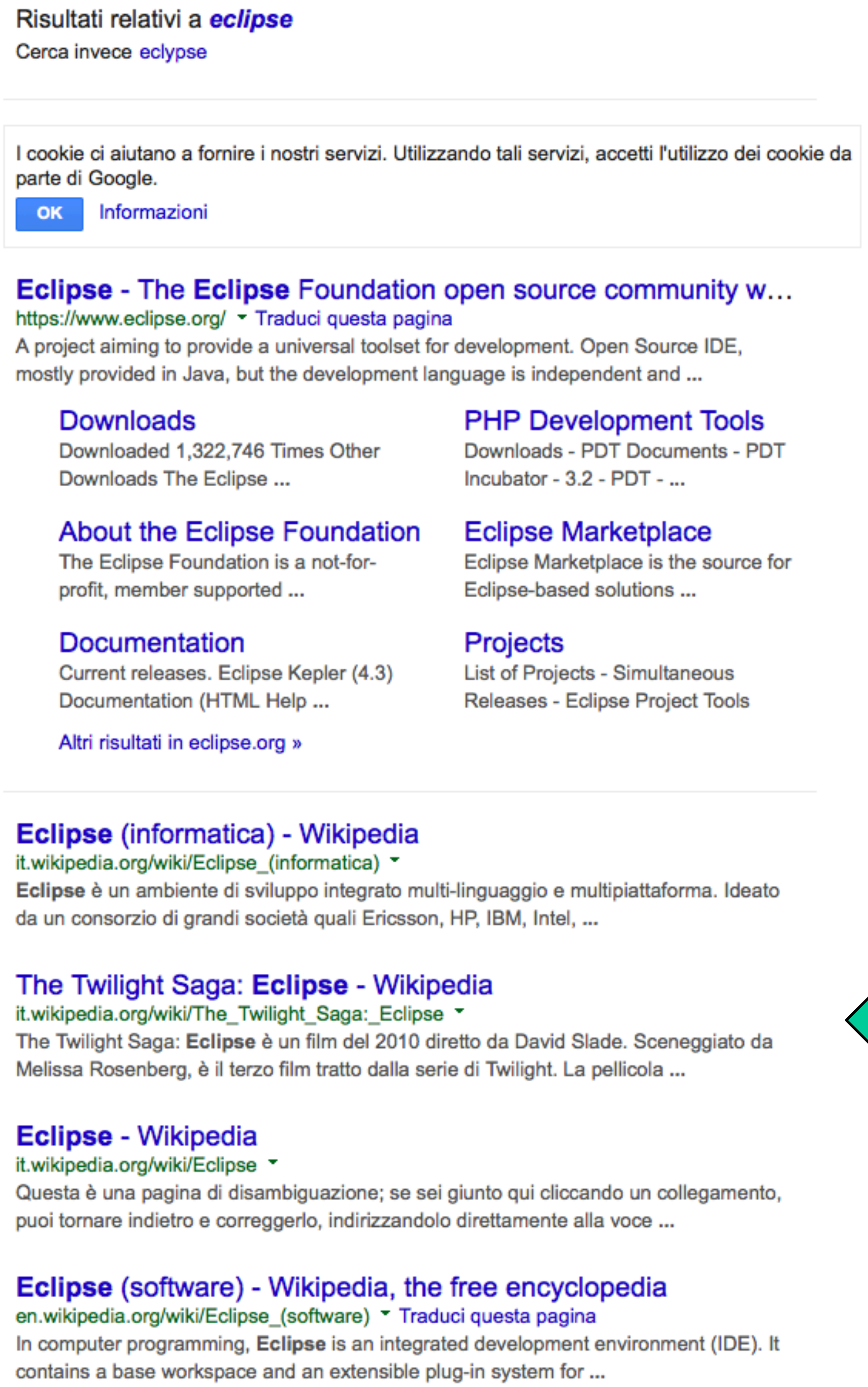
$$(1,-1,1,1,-0.5)$$

Standard Rocchio Method

- Previous method is not realistic since all relevant documents are unknown. Rocchio method uses the **known** relevant (D_r) and irrelevant (D_n) sets (among the first k ranked) of documents and include them in initial query *q*.

$$\vec{q}_m = \alpha\vec{q} + \frac{\beta}{|D_r|}\sum_{\forall \vec{d}_j \in D_r} \vec{d}_j - \frac{\gamma}{|D_n|}\sum_{\forall \vec{d}_j \in D_n} \vec{d}_j$$

α: Tunable weight for initial query.
β: Tunable weight for relevant documents.
γ: Tunable weight for irrelevant documents.



Risultati relativi a ***eclipse***

Cerca invece eclypse

I cookie ci aiutano a fornire i nostri servizi. Utilizzando tali servizi, accetti l'utilizzo dei cookie da parte di Google.

Eclipse - The **Eclipse** Foundation open source community w...
https://www.eclipse.org/ ▾ Traduci questa pagina
A project aiming to provide a universal toolset for development. Open Source IDE, mostly provided in Java, but the development language is independent and ...

Downloads
Downloaded 1,322,746 Times Other Downloads The Eclipse ...

PHP Development Tools
Downloads - PDT Documents - PDT Incubator - 3.2 - PDT - ...

About the Eclipse Foundation
The Eclipse Foundation is a not-for-profit, member supported ...

Eclipse Marketplace
Eclipse Marketplace is the source for Eclipse-based solutions ...

Documentation
Current releases. Eclipse Kepler (4.3) Documentation (HTML Help ...

Projects
List of Projects - Simultaneous Releases - Eclipse Project Tools

Altri risultati in eclipse.org »

Eclipse (informatica) - Wikipedia
it.wikipedia.org/wiki/Eclipse_(informatica) ▾
Eclipse è un ambiente di sviluppo integrato multi-linguaggio e multipiattaforma. Ideato da un consorzio di grandi società quali Ericsson, HP, IBM, Intel, ...

The Twilight Saga: **Eclipse** - Wikipedia
it.wikipedia.org/wiki/The_Twilight_Saga:_Eclipse ▾
The Twilight Saga: **Eclipse** è un film del 2010 diretto da David Slade. Sceneggiato da Melissa Rosenberg, è il terzo film tratto dalla serie di Twilight. La pellicola ...

Eclipse - Wikipedia
it.wikipedia.org/wiki/Eclipse ▾
Questa è una pagina di disambiguazione; se sei giunto qui cliccando un collegamento, puoi tornare indietro e correggerlo, indirizzandolo direttamente alla voce ...

Eclipse (software) - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Eclipse_(software) ▾ Traduci questa pagina
In computer programming, **Eclipse** is an integrated development environment (IDE). It contains a base workspace and an extensible plug-in system for ...

Query is: *eclipse saga*

D_r: *saga, movie, director, david slade, licantropus, melissa rosenberg..*

D_n: *foundation,software,development, tool,environment....*

Ide Regular Method

- Since more feedback should perhaps increase the degree of reformulation, do not normalize :

$$\vec{q}_m = \alpha\vec{q} + \beta \sum_{\forall \vec{d}_j \in D_r} \vec{d}_j - \gamma \sum_{\forall \vec{d}_j \in D_n} \vec{d}_j$$

α: Tunable weight for initial query.
β: Tunable weight for relevant documents.
γ: Tunable weight for irrelevant documents.

Example

- q $(w_{1q},w_{2q},w_{3q},w_{4q})$
- Dr: $[d_1(w_{11},w_{21},w_{31},w_{41}),$ $d_2(w_{12},w_{22},w_{32},w_{42})]$
- Dn: $d_3(w_{13},w_{23},w_{33},w_{43})$

- q_{exp}:$((\alpha w_{1q}+\beta(w_{11}+w_{12})-\gamma w_{13}),$ $(\alpha w_{2q}+\beta(w_{21}+w_{22})-\gamma w_{23}),$ $(\alpha w_{3q}+\beta(w_{31}+w_{32})-\gamma w_{33}),$ $(\alpha w_{4q}+\beta(w_{41}+w_{42})-\gamma w_{43}))$

Ide “Dec Hi” Method

- Bias towards rejecting **just** the highest ranked of the irrelevant documents:

$$\vec{q}_m = \alpha\vec{q} + \beta \sum_{\forall \vec{d}_j \in D_r} \vec{d}_j - \gamma \max_{non-relevant}(\vec{d}_j)$$

α: Tunable weight for initial query.
β: Tunable weight for relevant documents.
γ: Tunable weight for irrelevant document.

Comparison of Methods

- Overall, experimental results indicate no clear preference for any one of the specific methods.
- All methods generally improve retrieval performance (recall & precision) with feedback.
- Generally tunable constants α, β, γ equal 1.

Example

- Initial query: “*New space satellite applications*”

+ 1. 0.539, 08/13/91, NASA Hasn’ t Scrapped Imaging Spectrometer

+ 2. 0.533, 07/09/91, NASA Scratches Environment Gear From Satellite Plan

3. 0.528, 04/04/90, Science Panel Backs NASA Satellite Plan, But Urges Launches of Smaller Probes

4. 0.526, 09/09/91, A NASA Satellite Project Accomplishes Incredible Feat: Staying Within Budget

5. 0.525, 07/24/90, Scientist Who Exposed Global Warming Proposes Satellites for Climate Research

6. 0.524, 08/22/90, Report Provides Support for the Critics Of Using Big Satellites to Study Climate

7. 0.516, 04/13/87, Arianespace Receives Satellite Launch Pact From Telesat Canada

+ 8. 0.509, 12/02/87, Telecommunications Tale of Two Companies

- User marks relevant documents with “+”.

