

# **Relevance Feedback & Query Expansion**

# Tema: come migliorare la recall di una query?

## 1. Relevance feedback

- Direct feedback
- Pseudo feedback

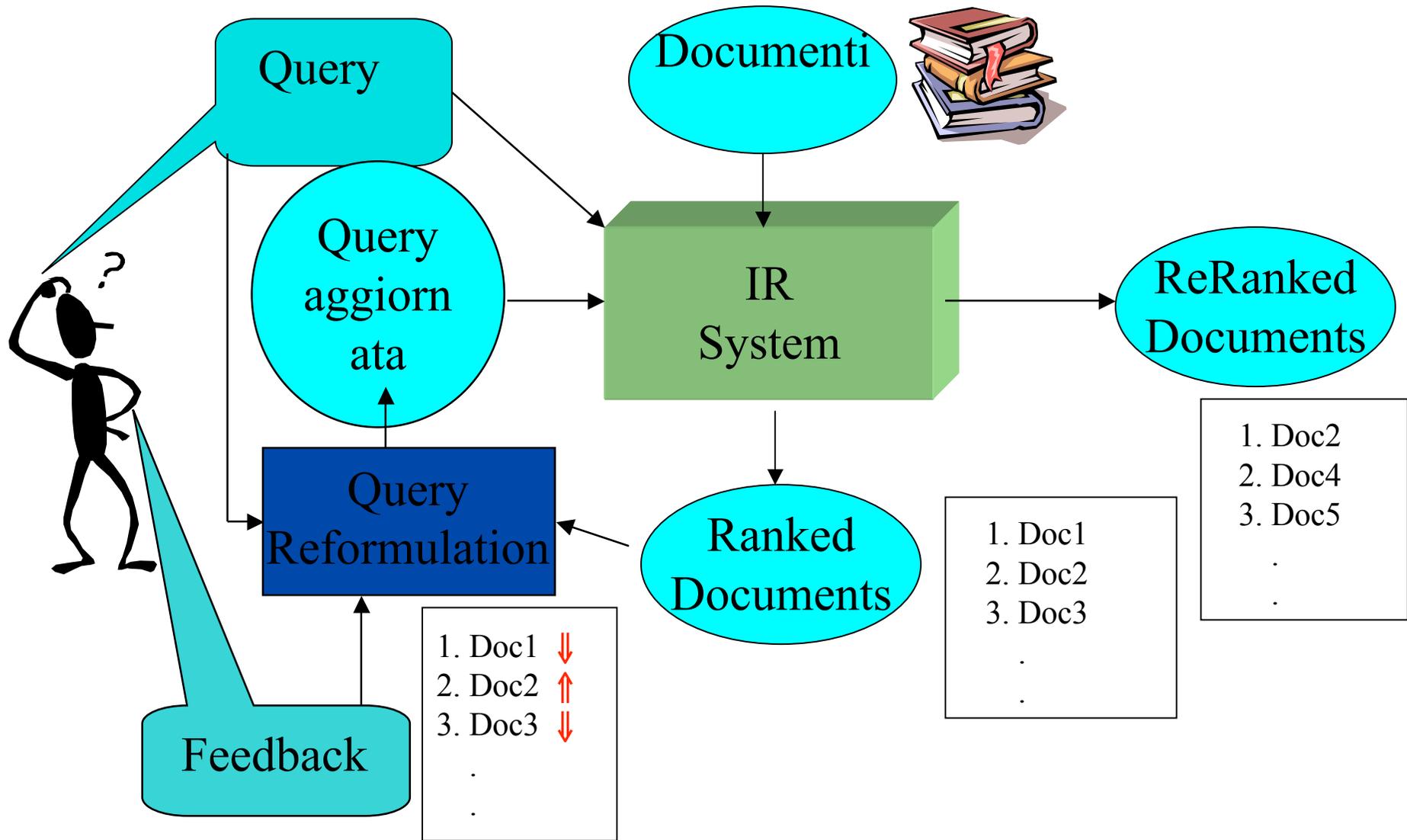
## 2. Query expansion

- Con co-occorrenze
- Usando un thesaurus

# Relevance Feedback

- Dopo la presentazione di un set iniziale di documenti, chiedi all'utente di selezionare i più rilevanti
- Usa questo **feedback** per riformulare la query
- Presenta nuovi risultati all'utente.
- Eventualmente, itera il processo.

# Relevance Feedback



# Query Reformulation

- Come tener conto del feedback?
  - **Query Expansion**: Aggiungi alla query **nuovi termini** estratti dai documenti prescelti dall'utente
  - **Term Reweighting**: Aumenta il peso dei termini che compaiono nei documenti rilevanti e diminuisci il peso di quelli che non vi compaiono.
- Diversi algoritmi per effettuare la riformulazione della query.

# Query Reformulation nel modello vettoriale

- Modifica il vettore della query:
  - **Aggiungi** i vettori dei documenti **rilevanti** al vettore della query.
  - **Sottrai** i vettori dei documenti **irrelevanti** al vettore della query.

