Information Retrieval Systems

These include:

- Web search engines, e.g. Google
- Question-answering systems, e.g. Ask.com
- Pseudo-NL document retrieval, e.g. Canadian World Encyclopedia, MS Office Assistant
- Legal/Medical professional research services, e.g. Lexis-Nexis
- TREC (Text REtrieval Conference) systems, more recently DUC (Document Understanding Conference) has been good showcase of these
What's the name of the disease that causes red bumps all over your face?

Name Definition
Trouble finding the perfect name? We can help with our new BabyName!
www.todayparen.com

Name Definitions
Search from 100s of names for free by gender, culture, meaning & more.
www.BabyZone.com

Diagnose Your Symptoms
Tell us about All your symptoms. Let us give you the full picture.
www.DiagnoseMe.com

Little under the skin bumps all over my face! [Archive] - ...
[Archive] Little under the skin bumps all over my face! Acne...
www.healthboards.com/boards/archive/index.php/t-255835.... · Cached

Red bumps all over - Hepatitis C - MedHelp
This is a discussion on MedHelp about Red bumps all over. Community members of MedHelp provide help, support, guidance and discussion around ...
www.medhelp.org/forums/hepatitis/messages/45284.html

Red itchy bumps face, legs, arms...now itchy rash on abdomen. - ...
This is a discussion on MedHelp about Red itchy bumps face, legs, arms...now itchy rash on abdomen. Community members of MedHelp provide help, ...
www.medhelp.org/forums/dermatology/messages/34650.html · Cached
More Results from www.medhelp.org

Chicken Pox | Diseases | Illnesses | Virus | Contagious
Infections
Ask Jeeves—example (1)

You asked: disease that causes red bumps all over your face

I have found answers to the following questions:

What disease or condition could be indicated by the symptom swelling or lump?

Where can I find the box score & recap for the Major League Team the Cincinnati Reds?

Where can I find the official Web site of the Cincinnati Reds?

Where can I find information about diseases and conditions?

Where can I see an image of the human facial nerves?

Where can I buy Cincinnati Reds merchandise?

Where can I buy tickets to an upcoming Cincinnati Reds home game?

What's the disease that causes red bumps all over your face?
What's the name of the disease that causes red bumps?

Red, Itchy Bumps In Various Areas of Body - Dermatology...
Over the past 2 months these red bumps have appeared in clumps on almost every... What's the best HIV Test? Mar 04 by Jose Gonzalez... Only your physician can provide specific... www.medhelp.org/posts/Dermatology/Red-itchy-Bumps-In-Various-Areas-of-Body/show244942 - Cached page

Bumps Upper Lip - Wellsource
... disease can this have caused it Recan gum disease cause white bumps ... cracks in the corners of my mouth and also red bumps ... Care - brown pigment and loss of pinnaent of face I all... stanford.wellsphere.com/wellmix969/bumps-upper-lip - Cached page

Sore red pimple like bumps on vagina - Dermatology...
ALWAYS check with your personal ... I noticed red bumps after shaving one day and it didn't really bother me cause I got that suntimes ... and all that; I've had red bumps all over the... www.medhelp.org/posts/Dermatology/Sore-red-pimple-like-bumps-on-vagina/show6865 - Cached page

How to Get Rid of Razor Bumps - howtogogetridofstuff.com
Be Gentle On Your Face. The reaction to razor bumps that many men ... it scrape yourself or shave your hair unevenly causing those red razor bumps. ... for razor burn, > just rub all over... www.howtogogetridofstuff.com/health/how-to-get-rid-of-razor-bumps - Cached page

Chicken Pox | Diseases | Illnesses | Virus | Contagious...
Chicken pox is a human disease, so how did it get its name? Well, the disease causes the skin to break out in red bumps ... suck because pimples-like bumps grow all over your face... www.kidzworld.com/article/4854-chicken-pox-101 - Cached page

Under Skin Bumps | LIVESTRONG.COM
Razor bumps can cause a red rash and make your skin feel chafed; ...ed up with a face full of red, irritated bumps instead of a close... What's Your Goal for the Weekend?? Alright... www.livestrong.com/under-skin-bumps - Cached page

Vaginal Bumps and Lumps
I have had a small under the skin red bump ... Okay, mine are all scabbed cause I scratched these bumps; ... it is and it is better to know what's going on with your... womenandhealth.about.com/0/2006/04/06/vaginal-bumps-and-lumps.htm - Cached page