Expanded query after relevance feedback

2.074	new	15.106	space
30.816	satellite	5.660	application
5.991	nasa	5.196	eos
4.196	launch	3.972	aster
3.516	instrument	3.446	arianespace
3.004	bundespost	2.806	ss
2.790	rocket	2.053	scientist
2.003	broadcast	1.172	earth
0.836	oil	0.646	measure

Results for expanded query

2 1. 0.513, 07/09/91, NASA Scratches Environment Gear From Satellite Plan

1 2. 0.500, 08/13/91, NASA Hasn’t Scrapped Imaging Spectrometer

3. 0.493, 08/07/89, When the Pentagon Launches a Secret Satellite, Space Sleuths Do Some Spy Work of Their Own

4. 0.493, 07/31/89, NASA Uses ‘Warm’ Superconductors For Fast Circuit

8 5. 0.492, 12/02/87, Telecommunications Tale of Two Companies

6. 0.491, 07/09/91, Soviets May Adapt Parts of SS-20 Missile For Commercial Use

7. 0.490, 07/12/88, Gaping Gap: Pentagon Lags in Race To Match the Soviets In Rocket Launchers

8. 0.490, 06/14/90, Rescue of Satellite By Space Agency To Cost $90 Million

Relevance Feedback on the Web

- Some search engines offer a *similar/related* pages feature (this is a trivial form of relevance feedback)
  - Google (link-based, but is now hidden). It rather shows "related search"
  - But some don't because it's hard to explain to average user why a page is suggested
- Specialized search engines are those who more often use feedback

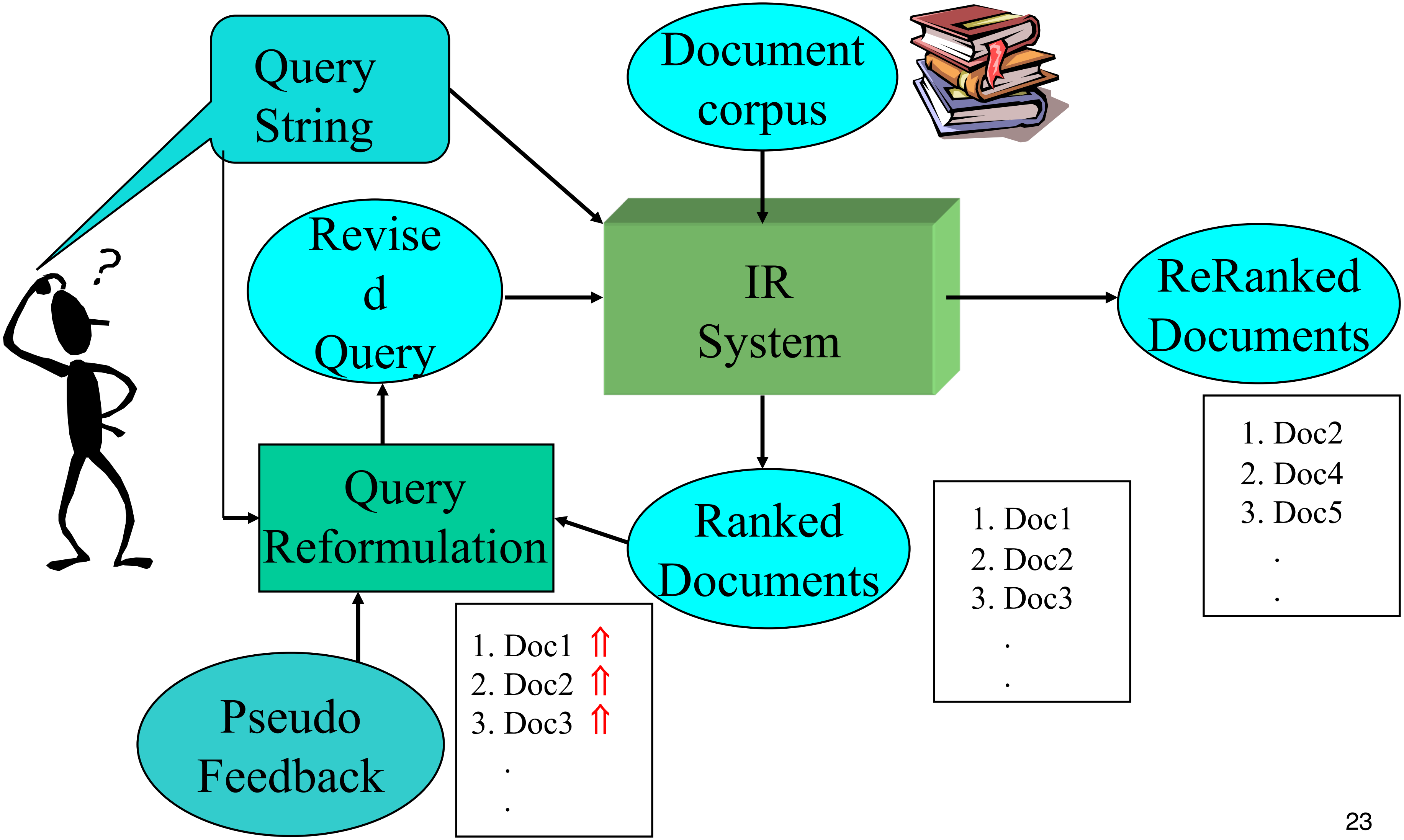
Why is Feedback Not Widely Used

- Users sometimes reluctant to provide explicit feedback.
- Results in long queries that require more computation to retrieve documents: search engines process lots of queries and allow little time for each one.
- Makes it harder to understand why a particular document was retrieved.

Pseudo Feedback

- Use relevance feedback methods **without explicit user input**.
- Just **assume** the top *m* retrieved documents are relevant, and use them to reformulate the query.
- Allows for query expansion that includes terms that are correlated with the query terms.
- Would not work well for previous "Eclypse" example but common queries are less ambiguous,
- E.g. Eclypse licantropous, Eclypse moon

Pseudo Feedback Architecture



Query String

Document corpus

Revise
d
Query

IR
System

ReRanked Documents

Query Reformulation

Ranked Documents

Pseudo Feedback

1. Doc1 ⇑
2. Doc2 ⇑
3. Doc3 ⇑
.
.

1. Doc1
2. Doc2
3. Doc3
.
.

1. Doc2
2. Doc4
3. Doc5
.
.

PseudoFeedback Results

- Found to improve performance on public IR competitions (ad-hoc retrieval task).
- Works even better if top documents must also satisfy additional boolean constraints in order to be used in feedback (especially negative constraints like

eclipse AND (licantropus OR not moon).

OTHER METHODS FOR QUERY EXPANSION

Query expansion methods

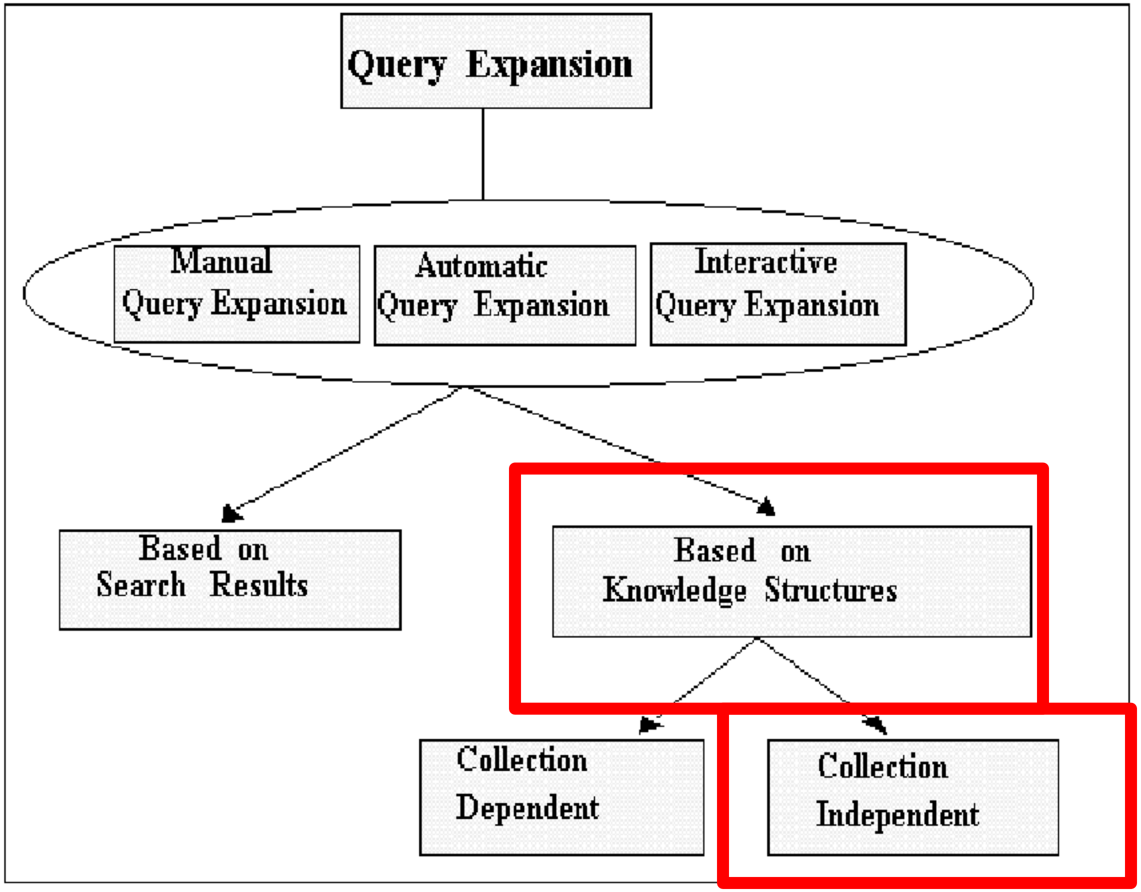


Figure 1: Query Expansion: Methods and Sources

Thesaurus

- A thesaurus provides information on synonyms and semantically related words and phrases.
- Example:

```
physician
   syn: ||croaker, doc, doctor, MD,
medical, mediciner, medico, ||sawbones
   rel: medic, general practitioner,
surgeon,
```