# Optimal Query

- Sia  $C_r$  il set dei vettori dei documenti rilevanti (selezionati dall'utente).
- Allora la migliore query che classifichi **tutti e solo** i documenti giudicati rilevanti è:

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

Dove  $N$  è il numero totale di documenti sottoposti all'utente per giudizio.

# Esempio

- d1:(1,1,0,0,0)

- d2: (1,1,0,0,1)

- d3: (0,0,0,0,1)

- d4: (0,0,0,1,0)

- d1,d2 rilevanti, d3 d4 non rilevanti

- $q = 1/2 (2,2,0,0,1) - 1/2(0,0,0,1,1) = 1/2(2,2,0,-1,0)$

$= (1,1,0,-1/2,0)$

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

# Metodo di Rocchio

- Ovviamente non si conoscono tutti i documenti rilevanti, ma solo, fra quelli proposti all'utente, la frazione dei rilevanti ( $D_r$ ) e irrilevanti ( $D_n$ ) rispetto alla query iniziale  $q$ . **Formula di Rocchio:**

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

$\alpha$ : Un peso (regolabile) per la query iniziale.

$\beta$ : peso dei documenti rilevanti.

$\gamma$ : peso dei documenti irrilevanti.

I tre parametri sono regolabili

# Ide Regular Method

- In questa variante, si evita la normalizzazione, nell'ipotesi che ulteriore feedback migliori il grado di riformulazione :

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

# Ide “Dec Hi” Method

- Si utilizza solo, fra gli irrilevanti, quello con più alto rank :

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

# Paragone dei metodi

- Dati sperimentali non indicano sostanziali differenze.
- Tutti e 3 i metodi migliorano sia la recall che la precisione.
- In generale:

$$\alpha = \beta = \gamma = 1$$

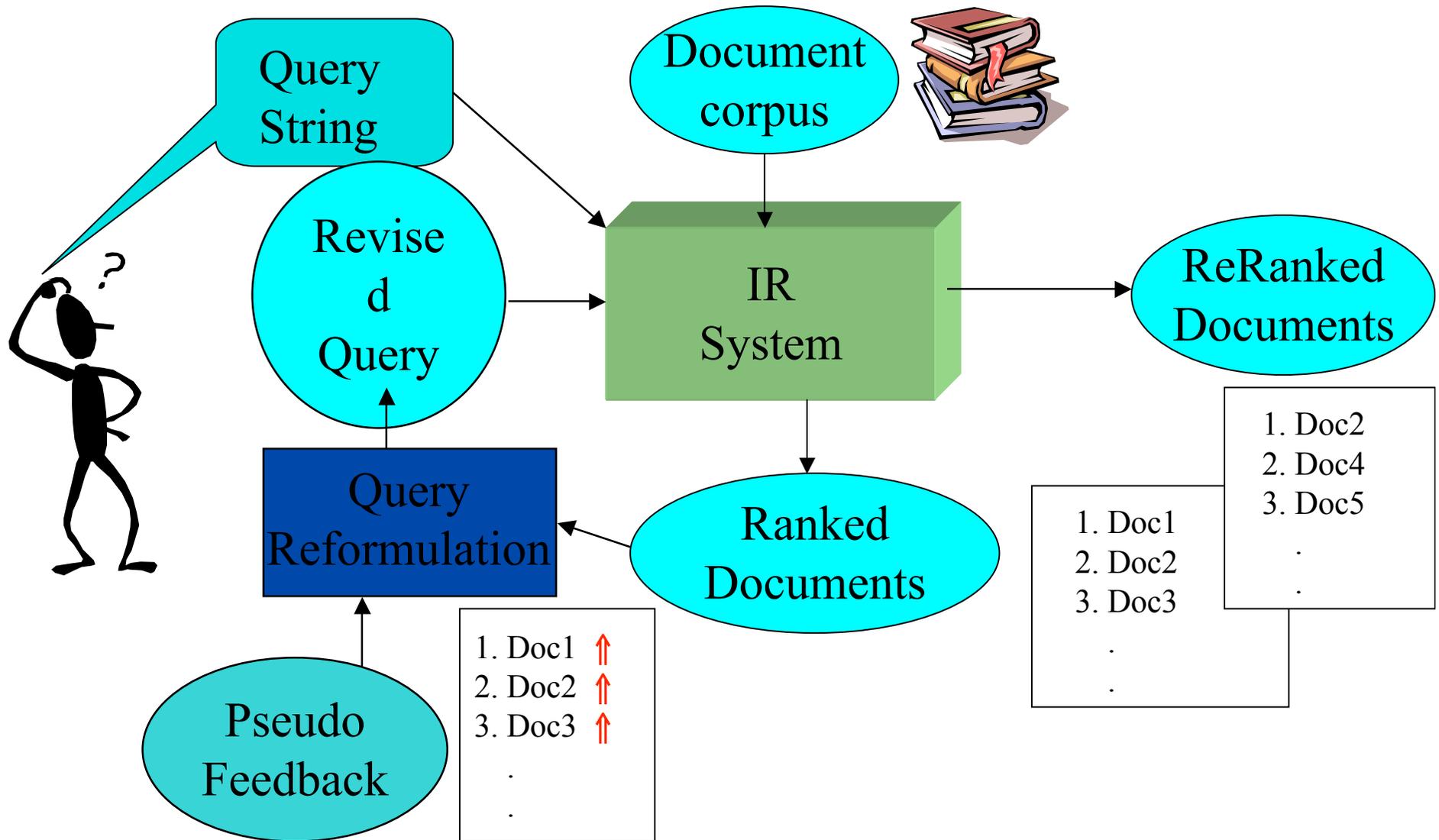
# Tuttavia il feedback esplicito non è molto usato