Neck pain - WrongDiagnosis.com
Face/Neck/Chest pain for over 3 yrs after ... Itchy red bumps all over body ... related to your Neck pain Take Survey View Results. Causes of Neck pain listed in Disease Database... www.wrongdiagnosis.com/sym/neck_pain.htm - Cached page
I had one but the wheel fell off

Head fell off
Answers to all your questions about the Head.
www.righthealth.com/Head

2ND LD: Yurikamome wheel fell off as result of metal fatigue
govt
2ND LD: Yurikamome wheel fell off as result of metal fatigue: gov't ... linked to the wheel falling off right after the train took off from ...
asia.news.yahoo.com/060417/kyodo/d8h1qs7g9.html · Cached

Dave petrucci's Profile - GameSpot
... but critics will shove one award in its hands despite multiple nominations--best ... Anyways a short one this time, see you around.
www.gamespot.com/users/Dave_petrucci/

my wheel fell off oh man - Ford Truck Enthusiasts Forums
... needless to say one of the tires on the trailer (tandem axle) came off. we had to catch it and knock it ... only one lug nut which was ...
www.ford-trucks.com/forums/193334-my-wheel-fell-off-oh-

HELP my wheel fell off - Ford Truck Enthusiasts Forums
... the wheel fell off again. This time it was different though. The lugnuts were still hooked to the studs. The studs had actually pulled ...
More Results from www.ford-trucks.com

And then my wheel fell off. Or, There's Madness in My Method.
The Wheelbarrow song - Wikipedia, the free encyclopedia
I had a wheelbarrow, the wheel fell off. I had a ... pitch near the corner flag, the wheel fell off and he spoke to both County fans and a terrace chant was born, sang from one ...
en.wikipedia.org/wiki/The_Wheelbarrow_song · Cached page

London Slang - I
I had one but the wheels fell off - a bizarre sarcastic reply that you give when you don't understand what the other person is saying (usually technical). e.g. www.londoonslang.com/ctlf/ · Cached page

Trailer Wheel Almost Fell Off - Close Call [Archive]
[Archive] Trailer Wheel Almost Fell Off - Close Call Trailers ... One lug bolt was sheared off. Another was nearly gone. Another had a missing lug nut. www.mastercraft.com/teamtalk/archive/index.php/t-5109.html · Cached page

New land rover in the family. [Archive] - Land Rover...
My dad just bought a "new" 1987 88 with a soft top. Just about wrecked it on the way home when he got off the freeway and his left rear wheel almost fell off. had one stud left on ... forum.landrover.net/archive/index.php/t-74377.html · Cached page

The Wheel Fell Off « Mellotone70Up Blog
For reasons I've never understood, the song Notts County fans sing most often begins, "I had a wheelbarrow, the wheel fell off." Or is it wheels? mellotone70up.wordpress.com/2013/02/19/...-the-wheel-fell-off · Cached page

I fell off the ferris wheel and now I am paralyzed and ... I had just shaved. I smelled like musk. I exuded ... or reciprocate, or give me some. I fell off the ferris wheel and now I am ... Things that no one told you... until now, because I'm ... www.everything2.com/title/I+fell+off+the+ferris+wheel+and+now+I+am+paralyzed+and+paralyzed · Cached page

Renault gets race ban for wheel loss - F1 | ITV Sport ...
... from the pit stop position without one of the retaining devices for the wheel ... lot of vibration and it felt like I had a puncture, but then a few corners later the wheel flew off ... www.itv.com/news_article.aspx?id=46489 · Cached page

Safety fears raised after plane wheel falls off - ABC ...
... one of the nose wheels fell off a ... fell off a plane at Melbourne Airport on Saturday. The plane was preparing for take-off when a ground engineer noticed one of the wheels had ... www.abc.net.au/news/stories/2009/07/27/2833051.htm · Cached page

Wheelbase - Wheelbase operator loses O-licence for 3 ...
The stud threads were virtually undamaged, indicating that the wheelnuts had unwound gradually until the wheels fell off. Only one wheel.
Microsoft Word - Dissertation - Final Draft - B.doc (PDF File)
They actively courted Diefenbaker, who then saw an opportunity to expand Canadian economic and political interests in the region. The personal relationships that Diefenbaker was able to establish with some of these leaders would be instrumental in developing trade and other linkages.