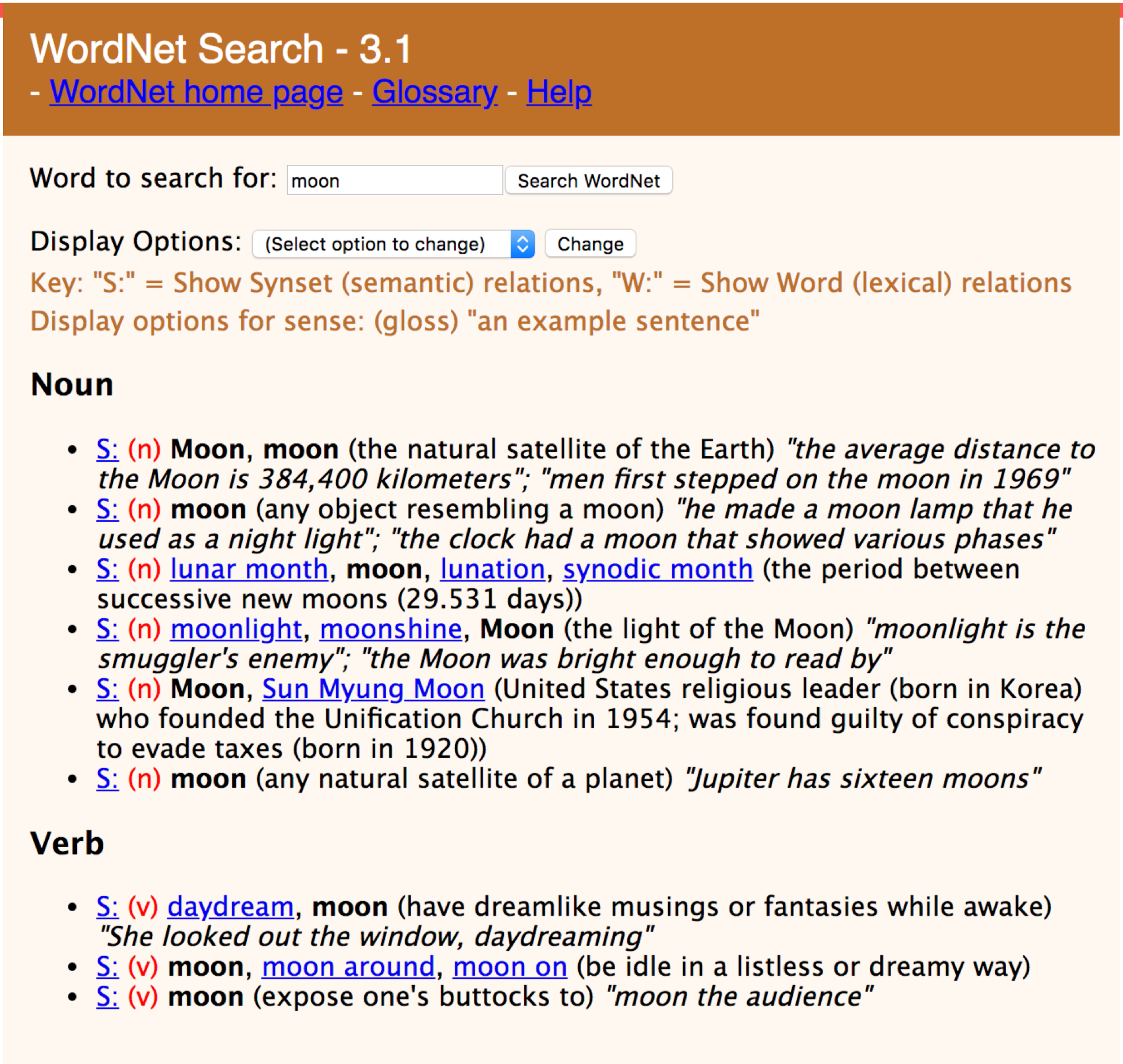
Thesaurus-based Query Expansion

- For each term, *t*, in a query, expand the query with synonyms and related words of *t* from the thesaurus.
- Can weight added terms less than original query terms (= *discount factor* for terms not in original query).
- Generally increases recall.
- May significantly decrease precision, particularly with ambiguous terms.
  - "interest rate" → "interest rate fascinate evaluate"

WordNet

- A more detailed database of semantic relationships between English words.
- Developed by famous cognitive psychologist George Miller and a team at Princeton University.
- About 144,000 English words.
- Nouns, adjectives, verbs, and adverbs grouped into about 109,000 synonym sets called *synsets*.

Wordnet



WordNet Search - 3.1

- WordNet home page - Glossary - Help

Word to search for: moon

Display Options: (Select option to change)

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Display options for sense: (gloss) "an example sentence"

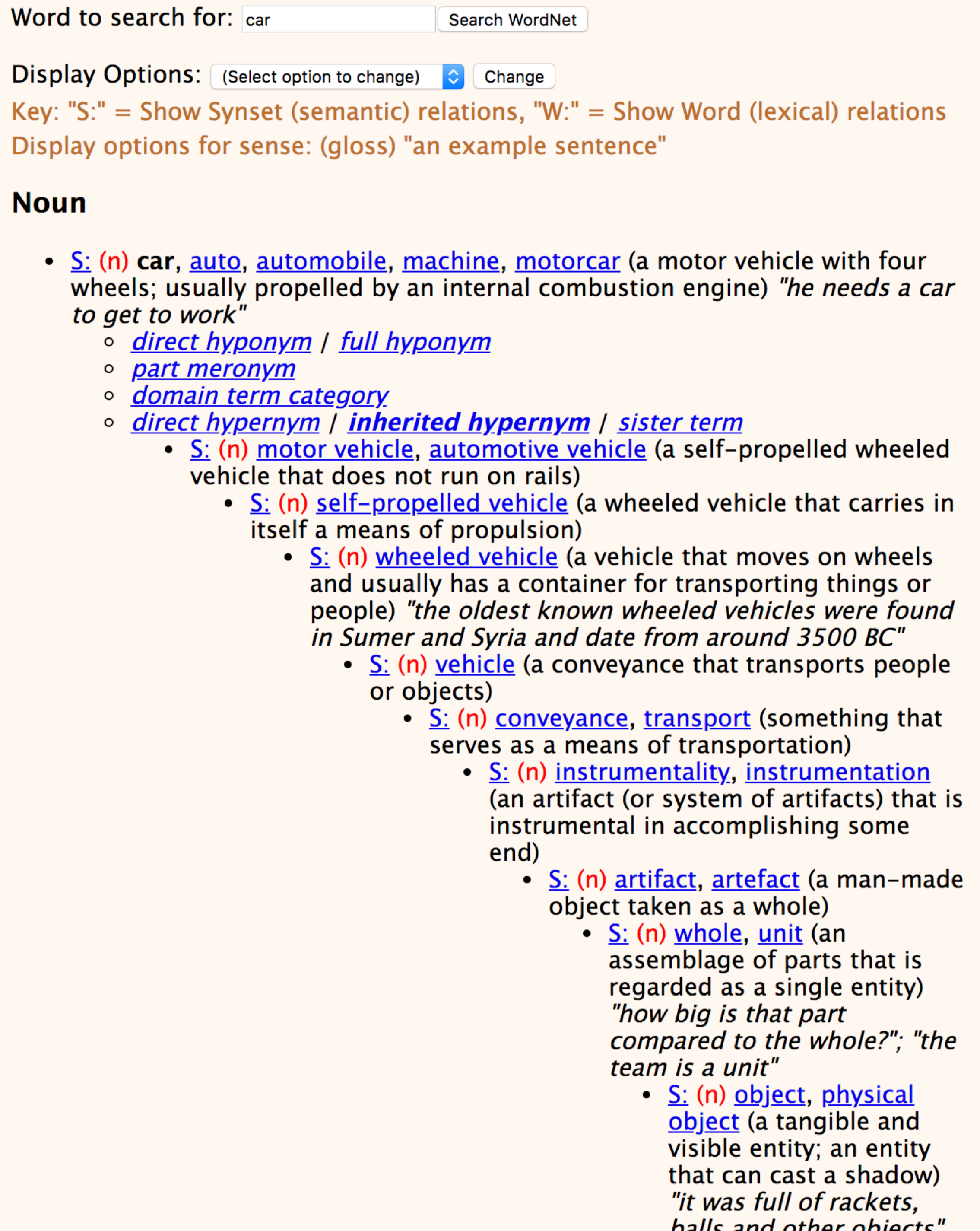
Noun

- S: (n) **Moon**, **moon** (the natural satellite of the Earth) *"the average distance to the Moon is 384,400 kilometers"; "men first stepped on the moon in 1969"*
- S: (n) **moon** (any object resembling a moon) *"he made a moon lamp that he used as a night light"; "the clock had a moon that showed various phases"*
- S: (n) lunar month, **moon**, lunation, synodic month (the period between successive new moons (29.531 days))
- S: (n) moonlight, moonshine, **Moon** (the light of the Moon) *"moonlight is the smuggler's enemy"; "the Moon was bright enough to read by"*
- S: (n) **Moon**, Sun Myung Moon (United States religious leader (born in Korea) who founded the Unification Church in 1954; was found guilty of conspiracy to evade taxes (born in 1920))
- S: (n) **moon** (any natural satellite of a planet) *"Jupiter has sixteen moons"*