- Gli utenti sono a volte riluttanti.
- E' più difficile capire perché un documento sia stato selezionato (l'utente può rendersi conto di aver mal formulato la query e le sue selezioni appaiono inconsistenti con i primi risultati proposti).

# Pseudo Feedback

- Non chiedere esplicito aiuto all'utente.
- Assumi che i primi  $m$  top-ranked siano i più interessanti.
- Espandi la query includendo termini correlati con i termini della query, usando gli  $m$  top-ranked.

# Pseudo Feedback Architecture



# PseudoFeedback

- In alcune competizioni internazionali (TREC) lo pseudo-feedback si è dimostrato utile.
- Funziona ancor meglio se si usa un metodo booleano esteso (ad esempio, i termini correlati vengono aggiunti in “or”)

## Relevance Feedback on the Web

- Alcuni motori di ricerca offrono una facility *similar/related* pages (che è una forma semplificata di relevance feedback)
  - Google, Altavista (ma funziona molto male!!)
- Altri motori preferiscono non sovraccaricare l'utente:
  - Alltheweb
  - msn
  - Yahoo

# Query Expansion utilizzando un Thesaurus

- Cos'è un *thesaurus*: un glossario, con informazioni aggiuntive
- Ad esempio, oltre alle definizioni, un *thesaurus* fornisce informazioni di sinonimia e correlazione fra termini

# WordNet 3.0

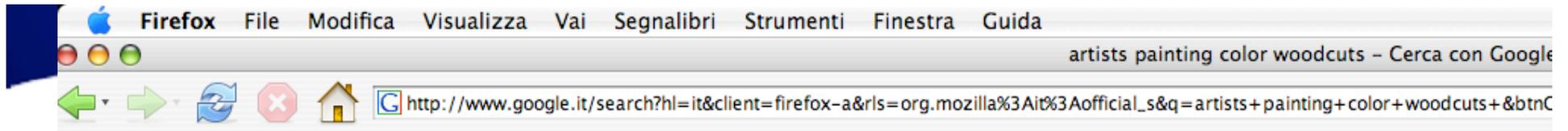
Iponimi: struttura gerarchica

- **S: (n) organization, organisation** (a group of people or institutions that are organized for a common purpose)
  - direct hyponym / full hyponym
  - member meronym
  - has instance
  - direct hypernym / inherited hypernym / sister term
    - **S: (n) social group** (people sharing some social relation)
      - **S: (n) group, grouping** (any number of entities (members) considered as a unit)
        - **S: (n) abstraction, abstract entity** (a general concept formed by extracting common features from specific instances)
        - **S: (n) entity** (that which is perceived as existing)
  - derivationally related form
- **S: (n) arrangement, organization, organisation, system** (an organized structure for arranging or classifying) "*he tried to understand their system of classification*"
  - direct hyponym / full hyponym
  - direct hypernym / inherited hypernym / sister term
    - **S: (n) structure** (the complex composition of knowledge as elements and their combinations) "*he tried to understand their system of classification*"
      - **S: (n) arrangement, organization, organisation, system** (an organized structure for arranging or classifying) "*he tried to understand their system of classification*"
      - **S: (n) form, shape, pattern** (a perceptual structure) "*the composition presents problems for the eye*"
      - **S: (n) syntax, sentence structure, phrase structure** (the grammatical arrangement of words in a sentence)
      - **S: (n) morphology, sound structure, syllable structure, word structure** (the admissible arrangements of the parts of a word)
  - derivationally related form

“fratelli”: termini correlati

# Iponimi e sinonimi migliorano la ricerca??

- Esempi nel seguito
- In generale:
  - **NO** se i termini della ricerca sono pochi e poco specifici (ambiguità genera rumore)
  - **SI** se i termini non sono ambigui (dominii tecnici)
  - **NI** se si applicano algoritmi di *word sense disambiguation*:
    - **SI** per query lunghe (molto “contesto” migliora WSD)
    - **NO** per query corte e generiche (poca precisione nella disambiguazione)



Google [Web](#) [Immagini](#) [Gruppi](#) [News](#) [altro »](#)  
artists painting color woodcuts  Cerca [Ricerca avanzata](#)  
Cerca:  il Web  pagine in Italiano  pagine provenienti da: Italia [Preferenze](#)

## Web

### [The Technique of the Color Wood-cut](#) - [ [Traduci questa pagina](#) ]

In **painting**, the **artist** has to struggle with three aspects simultaneously – his idea, form, and **color**. In the **woodcut** the two former, design and drawing, ...