Spyware
Get Enterprise-Wide Protection with Microsoft® Forefront. Try it Free.
Microsoft.com/Forefront

Wax
We've Got Your Style! Hair Styling Videos & Step-by-step instructions
www.garnier.ca/yourstyle

Cheating Spouse
Love'em but Don't Trust'em? Do you want to know for sure?
www.TattleTalesInvestigations.com

Who is Lookup
Global Who is Lookup for domain names and their owners
www.who.is

Microsoft Word - 17 - gomezko outline (PDF File)
Dief knew...but did not fire him) Pearson and Diefenbaker - blackmail March 1966 Liberals raise the "Monsieurneige case" in House of Commons: Refers to GERDA MUNSGINGE (Prostitute? German spy?

The Nexus of Ashes: Moral Puritans of the Hateful Left
May 9, 2008 ... Conservative Pierre Savigny, one of John Diefenbaker's ministers, ... Now, Lulu's provided us with an absolute embarrassment of riches in ... Pierre Savigny sleeping with an East German Spy is pretty juicy stuff — but ...

Famous Foreign Masons
John Diefenbaker ... Eighteenth-century songwriters in search of easy rhymes fell in love with his name, feminized with an "o" to make it agree with "machina." Try doing that with "Schmidt," the name of the drunken German carpenter who turned Guillotin's blueprints into a reality. ... Alchemist, cabalist, and spy,

Tories dismiss 'wild' allegations of secret Bernier meeting
OTTAWA — The Liberal opposition suggested Wednesday that Canada's spy agency met with the Prime Minister's Office in early May to discuss the "conduct" of former Foreign Affairs minister Maxime Bernier. ... It is that way because of Conservative governments, people like John Diefenbaker and Brian Mulroney, who led on the...

CBC News Indepth: Canadian Government
3. The Gorda Munsinger scandal: The guilty parties were already out of office by the time the Canadian public learned that some Progressive Conservative cabinet ministers had been consorting with an East German played girl who may have been a KGB spy.

Citizen Fink
Who was the German spy who embarrassed Diefenbaker

April - John F. Kennedy Presidential Library & Museum
The East German press service said the move was in response to the U.S. embargo on spy charges. (pg. 12)
www.jfklibrary.org/Historical-Resources/Archives/Reference-Desk/New-York-Times+-Chronology/... Cached page

Cold War Times - August 2008 - Vol. 8, Issue 3
... Museum continues to work with the Diefenbaker Museum in Ottawa, Canada, the Atomic Burker in Hamekop, Germany, and the International Spy ... Simon - The Cold War Museum - German

Harry S. Truman - Wikipedia, the free encyclopedia
On November 11, 1918 his artillery unit fired some of the last shots of World War I into German positions. The war was a transformative experience that brought out Truman's
on.wikipedia.org/wiki/Harry_S._Truman  Cached page

Pay Articles from June 1958 Part 6 - Site Map - The...
DIEFENBAKER GETS ASSURANCE BY U.S.: Eisenhower Pledges Backing...SOVIET SPY DEVICE WAS FOUND IN 1958, Planted in U.S. Eight-Story Company Regains Permit – New German Ship ...

shwi.alternatehistory.com
Sorge, alleged Soviet spy and former German diplomat. It is seemingly a season for invasions; on August 9, 1945, the invasion of Italy begins.
shwi.alternatehistory.com/For%20All%20Time.txt  Cached page

Pay Articles from November 1958 Part 6 - Site Map...
German Flies for Remark: Plan Offered by Britain: Soviet Complain to U.N. On Budget ... COURT INTRUDER THREATENS SOBUE; Convicted Spy: a Witness in Pejury Trial. Accosted ...

Dear Readers
If a Canadian spy could not believe it, then how much ... in the U.S. campaign against Prime Minister Diefenbaker ... D.Q. Cスター, “Behind the German Lines,” Readers Digest 37 ...
www.peaco.ca/questionsofushostility.htm  Cached page

My Autobiography ... From Jimmy to Jim
I was terribly embarrassed. It cost me a pound, but it was ... German bombers came over the first night and their very ... He was arrested as a spy in the Spanish-American War ...
www.permanent.com/genealogy/jimmy.html  Cached page

For All Time - The Postwar World (parts 35-79)
Who was the German spy who embarrassed Diefenbaker?
The close relationship between human and animal has persisted since prehistoric times but societies for the protection of animals have been in existence for only about 125 years in the Western world.

Development of Animal Rights

During the Victorian era, the antivissection movement developed to oppose science's use of animals in biomedical research. The battle between vivisectionists and antivivisectionists resulted in the UK's Cruelty to Animals Act (1876). This legislation heralded a clear victory for the antivivisection movement and imposed upon Britain's medical scientists a system of licensing, inspection and bureaucratic practices. The need for the act has been debated ever since and in 1986 was strengthened and replaced by the Animals and Scientific Procedures Act.