Verb

- S: (v) daydream, **moon** (have dreamlike musings or fantasies while awake) *"She looked out the window, daydreaming"*
- S: (v) **moon**, moon around, moon on (be idle in a listless or dreamy way)
- S: (v) **moon** (expose one's buttocks to) *"moon the audience"*

Wordnet hierarchy



Word to search for: car

Display Options: (Select option to change)

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Display options for sense: (gloss) "an example sentence"

Noun

- S: (n) **car**, auto, automobile, machine, motorcar (a motor vehicle with four wheels; usually propelled by an internal combustion engine) *"he needs a car to get to work"*
    - *direct hyponym* / *full hyponym*
    - *part meronym*
    - *domain term category*
    - *direct hypernym* / ***inherited hypernym*** / *sister term*
        - S: (n) motor vehicle, automotive vehicle (a self-propelled wheeled vehicle that does not run on rails)
            - S: (n) self-propelled vehicle (a wheeled vehicle that carries in itself a means of propulsion)
                - S: (n) wheeled vehicle (a vehicle that moves on wheels and usually has a container for transporting things or people) *"the oldest known wheeled vehicles were found in Sumer and Syria and date from around 3500 BC"*
                    - S: (n) vehicle (a conveyance that transports people or objects)
                        - S: (n) conveyance, transport (something that serves as a means of transportation)
                            - S: (n) instrumentality, instrumentation (an artifact (or system of artifacts) that is instrumental in accomplishing some end)
                                - S: (n) artifact, artefact (a man-made object taken as a whole)
                                    - S: (n) whole, unit (an assemblage of parts that is regarded as a single entity) *"how big is that part compared to the whole?"; "the team is a unit"*
                                        - S: (n) object, physical object (a tangible and visible entity; an entity that can cast a shadow) *"it was full of rackets, balls and other objects"*

WordNet Synset Relationships

- Antonym: front → back
- Attribute: benevolence → good (noun to adjective)
- Pertainym: alphabetical → alphabet (adjective to noun)
- Similar: unquestioning → absolute
- Cause: kill → die
- Entailment: breathe → inhale
- Holonym: chapter → text (part-of)
- Meronym: computer → cpu (whole-of)
- Hyponym: plant → tree (specialization)
- Hypernym: apple → fruit (generalization)

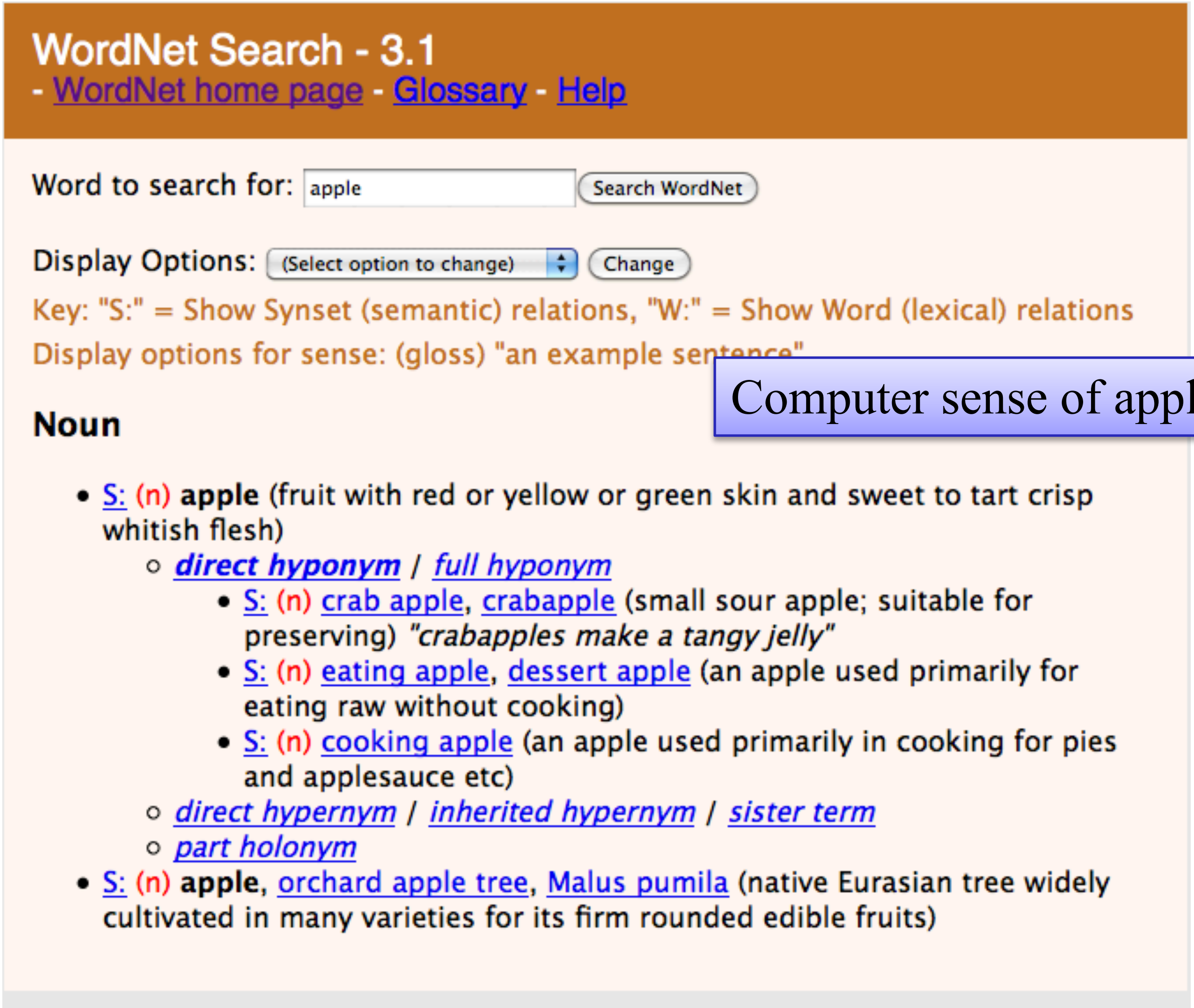
WordNet Query Expansion

- Add synonyms in the same synset.
- Add hyponyms to add specialized terms.
- Add hypernyms to generalize a query.
- Add other related terms to expand query.
- In case of ambiguity, which synset?

Example query: car rental

Expanded query: (car OR automonile OR machine OR ..)AND
(rental OR leasing OR..)

Not all senses available



WordNet Search - 3.1
- WordNet home page - Glossary - Help
Search WordNet
(Select option to change)
Change

Word to search for: apple

Display Options: (Select option to change)

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

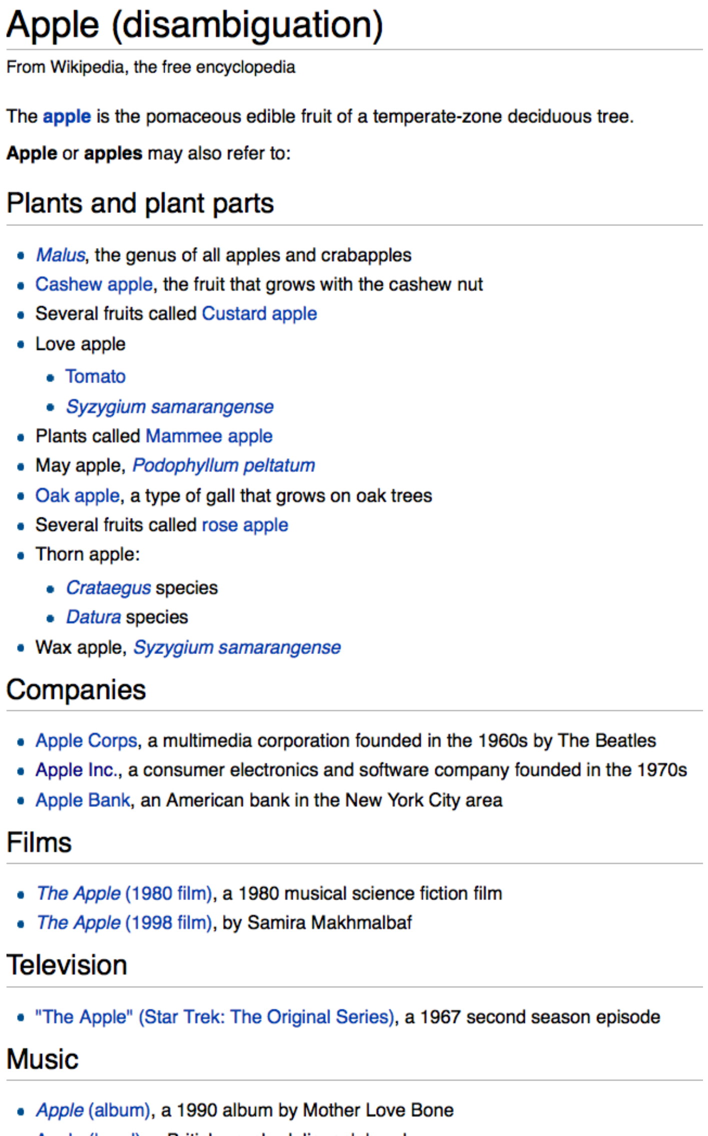
Display options for sense: (gloss) "an example sentence"

Computer sense of apple is missing

Noun

- S: (n) **apple** (fruit with red or yellow or green skin and sweet to tart crisp whitish flesh)
    - ***direct hyponym*** / *full hyponym*
        - S: (n) crab apple, crabapple (small sour apple; suitable for preserving) *"crabapples make a tangy jelly"*
        - S: (n) eating apple, dessert apple (an apple used primarily for eating raw without cooking)
        - S: (n) cooking apple (an apple used primarily in cooking for pies and applesauce etc)
    - *direct hypernym* / *inherited hypernym* / *sister term*
    - *part holonym*
- S: (n) **apple**, orchard apple tree, Malus pumila (native Eurasian tree widely cultivated in many varieties for its firm rounded edible fruits)

A better source: Wikipedia (disambiguation page)



Apple (disambiguation)

From Wikipedia, the free encyclopedia

The **apple** is the pomaceous edible fruit of a temperate-zone deciduous tree.