[www.sharecom.ca/phillips/technique.html](http://www.sharecom.ca/phillips/technique.html) - 10k - [Copia cache](#) - [Pagine simili](#)

### [MassArt Continuing Ed | Art New England | Workshop | Layered Color ...](#) - [ [Traduci questa pagina](#) ]

Layered **Color Woodcut**/Relief Print with Jim Lee. By Jim Lee ... of **Artists'** Books, Metropolitan Museum of Art; New American **Woodcuts**, Galerie Im Hollental, ...

[www.massart.edu/at\\_massart/academic\\_prgms/continuing/ane/workshops/Lee/](http://www.massart.edu/at_massart/academic_prgms/continuing/ane/workshops/Lee/) - 8k - [Copia cache](#) - [Pagine simili](#)

### [Original woodcut relief prints, modern color woodcuts, pictures of ...](#) - [ [Traduci questa pagina](#) ]

Art for sale, **Woodcut** gallery with **color woodcuts**, english **artists**, english **woodcuts**, ... **woodcut painting**, **woodcut** landscapes, **woodcut** images, ...

[www.dreamgallery.co.uk/ain\\_dream/woodcut.htm](http://www.dreamgallery.co.uk/ain_dream/woodcut.htm) - 18k - [Copia cache](#) - [Pagine simili](#)

### [ART REVIEW: Painting With Wood and Ink, as a Master Did It - New ...](#) - [ [Traduci questa pagina](#) ]

Domjan, who began his career as a painter, treated the **color woodcut** like a **painting**. His carved blocks were covered with thick oil **paint** instead of the ...

[query.nytimes.com/gst/fullpage.html?res=9A02EFDF103FF932A15751C1A961958260](http://query.nytimes.com/gst/fullpage.html?res=9A02EFDF103FF932A15751C1A961958260) - 21k - [Copia cache](#) - [Pagine simili](#)

### [Glossary-Meyerovich Gallery](#) - [ [Traduci questa pagina](#) ]

The **artist** begins by **painting** the metal plate with a solution of sugar and ... In contemporary art, Helen Frankenthaler's **color woodcuts** are considered ...

[www.meyerovich.com/info/glossary.html](http://www.meyerovich.com/info/glossary.html) - 11k - [Copia cache](#) - [Pagine simili](#)

### [Untitled Document](#) - [ [Traduci questa pagina](#) ]

The actor Nakamura Tosha as Okaru in the drama Yoigoshin. 1928 15 **colour woodcuts** Gift of Jennifer Gordon 1998. STELLA, Frank **artist** United States born 1936 ...

[www.nga.gov.au/newacquisitions/98-99/99Inter.htm](http://www.nga.gov.au/newacquisitions/98-99/99Inter.htm) - 21k - [Copia cache](#) - [Pagine simili](#)

### [Against the grain : Helen Frankenthaler woodcuts](#) - [ [Traduci questa pagina](#) ]

DETAIL: Helen FRANKENTHALER Freefall 1993 **colour woodcut** Purchased with the ...

Firefox File Modifica Visualizza Vai Segnalibri Strumenti Finestra Guida

Taxonomy Validator

http://lcl.di.uniroma1.it/tav/index.jsp?query=color+woodcut&op=&fe=300041412

# Taxonomy Validator

Linguistic Computing LABORATORY  
University of Rome "La Sapienza"

Taxonomy

- technique
  - renderings
  - tracings
  - thumbnail sketches
  - blot drawings
  - line drawings
    - contour drawings
    - outline drawings
  - mechanical drawings
  - tool-aided drawings
  - three-tone drawings
  - tone drawings
  - cadavres exquis
  - sandpaper paintings
  - animation cels
  - animated cartoons
  - animation drawings
  - striking
  - drawings by subject type
  - prints by form
  - prints by function
  - prints by process or technique
  - color woodcuts**
    - nishiki-e
    - benizuri-e
  - sculpture by form

Concept label	color woodcuts		
Concept id	300041412		
Subclass of	Visual Works (AAT)		
Definition	(none)		
Changes	Author	Superclass	Date
	(initial)	Visual Works (AAT)	Fri Oct 20 07:30:35 CEST 2006
Options	<a href="#">Expand subtree</a> <a href="#">Collapse subtree</a> <a href="#">Up one level</a> <a href="#">Add conceptual relation</a> <a href="#">Add concept</a> <a href="#">Delete concept</a> <a href="#">Move concept</a> <a href="#">Move subtree</a>		

color woodcut search Download ontology Change ontology Logout

## Art and Architecture Thesaurus





Web Immagini Gruppi News altro »

examples of flower pieces

Cerca

Ricerca avanzata  
Preferenze

Cerca: il Web pagine in Italiano pagine provenienti da: Italia

## Web

### [Articoli accademici per examples of flower pieces](#)

 [Fractal study of magnetic domain patterns](#) - Han - Citato da 4

### [How To Press Flowers, Leaves and Herbs](#) - [ [Traduci questa pagina](#) ]

Cut **pieces** of cardboard and blotting paper (or newsprint) to fit between the ... There are many **examples of flowers** that press well on other pages of this ...