After WWII, biomedical research escalated tremendously. New drugs, treatments and techniques were...
Central Experimental Farm
The Central Experimental Farm (CEF) was created in 1886. It is located along the historic Rideau Canal in central Ottawa and covers an open space of 427 ha. The Farm site was initially established on the outskirts of Ottawa but...
Article from The Canadian Encyclopedia

Devon Island Miocene Fossils
Near the north coast of DEVON Island (NU) at about 750 north latitude, there occurs an unusual, roughly circular geologic structure originally called Haughton Dome; its origin puzzled scientists for many years. Finally, in 1978...
Article from The Canadian Encyclopedia

Harvestmen
Harvestmen are ARACHNIDS of the order Opiliones. Harvestmen is the preferred term but they are commonly called daddy-long-legs - even though many have short legs. Moreover, "daddy-long-legs" is also used for cellar...
Article from The Canadian Encyclopedia

First Peoples' Music
Introduction The Aboriginal people (FIRST NATIONS, INUIT, MÉTIS) who reside in what is now Canada have diverse cultural traditions that are reflected in the variety of their musical genres and styles. Aboriginal...
Article from The Canadian Encyclopedia

Architectural History: Early First Nations
The people of Canada's FIRST NATIONS developed rich building traditions thousands of years before the arrival of the first Europeans. Canada contained five broad cultural regions, defined by common climatic, geographical and...
Article from The Canadian Encyclopedia

Theatre Design to 1950
Canada's earliest theatre buildings are known only through written accounts and other archival sources. The first is thought to have been the New Grand Theatre, erected in 1789 for the British military garrison in Halifax. It...
Article from The Canadian Encyclopedia

Native People: Subarctic
The area of Subarctic cultures lies largely within the 5 million km 2 zone of northern or boreal coniferous forest that extends from the arctic tundra to the mountains, plains or deciduous forest in the south and across North...
Article from The Canadian Encyclopedia

Native People: Northwest Coast
The Canadian portion of the Northwest Coast is a region of extremes in topography, from wide beaches to deep fjords and snowcapped mountains. Temperatures are moderate, the January mean above freezing and July less than 18º...
Article from The Canadian Encyclopedia

Yellowknife (NWT)
Yellowknife, NWT. Yellowknife is the capital of the NORTHWEST TERRITORIES and the territory's only city. It sits on the ancient rock of the Canadian SHIELD, on the north shore of GREAT SLAVE LAKE. Due to its northerly location...
Article from The Canadian Encyclopedia

Work
At different times in different cultures, work has been considered a divine
What is the name of the disease where you get red bumps all over your face?
Example TREC Queries

• How often were the peace talks in Ireland delayed or disrupted as a result of acts of violence?

• Who was involved in the Schengen agreement to eliminate border controls in Western Europe and what did they hope to accomplish?

• Is there contemporary interest in the Greek philosophy of stoicism?

• Identify instances of attacks on humans by Africanized (killer) bees.
Terminology

Document $\longrightarrow$ one of book, article, paragraph, etc.
Collection $\longrightarrow$ corpus (of documents)
Term $\longrightarrow$ word type
Stop word $\longrightarrow$ function word (generally)
Queries

• Specification of subset of documents in a collection.

• Historically, conceived of as either very structured logical language (Boolean connectives etc.) or (in research contexts) well-formed NL.

• Now, particularly in web search engines, they’re often just streams of syntactically disconnected keywords.

• Usually given by user, but can be enhanced by system, e.g.:
  Q: shipwreck salvaging (the recovery or attempted recovery of treasure from sunken ships)
  – dictionary definition
  – thesaurus
  – WordNet

• IR systems depend heavily on quality of the query.
Major IR Tasks


2. Question Answering: Find document with answer to question, or find answer itself.

3. Information Extraction: Find answers to specific pre-specified questions within selected bulk documents (not actually IR).
Evaluation

Precision and recall aren’t quite good enough — assuming there are 5 relevant documents to a query out of a collection of 100, all of these systems have 50% precision and 50% recall:

<table>
<thead>
<tr>
<th>Rank</th>
<th>System A</th>
<th>System B</th>
<th>System C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>2.</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>3.</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>4.</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>5.</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>6.</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>7.</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>8.</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>9.</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>10.</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
</tbody>
</table>
Cutoff precision

Measure precision down to some cutoff, e.g. first 5 docs

<table>
<thead>
<tr>
<th>Rank</th>
<th>System A</th>
<th>System B</th>
<th>System C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>2.</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>3.</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>4.</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>5.</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>6.</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>7.</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8.</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>9.</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10.</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CP</td>
<td>0%</td>
<td>100%</td>
<td>60%</td>
</tr>
</tbody>
</table>
Uninterpolated Average Precision