Apple or **apples** may also refer to:

Plants and plant parts

- *Malus*, the genus of all apples and crabapples
- Cashew apple, the fruit that grows with the cashew nut
- Several fruits called Custard apple
- Love apple
  - Tomato
  - *Syzygium samarangense*
- Plants called Mammee apple
- May apple, *Podophyllum peltatum*
- Oak apple, a type of gall that grows on oak trees
- Several fruits called rose apple
- Thorn apple:
  - *Crataegus* species
  - *Datura* species
- Wax apple, *Syzygium samarangense*

Companies

- Apple Corps, a multimedia corporation founded in the 1960s by The Beatles
- Apple Inc., a consumer electronics and software company founded in the 1970s
- Apple Bank, an American bank in the New York City area

Films

- *The Apple* (1980 film), a 1980 musical science fiction film
- *The Apple* (1998 film), by Samira Makhmalbaf

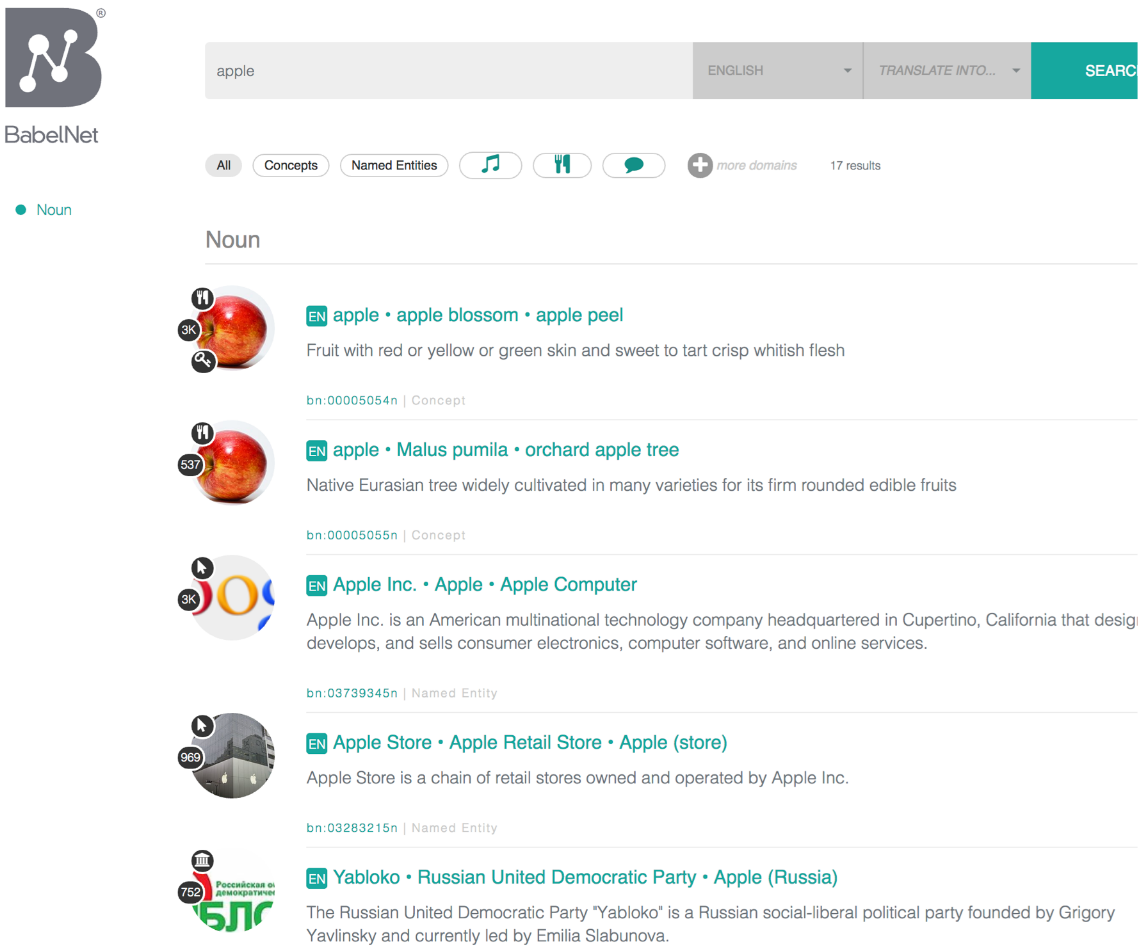
Television

- "The Apple" (Star Trek: The Original Series), a 1967 second season episode

Music

- *Apple* (album), a 1990 album by Mother Love Bone

An even better source: BabelNet



Noun

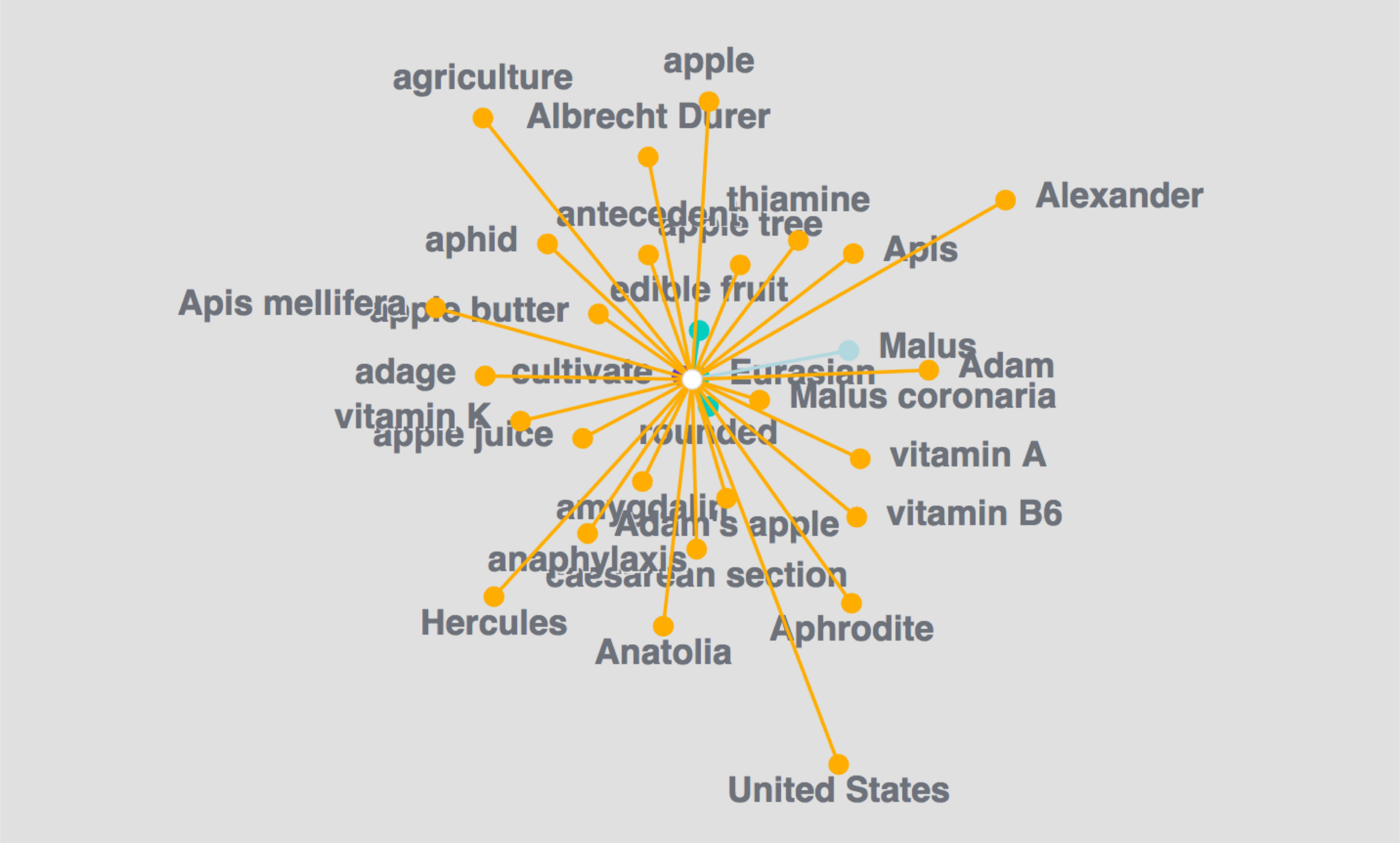
EN apple • apple blossom • apple peel
Fruit with red or yellow or green skin and sweet to tart crisp whitish flesh
bn:00005054n | Concept

EN apple • Malus pumila • orchard apple tree
Native Eurasian tree widely cultivated in many varieties for its firm rounded edible fruits
bn:00005055n | Concept

EN Apple Inc. • Apple • Apple Computer
Apple Inc. is an American multinational technology company headquartered in Cupertino, California that desig
develops, and sells consumer electronics, computer software, and online services.
bn:03739345n | Named Entity