[www.preservedgardens.com/how-to-press.htm](http://www.preservedgardens.com/how-to-press.htm) - 27k - [Copia cache](#) - [Pagine simili](#)

### [JSTOR: An Exhibition of Dutch and Flemish Flower-Pieces](#) - [ [Traduci questa pagina](#) ]

All the **examples** exhibited are in exceptionally good condition, and still glow with ... EUGENE SCHLICHTER) AN EXHIBITION OF DUTCH AND FLEMISH FLOWER-PIECES.

[links.jstor.org/sici?sici=0951-0788\(193704\)70%3A409%3C192%3AAEODAF%3E2.0.CO%3B2-R](http://links.jstor.org/sici?sici=0951-0788(193704)70%3A409%3C192%3AAEODAF%3E2.0.CO%3B2-R) - [Pagine simili](#)

### [TSR Ceramics](#) - [ [Traduci questa pagina](#) ]

A third **example** with **flowers** and garlands is painted in cobalt-blue and green ... The Tareq Rajab Museum possesses a rare and unique laqabi **piece**, a jug, ...

[www.trmkt.com/cera.html](http://www.trmkt.com/cera.html) - 15k - [Copia cache](#) - [Pagine simili](#)

### [Splitcoast Technique Tutorials: Paper Tole](#) - [ [Traduci questa pagina](#) ]

Place one of the stamped **flower pieces** on a mouse pad. ... Want to see more **examples** from our gallery using similar techniques and materials? ...

[www.splitcoaststampers.com/resources/papertole.php](http://www.splitcoaststampers.com/resources/papertole.php) - 18k - [Copia cache](#) - [Pagine simili](#)

### [Flowers at EnchantedLearning.com](#) - [ [Traduci questa pagina](#) ]

Draw two **flowers** and then answer simple questions comparing them, for **example**, "Which **flower** is more colorful?" Spelling words **Flower Word Pieces** Puzzle ...

[www.enchantedlearning.com/themes/flowers.shtml](http://www.enchantedlearning.com/themes/flowers.shtml) - 45k - 18 ott 2006 - [Copia cache](#) - [Pagine simili](#)

### [The Vincent van Gogh Gallery](#) - [ [Traduci questa pagina](#) ]

2. The center "eye" of the **flower** contains different colors. In the first **example** the center is

# TAXonomy Validator



Taxonomy

- historiated portraits
- double-sided portraits
- composite portraits
- [-] **still lifes**
  - vanitas
  - ontbijtje
  - [-] banquet pieces
    - flower pieces
    - fruit pieces
    - game pieces
  - hammetje
  - tabakje
- history paintings
- shunga
- sporting prints
- seascapes
- [-] genre
- [-] figures (representations)
- [-] crosses (visual works)
- marines visual works
- townscapes
- representations
- drapery
- representations
- Buddhas
- animal paintings
- maesta
- meisho-e
- lineage

Concept label	still lifes						
Concept id	300015638						
Subclass of	visual works by subject type						
Definition	Images in which the focus is a depiction of inanimate objects, as distinguished from fruit, flowers, meat or dead game, vessels, eating utensils, and other objects, including ancient Greece and Rome; however, the subject was exploited by some 16th-century texture, and composition were valued, and the images were intended to relay alle prints, collages, and photographs. The term originally included paintings in which 1						
Carried-out-by	Italian painters						
Depicts	flowers, fruit, eating utensils, meat, vessels, dead game, inanimate objects						
Location	ancient , table, Greece, Rome						
Has-time-span	17th-century, 16th-century						
Has-type	Images						
Is-composed-of	focus, paintings, objects						
Changes	<table border="0"> <tr> <td>Author</td> <td>Superclass</td> <td>Date</td> </tr> <tr> <td>(initial)</td> <td>visual works by subject type</td> <td>Fri Oct 20 07:30:36 CEST 2006</td> </tr> </table>	Author	Superclass	Date	(initial)	visual works by subject type	Fri Oct 20 07:30:36 CEST 2006
Author	Superclass	Date					
(initial)	visual works by subject type	Fri Oct 20 07:30:36 CEST 2006					
Options	<ul style="list-style-type: none"> <li> Expand subtree  Collapse subtree  Up one level</li> <li> Add conceptual relation  Add concept  Delete concept</li> <li> Move concept  Move subtree</li> </ul>						

Firefox File Modifica Visualizza Vai Segnalibri Strumenti Finestra Guida

examples of flower pieces still life - Cerca con Google

http://www.google.it/search?hl=it&client=firefox-a&rls=org.mozilla%3Ait%3Aofficial\_s&q=examples+of+flower+pieces+still+life&btnC