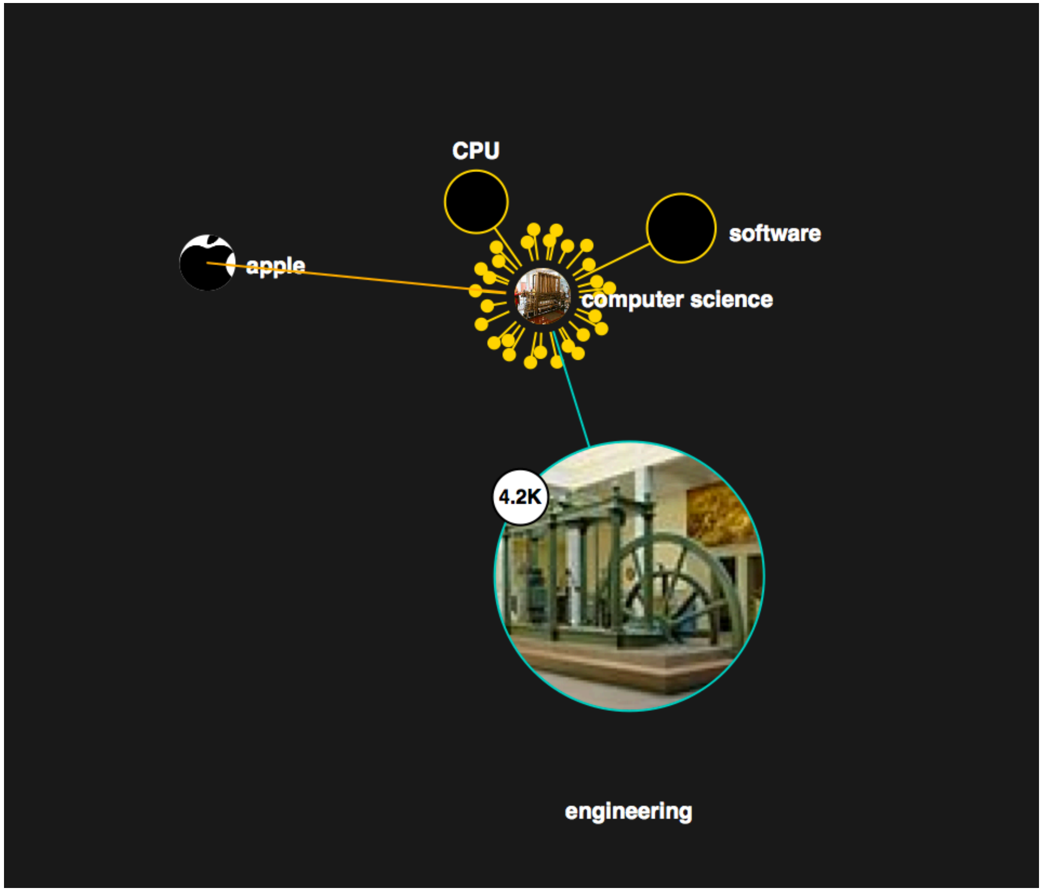
EN Apple Store • Apple Retail Store • Apple (store)
Apple Store is a chain of retail stores owned and operated by Apple Inc.
bn:03283215n | Named Entity

EN Yabloko • Russian United Democratic Party • Apple (Russia)
The Russian United Democratic Party "Yabloko" is a Russian social-liberal political party founded by Grigory Yavlinsky and currently led by Emilia Slabunova.


apple
agriculture
Albrecht Durer
thiamine
antecedent
apple tree
Alexander
aphid
Apis
Apis mellifera
apple butter
edible fruit
Malus
adage
cultivate
Eurasian
Adam
Malus coronaria
vitamin K
apple juice
rounded
vitamin A
amygdalin
Adam's apple
vitamin B6
anaphylaxis
caesarean section
Hercules
Aphrodite
Anatolia
United States

Use context in query to disambiguate (e.g. “apple computer”) or click on right sense

Apple Inc. is an American multinational corporation that designs and sells consumer electronics, **computer** software, and personal computers.



Explore: apple computer

apple, CPU,engineering , computer science, software,

Query expansion methods

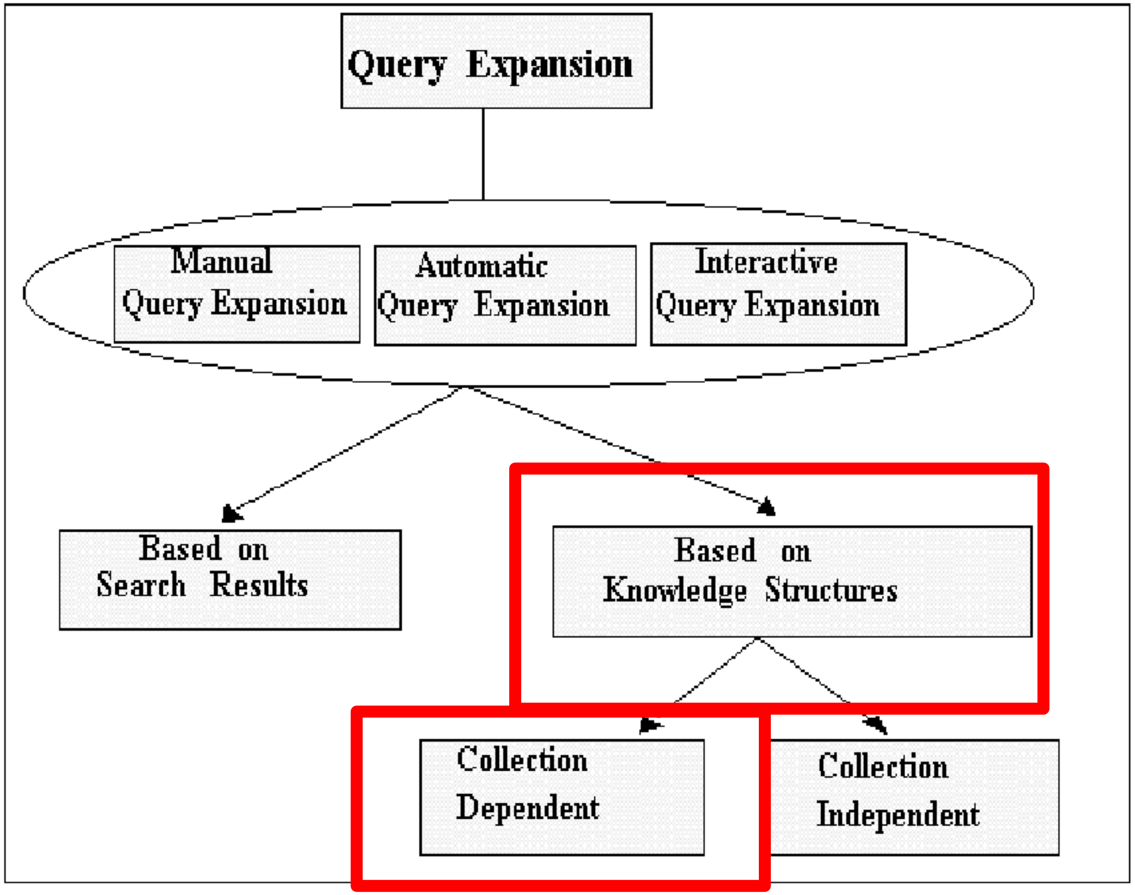


Figure 1: Query Expansion: Methods and Sources

STATISTICAL QUERY EXPANSION

Statistical Expansion

- Existing human-developed thesauri are not easily available in all languages (even though now BabelNet has 100 languages).
- More importantly: *Semantically related terms can be more easily discovered from statistical analysis of corpora.*
- E.g. “licantrope” and “eclipse” may not co-occur in thesauri: “*The Twilight Saga: Eclipse, commonly referred to as Eclipse, is a 2010 American romantic fantasy film based on Stephenie Meyer's 2007 novel, Eclipse*” but they do co-occur in texts (more free texts available than thesauri..)

Automatic Global Analysis

- Determine term similarity through a **pre-computed statistical analysis** of the complete corpus.
- Compute association matrices which quantify term correlations in terms of how frequently they co-occur.
- Expand queries with statistically most similar terms.

Association Matrix

$$\begin{array}{c|cccccc|}
 & w_1 & w_2 & w_3 & \ldots\ldots\ldots\ldots\ldots\ldots & w_n \\ \hline
w_1 & c_{11} & c_{12} & c_{13} & \ldots\ldots\ldots\ldots\ldots\ldots & c_{1n} \\
w_2 & c_{21} & & & & \\
w_3 & c_{31} & & & & \\
. & . & & & & \\
. & . & & & & \\
w_n & c_{n1} & & & & \\ \hline
\end{array}$$

c_{ij}: Correlation factor between term *i* and term *j*

$$c_{ij} = \sum_{d_k \in D} f_{ik} \times f_{jk}$$

c_{ij}=0 if either i or j do not occur in d_k
c_{ii}= sum of quadratic frequencies

f_{ik} : Frequency of term *i* in document *k*

Normalized Association Matrix

- Frequency based correlation factor favors more frequent terms: need to discriminate chance (they co-occur because they are very frequent) from genuine relatdness
- **Normalize** association scores:

$$s_{ij} = \frac{c_{ij}}{c_{ii} + c_{jj} - c_{ij}}$$

Numerator: SUM(product of *i-j* frequencies)
Denominator: SUM(frequency of *i*)^2 + SUM(frequency of *j*)^2 – numerator

- Normalized score is 1 if two terms have the same frequency in all documents.

Example (assuming freq=1 or 0 in all docs)

- Documents with “information” : 5500
- Documents with “retrieval” : 2600
- Documents with both: 2500

$$\frac{2500}{5500+2600-2500}=\frac{2500}{5600}=0,45$$

Metric Correlation Matrix

- Association correlation does not account for the proximity of terms in documents, just co-occurrence frequencies within documents.
- Metric correlations account for term **proximity**.

$$c_{ij} = \sum_{k_u \in V_i} \sum_{k_v \in V_j} \frac{1}{r(k_u, k_v)}$$

V_i: Set of all occurrences of term i in any document.