Google Web Immagini Gruppi News altro »

examples of flower pieces still life Cerca Ricerca avanzata Preferenze

Cerca: il Web pagine in Italiano pagine provenienti da: Italia

### Web

[ArtLex on Still Life](#) - [ [Traduci questa pagina](#) ]  
Still life defined with images of **examples** from art history, ... **life** image of things associated with breakfast is called an ontbijt or breakfast **piece**. ...  
[www.artlex.com/ArtLex/s/still-life.html](http://www.artlex.com/ArtLex/s/still-life.html) - 11k - [Copia cache](#) - [Pagine simili](#)

[ArtLex on Still Lifes by Artists Born Between 1701 and 1850](#) - [ [Traduci questa pagina](#) ]  
**Examples** of still life pictures by artists born between 1701 and 1850: ... see thumbnail to right Peter Faes (Flemish, 1750-1814), **Flowers** by a Stone Vase, ...  
[www.artlex.com/ArtLex/s/still-life/1701-1850.html](http://www.artlex.com/ArtLex/s/still-life/1701-1850.html) - 37k - [Copia cache](#) - [Pagine simili](#)

[Flowers in a Rococo Vase](#)  
Notable **flower** painters include the 18th-century botanical artist Pierre-Joseph Redoute; the **still life** master Jean-Baptiste-Simeon Chardin, ...  
[jama.ama-assn.org/cgi/content/full/283/16/2072](http://jama.ama-assn.org/cgi/content/full/283/16/2072) - [Pagine simili](#)

[Still Life With Monkey, Fruits, and Flowers](#)  
The original of **Still Life With Monkey, Fruits, and Flowers** was a buffet **piece** more than 8 feet high by 6 feet wide, and it was priced at 1200 livres. ...  
[jama.ama-assn.org/cgi/content/full/282/23/2193](http://jama.ama-assn.org/cgi/content/full/282/23/2193) - [Pagine simili](#)

[Coastal Antiques and Art](#) - [ [Traduci questa pagina](#) ]  
There is likely no more commanding **piece** of furniture than the classic highboy, ... **Still life** subjects of **flowers** and fruit were first evident in Greek ...  
[www.coastalantiques.com/archives/september2002/ANTstilllife.html](http://www.coastalantiques.com/archives/september2002/ANTstilllife.html) - 12k - [Copia cache](#) - [Pagine simili](#)

[Title Still Life with Flowers Year 1723 Artist Jan van Huysum ...](#) - [ [Traduci questa pagina](#) ]  
This **still life** from 1723 is typical of the work of Jan van Huysum, ... He was trained by his father Justus, who painted **flower pieces** and interior ...  
[www.blogg.org/blog-43176-billet-446377.html](http://www.blogg.org/blog-43176-billet-446377.html) - 23k - [Copia cache](#) - [Pagine simili](#)

[Still Life - MSN Encarta](#) - [ [Traduci questa pagina](#) ]  
The earliest **still-life** subjects were arrangements of skulls or **flowers**, suggesting a vanitas

# Query Expansion con un Thesaurus (2)

- Per ogni termine  $t$ , in una query, espandi la query con **sinonimi e termini correlati** nel thesaurus.
- In genere i pesi dei termini aggiunti sono **più bassi**.
- In genere questo metodo aumenta la **recall**.
- Ma diminuisce la precisione, per via **dell'ambiguità semantica**

# Esempio

Google   [Ricerca avan...](#)  
Cerca:  nel Web  pagine in Italiano  pagine provenienti da: Italia

Web [+ Mostra opzioni...](#) Risultati 1 - 10 su circa 73.100.000 per

Suggerimento: [Cerca risultati solo in Italiano](#). Puoi specificare la lingua di ricerca su [Preferenze](#)

[WHO | World Health Organization](#) - [ [Traduci questa pagina](#) ]

The United Nations public **health** arm. Monitors disease outbreaks, assesses the performance of **health** systems around the globe, maintains **world** ...

[www.who.int/](http://www.who.int/) - [Copia cache](#) - [Simili](#)

[Pandemic \(H1N1\) 2009](#)

[Employment](#)

[Data and statistics](#)

[Health topics](#)

[Publications](#)

[Countries](#)

[World Health Report](#)

[Programmes and projects](#)

[Altri risultati in who.int »](#)

[WHO | Pandemic \(H1N1\) 2009](#) - [ [Traduci questa pagina](#) ]

The Review Committee, convened under the International **Health** Regulations, will set its methods of work and advise WHO of its findings. ...

[www.who.int](http://www.who.int) > ... > [Pandemic \(H1N1\) 2009](#) - [Copia cache](#) - [Simili](#)

[World Health Organization - Wikipedia, the free encyclopedia](#) - [ [Traduci questa pagina](#) ]

The **World Health Organization** (WHO) is a specialized agency of the United Nations (UN) that acts as a coordinating authority on international public health. ...

[en.wikipedia.org/wiki/World\\_Health\\_Organization](http://en.wikipedia.org/wiki/World_Health_Organization) - [Copia cache](#) - [Simili](#)

# Aggiungere sinonimi in AND peggiora



world health organization administration governanc

Cerca

[Ricerca avanzata](#)

Cerca:  nel Web  pagine in Italiano  pagine provenienti da: Italia

Web [+ Mostra opzioni...](#)

Risultati 1 - 10 su circa 6.800.000 per world health organization administrati

Suggerimento: [Cerca risultati solo in Italiano](#) . Puoi specificare la lingua di ricerca su [Preferenze](#)

## [Articoli accademici per world health organization administration governance](#)



[The concept of stewardship in health policy](#) - Saltman - Citato da 102

[... private partnerships: part la new development in health ...](#) - Buse - Citato da 136

[Mental and social health during and after acute ...](#) - Van Ommeren - Citato da 107

## [WHO | Governance](#) - [ [Traduci questa pagina](#) ]

**Governance.** The **World Health Assembly** is the supreme decision-making body for WHO. ... Its main function is to determine the policies of the **Organization**. ... Programme, Budget and **Administration** Committee · Former committees ...

[www.who.int](#) > ... > [Governance](#) - [Copia cache](#) - [Simili](#)

[CONSTITUTION OF THE WORLD HEALTH ORGANIZATION THE STATES Parties ...](#)

[health administration](#) of the Member. Article 12. Alternates and advisers may ...

[www.who.int/governance/eb/who\\_constitution\\_en.pdf](#)

[WHO | WHO Constitution](#)

[Health](#) topics · Publications · Data and statistics · Programmes and projects ...

[www.who.int/entity/governance/eb/constitution/en/index.html](#)

[Altri risultati in who.int](#) »

# ..un po' meglio se si espande con "OR"



Web [Mostra opzioni...](#) Risultati 1 - 10 su circa 90.300.000 per world health organization OR administration OR

Suggerimento: [Cerca risultati solo in Italiano](#). Puoi specificare la lingua di ricerca su [Preferenze](#)

[WHO | World Health Organization](#) - [ [Traduci questa pagina](#) ]

KEY WHO INFORMATION. Director-General Director-General and senior management.

Governance of WHO WHO Constitution, Executive Board and **World Health Assembly** ...

[www.who.int/](http://www.who.int/) - [Copia cache](#) - [Simili](#)

[Pandemic \(H1N1\) 2009](#)

[Publications](#)

[Employment](#)

[Countries](#)

[Data and statistics](#)

[World Health Report](#)

[Health topics](#)

[Programmes and projects](#)

[Altri risultati in who.int »](#)

[PDF] [LA PREVENZIONE DEL SUICIDO NELLE CARCERI World Health Organization](#)

Formato file: PDF/Adobe Acrobat - [Visualizzazione rapida](#)

di DELS LA PREVENZIONE - 2007 - [Articoli correlati](#)

Le pubblicazioni della **World Health Organization** possono essere ... In nessun caso la **World**

**Health Organization** può essere imputata per ...

[www.who.int/entity/mental\\_health/.../resource\\_jails\\_prisons\\_italian.pdf](http://www.who.int/entity/mental_health/.../resource_jails_prisons_italian.pdf)

[hotel milano accommodation milano](#)

dormire a milano in camera hotel centro città città.

[www.advertown.com/](http://www.advertown.com/) - [Copia cache](#) - [Simili](#)

Risultati da Google News su [world health organization OR ...](#)

[World Health Organization Orders Swine Flu Probe](#) - 33 minuti fa

# **Espansione della query con co-occorrenze:**

## **Automatic Global Analysis**

1. Determina la similarità fra termini usando delle statistiche pre-calcolate sulla collezione di documenti.
2. Calcola delle matrici associative (matrici di co-occorrenze) che quantificano la correlazione fra termini.
3. Espandi la query con i termini più simili, sulla base di questa matrice.

# Matrice delle associazioni

	W <sub>1</sub>	W <sub>2</sub>	W <sub>3</sub>	.....	W <sub>n</sub>
W <sub>1</sub>	c <sub>11</sub>	c <sub>12</sub>	c <sub>13</sub>	.....	c <sub>1n</sub>
W <sub>2</sub>	c <sub>21</sub>				
W <sub>3</sub>	c <sub>31</sub>				
.	.				
.	.				
W <sub>n</sub>	c <sub>n1</sub>				

$c_{ij}$ : fattore di correlazione fra termine  $i$  e termine  $j$

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

$f_{ik}$ : Frequenza del termine  $i$  nel documento  $k$

<http://www.kilgarriff.co.uk/bnc-readme.html>

# Matrice Normalizzata

- La matrice delle frequenze favorisce i termini più frequenti.
- Normalizza i fattori di associazione (Jaccard):

$$S_{ij} = \frac{C_{ij}}{C_{ii} + C_{jj} - C_{ij}}$$

- Il fattore normalizzato di associazione è 1 se due termini hanno la stessa frequenza nei documenti.

## Altre metriche di correlazione

- Le misure di correlazione precedenti non tengono conto della **prossimità** dei termini correlati in un documento.
- La seguente metrica tiene conto della prossimità.

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

$V_i$ : Set delle occorrenze di un termine  $i$  in un qualsiasi documento

$r(k_u, k_v)$ : **Distanza** in parole fra due occorrenze  $k_u$  e  $k_v$

( $\infty$  se  $k_u$  e  $k_v$  capitano in documenti diversi).