$r(k_u, k_v)$: Distance in words between word occurrences k_u and k_v (∞ if k_u and k_v are occurrences in different documents).

Normalized Metric Correlation Matrix

- Normalize scores to account for term frequencies:

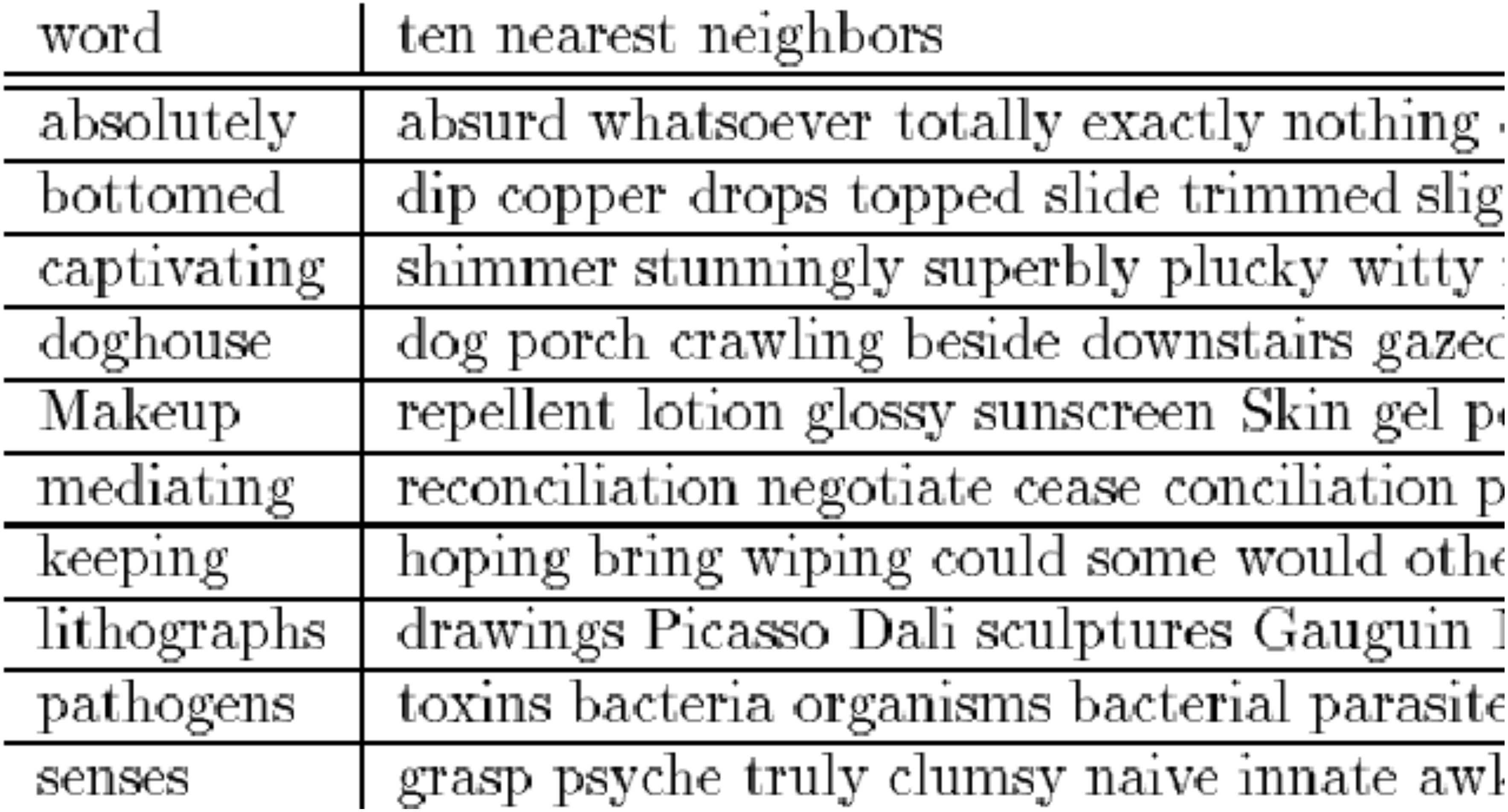
$$s_{ij} = \frac{c_{ij}}{|V_i| \times |V_j|} = \sum_{k_u \in V_i} \sum_{k_v \in V_j} \frac{1}{r(k_u, k_v)} / (|V_i| \times |V_j|)$$

V_i, V_j are the subset of documents in the collection including term i or term j

Query Expansion with Correlation Matrix

- For each term *i* in query, expand query with the *n* terms, *j*, with the highest value of c_{ij} (s_{ij}).
- **This adds related terms found in the "neighborhood" of the query terms.**

Co-occurrence table Example



word	ten nearest neighbors
absolutely	absurd whatsoever totally exactly nothing
bottomed	dip copper drops topped slide trimmed slig
captivating	shimmer stunningly superbly plucky witty
doghouse	dog porch crawling beside downstairs gazec
Makeup	repellent lotion glossy sunscreen Skin gel p
mediating	reconciliation negotiate cease conciliation p
keeping	hoping bring wiping could some would othe
lithographs	drawings Picasso Dali sculptures Gauguin
pathogens	toxins bacteria organisms bacterial parasite
senses	grasp psyche truly clumsy naive innate awk

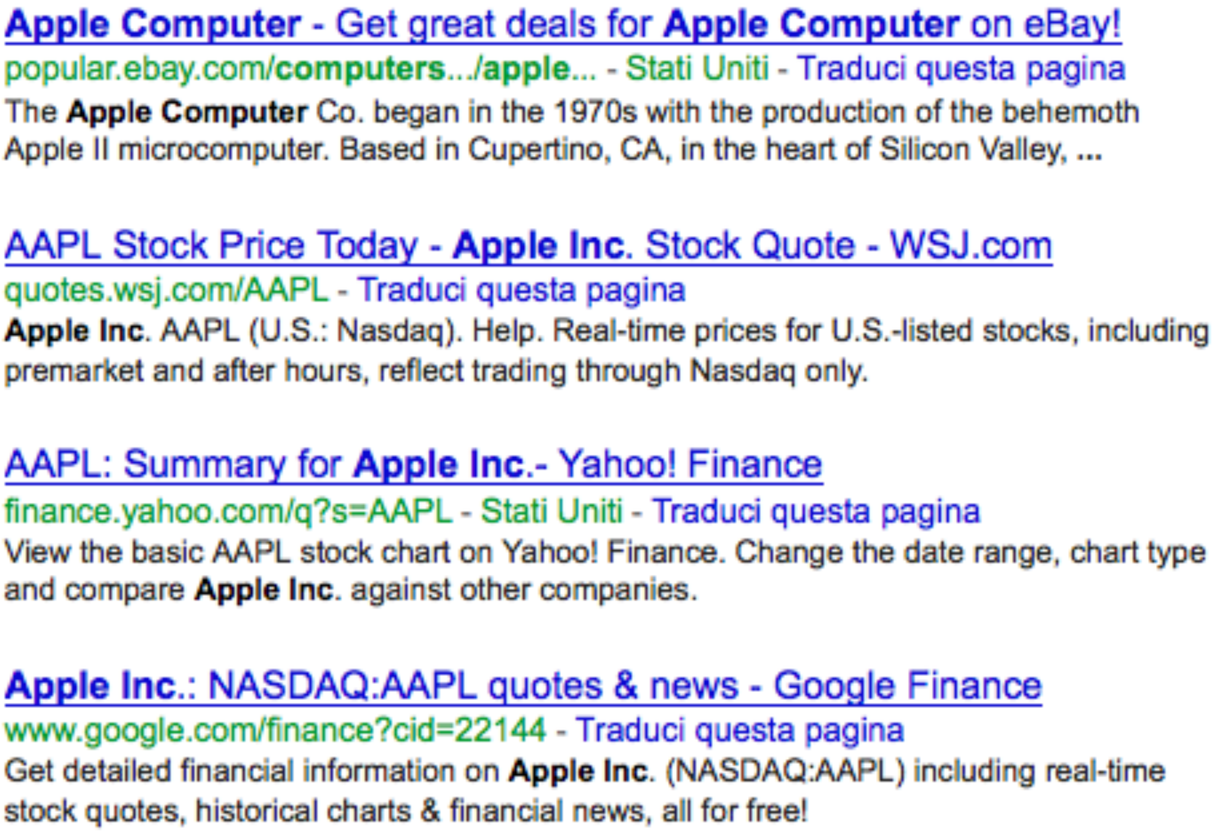
Problems with Global Analysis

- Term ambiguity may introduce irrelevant statistically correlated terms.
  - "Apple computer" → "Apple red fruit computer"
- Since terms are highly correlated anyway, expansion may not retrieve many additional documents.

Automatic Local Analysis

- At query time, dynamically determine similar terms based on analysis of top-ranked **retrieved documents**.
- Base correlation analysis on only the "local" set of retrieved documents for a specific query.
- Avoids ambiguity by determining similar (correlated) terms only within relevant documents.
  - "Apple computer" →
    "Apple computer Powerbook laptop"

Example (apple computer)



Apple Computer - Get great deals for **Apple Computer** on eBay!
popular.ebay.com/**computers**.../**apple**... - Stati Uniti - Traduci questa pagina
The **Apple Computer** Co. began in the 1970s with the production of the behemoth Apple II microcomputer. Based in Cupertino, CA, in the heart of Silicon Valley, ...

AAPL Stock Price Today - **Apple Inc**. Stock Quote - WSJ.com
quotes.wsj.com/AAPL - Traduci questa pagina
Apple Inc. AAPL (U.S.: Nasdaq). Help. Real-time prices for U.S.-listed stocks, including premarket and after hours, reflect trading through Nasdaq only.

AAPL: Summary for **Apple Inc**.- Yahoo! Finance
finance.yahoo.com/q?s=AAPL - Stati Uniti - Traduci questa pagina
View the basic AAPL stock chart on Yahoo! Finance. Change the date range, chart type and compare **Apple Inc**. against other companies.

Apple Inc.: NASDAQ:AAPL quotes & news - Google Finance
www.google.com/finance?cid=22144 - Traduci questa pagina
Get detailed financial information on **Apple Inc**. (NASDAQ:AAPL) including real-time stock quotes, historical charts & financial news, all for free!

microcomputer, company, stock quotes,..

Global vs. Local Analysis

- Global analysis requires intensive term correlation computation only occasionally.
- Local analysis requires intensive term correlation computation for every query at run time (although number of terms and documents is less than in global analysis).
- But local analysis gives better results.

Global Analysis Refinements

- Only expand query with terms that are similar to ***all* terms in the query**.

$$sim(k_i, Q) = \sum_{k_j \in Q} c_{ij}$$

  - “fruit” not added to “Apple computer” since it is far from “computer.”
  - “fruit” added to “apple pie” since “fruit” close to both “apple” and “pie.”
- Use more sophisticated term weights (instead of just frequency) when computing term correlations.

Query Expansion with co-occurrences: Conclusions

- Expansion of queries with related terms can improve performance, particularly recall (more terms=more documents with same rank threshold).
- However, must select similar terms very carefully to avoid problems, such as loss of precision (e.g. if unrelated terms are added, precision might considerably decrease).

Query expansion methods

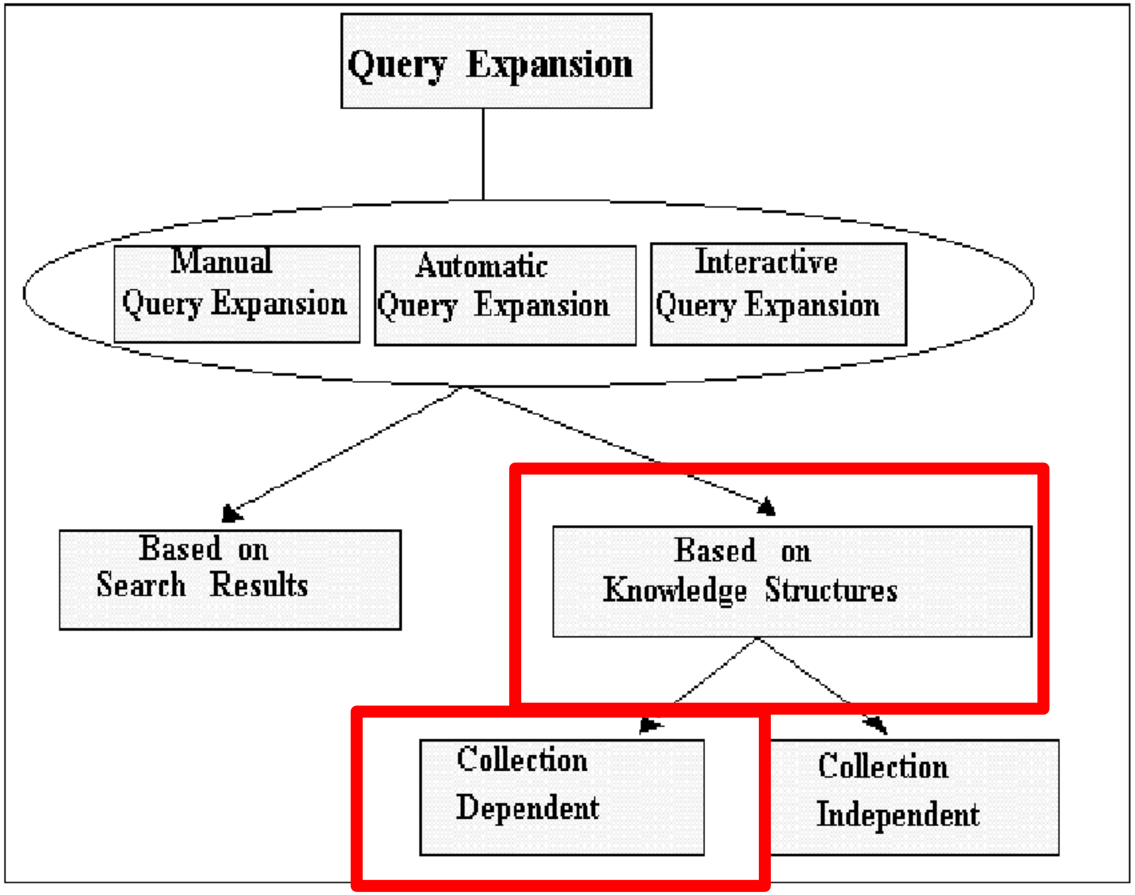
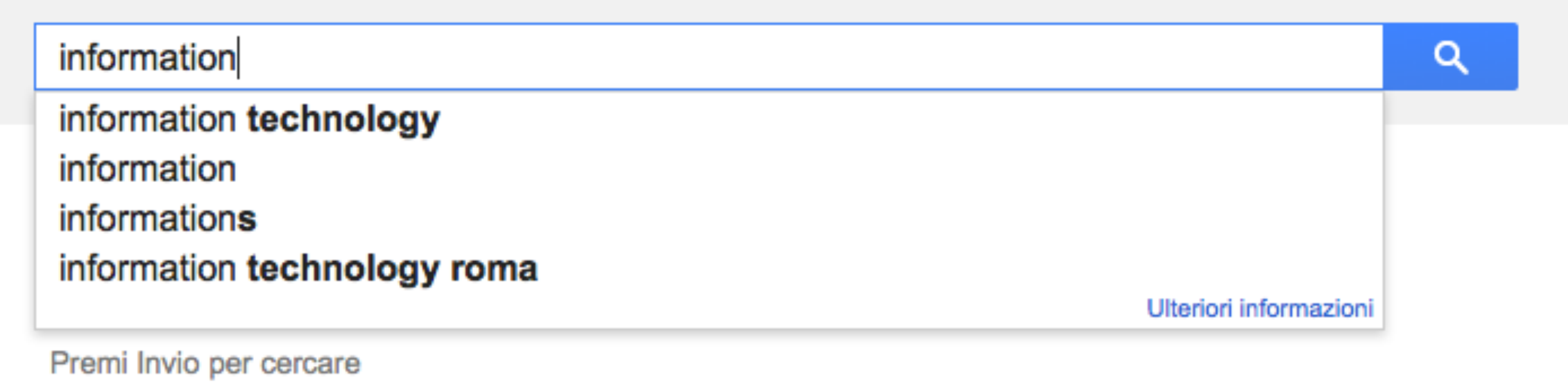


Figure 1: Query Expansion: Methods and Sources

Expansion qith query logs

- Google use query logs:



information

information **technology**

information

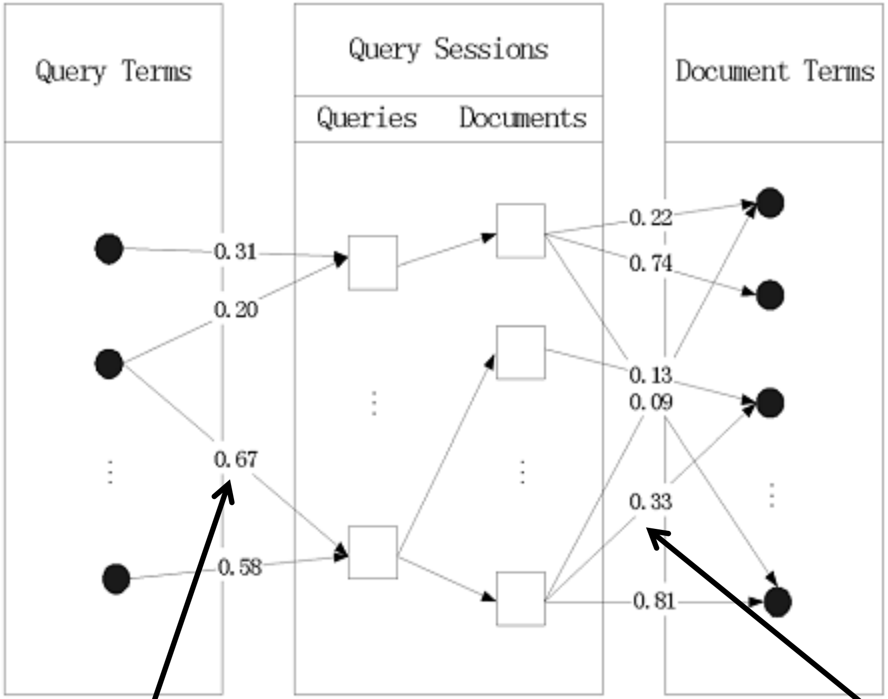
information**s**

information **technology roma**

Query is expanded with “hints” as you type words into the query window

Query expansion with query logs

Query logs: given a query, which documents have been accessed?



Relevance of term i in doc j

Prob of term j in query k

Learning from querylogs is important, however

- Google process 3 billion queries per day
- Lots of data, however, 20-25% of these queries are NEW, have never been done before
- 450ml previously “unseen” queries
- Therefore “brute force” is not enough, need to LEARN from previous query and GENERALIZE
- GENERALIZE= learn word meaning and word correlations

Is there anything more advanced than co-occurrences to learn correlations?

- To detect these similarities (next lessons):
  - Latent Semantic Indexing
  - Word embeddings (a.k.o. deep method)