# Metriche normalizzate

- Funzione normalizzata:

$$s_{ij} = \frac{c_{ij}}{|V_i| \times |V_j|}$$

# Espansione della query con matrici di correlazione

- Per ogni termine  $i$  della query, espandi con gli  $n$  termini con i valori più alti di  $c_{ij}$  ( $s_{ij}$ ).
- Questo aggiunge solo termini semanticamente correlati (sulla base della prossimità).

## Problemi:

- Resta il problema dell'ambiguità:
  - “Apple computer” → “Apple red fruit computer”
- Poiché i termini sono in ogni caso altamente correlati, l'espansione potrebbe non aggiungere molti nuovi documenti rispetto alla query non espansa!

# Un altro metodo di espansione: **Automatic Local Analysis**

- Al momento della query, determina dinamicamente i termini correlati usando i documenti top-ranked sulla base dei criteri classici.
- **L'analisi dei termini correlati non è basata sull'intera collezione, ma solo sui documenti “localmente” recuperati sulla base della query iniziale.**
- Questo riduce il problema della ambiguità semantica, perché i documenti, essendo recuperati sulla base di **tutti** termini della query, molto probabilmente contengono ogni termine nel senso corretto per l'utente
  - “Apple computer” → “Apple computer Powerbook laptop”

## Global vs. Local Analysis

- L'analisi globale richiede di fare dei calcoli una volta per tutte.
- L'analisi locale va fatta in tempo reale, sulla base di ogni query
- Ma fornisce risultati migliori.

# Miglioramenti del metodo di analisi globale

- Espandi solo i termini che hanno una similarità al di sopra di una soglia rispetto a TUTTI i termini della query.

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

Esempio:

- “fruit” non viene aggiunto a “Apple computer” perché è non correlato a “computer.”
  - “fruit” è aggiunto a “apple pie” poiché “fruit” è correlato sia con “apple” che con “pie.”
- Inoltre si usano funzioni di pesatura più sofisticate (rispetto alla frequenza) per calcolare la correlazione (es. *Dice factor*, *mutual information*..)

# Misure di correlazione

- $\text{Dice}(x,y) = \frac{2 \text{freq}(x,y)}{\text{freq}(x) + \text{freq}(y)}$
- $\text{Mutual Information}(x,y) = \text{freq}(x,y) \log \left( \frac{\text{freq}(x,y)}{\text{freq}(x)\text{freq}(y)} \right)$
- Nota: le due formule sono in realtà **stime**, poiché la definizione “vera” di queste misure è probabilistica. Le due formule si calcolano stimando la probabilità di occorrenza di un termine  $x$  come:

$$E(p(x)) = \frac{\text{freq}(x)}{\sum_{\forall y \in D} \text{freq}(y)} \quad D \text{ è la collezione di documenti}$$

# LSI per determinare correlazioni fra termini

$$L = U\Sigma V^T \approx U_k \Sigma_k V_k^T$$

$$L L^T \approx U_k \Sigma_k (U_k \Sigma_k)^T$$

Word	Nearest neighbors
absolutely	absurd, whatsoever, totally, exactly, nothing
bottomed	dip, copper, drops, topped, slide, trimmed
captivating	shimmer, stunningly, superbly, plucky, witty
doghouse	dog, porch, crawling, beside, downstairs
makeup	repellent, lotion, glossy, sunscreen, skin, gel
mediating	reconciliation, negotiate, case, conciliation
keeping	hoping, bring, wiping, could, some, would
lithographs	drawings, Picasso, Dali, sculptures, Gauguin
pathogens	toxins, bacteria, organisms, bacterial, parasite
senses	grasp, psyche, truly, clumsy, naive, innate

Il metodo LSI può essere utilizzato per determinare automaticamente “clusters” di termini simili (Shutze 1998)

# Google query expansion

- Word **stemming**: translator -> translator, translation
- **Acronimi**: NATO -> North Atlantic Treaty Organization (pericoloso... Northern Arts Tactical Offensive)
- **Errori** di digitazione: wigets -> widgets
- **Sinonimi**: solo se appare evidente che la parola è usata in modo improprio (information lost -> loss)
- **Traduzione** (organizzazione mondiale sanità -> world health organization)
- **Related Search** (migliorata dal 2009 dopo l'accordo con Orion)

Searches related to: **comedy of the 80s**

[eddie murphy comedy](#)

[80's movies comedy](#)

[80s comedy films](#)

[tom hanks comedy](#)

[bbc comedy](#)

L'algoritmo di Ori Allon non è noto, ma si basa sull'identificare pagine strettamente correlate alla keyword di ricerca e estrarre da queste altre Keywords.

# Conclusioni

- L'espansione delle query può migliorare le prestazioni, in particolare la **recall** (ridurre i “silenzio”).
- Tuttavia, il problema principale resta quello **dell'ambiguità semantica**, che può influire negativamente sulla precisione.
- Metodi di WSD (word sense disambiguation) possono essere usati (con dei limiti) per selezionare il senso corretto