Measure precision at every relevant document returned:

<table>
<thead>
<tr>
<th>Rank</th>
<th>System A</th>
<th>System B</th>
<th>System C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>2.</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>3.</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>4.</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>5.</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>6.</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>7.</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>8.</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>9.</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>10.</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>UAP</td>
<td>0.354</td>
<td>1.0</td>
<td>0.679</td>
</tr>
<tr>
<td>1 ↔ 2</td>
<td></td>
<td></td>
<td>0.579</td>
</tr>
<tr>
<td>9 ↔ 10</td>
<td></td>
<td></td>
<td>0.668</td>
</tr>
</tbody>
</table>
(Interpolated) Average Precision

Measure precision at various levels of recall and average.

11-point Average Precision: Use levels 0%, 10%, 20%, ..., 100%.

<table>
<thead>
<tr>
<th>Rank</th>
<th>System A</th>
<th>System B</th>
<th>System C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>2.</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>3.</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>4.</td>
<td>no</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>5.</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>6.</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>7.</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>8.</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>9.</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>10.</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>11-AP</td>
<td>0.337</td>
<td>1.0</td>
<td>0.708</td>
</tr>
<tr>
<td>1 ↔ 2</td>
<td></td>
<td></td>
<td>0.572</td>
</tr>
<tr>
<td>9 ↔ 10</td>
<td></td>
<td></td>
<td>0.698</td>
</tr>
</tbody>
</table>
Uninterpolated vs. Interpolated Average Precision

- Interpolated is *more sensitive* to recall: good to use when comparing 2 algorithms or datasets for which recall differs widely
- Uninterpolated is *less sensitive* to recall: good to use when only difference is relevance order, or when you only want to test precision.

...but the best “measure” is still a precision/recall graph.
Relevance Feedback

User judges results of the search; Query is adjusted accordingly.

Approaches:

• user can judge each retrieved document

• “click capture” on web-sites: clicks count as votes for relevance

• system automatically assumes top-ranked documents are relevant (*pseudo-relevance feedback*)

Objective: move query vector towards relevant ones and away from irrelevant ones:

\[
\vec{q}_{i+1} = \vec{q}_i + \frac{\beta}{R} \sum_{j=1}^{R} \vec{r}_j - \frac{\gamma}{S} \sum_{k=1}^{S} \vec{s}_k,
\]

where \(0 \leq \beta, \gamma \leq 1, \beta + \gamma = 1\)
Inverted Index

This is the basic data structure for document retrieval: given a word, return all documents that contain it in:

- title
- abstract or title
- abstract or title plus chosen index terms
- whole text
Vector Space Model for Retrieval

- Each document represented by vector
- Query also represented by vector
- Each dimension of vector corresponds to a non-stop-list term (word/phrase)
  - $0/1$: present/absent
  - $N$: number of tokens
    (of term $i$ in document $j$: $tf_{ij}$)
  - $\mathbb{R}_0^+$: dampened weight, e.g.
    $$= \begin{cases} 1 + \log t f_{ij} & t f_{ij} > 0 \\ 0 & t f_{ij} = 0 \end{cases}$$
  - Sometimes add “bonus” for being in title/abstract
- Once vectors are determined, can use cosine method etc. to determine similarity to query
**tf.idf**

Many frequent terms in one document occur frequently in all documents. These shouldn’t count for as much as a term that occurs frequently only in one document.

\[
\text{weight}(i, j) = tf_{ij} \times idf_i
\]

where

- \( idf_i = \log\left(\frac{N}{n_i}\right) \) (\textit{inverse document frequency})
- \( N = \text{number of documents} \)
- \( n_i = \text{number of documents in which term } i \text{ occurs} \)
Problems

- Size of query (part. with query expansion)
  – can normalise lengths
- Ambiguous words cause spurious matches
  – can add topic-related words
  – but this means we can determine user-intended sense
  – has to be very high accuracy or it hurts rather than helps
- Vectors of terms, regardless of actual value computed, don’t capture indirect connections through \textit{semantic relatedness}
  – what we need are vectors of ideas/concepts
  ...
Answer: Latent Semantic Indexing

Which concepts do we use? How do we assign terms to them? We don’t know, but we can guess:

- there are far fewer concepts than terms
- terms that co-occur frequently are likely to reify the same concepts
- so we reduce dimensionality of document matrix using co-occurrence
- use *singular value decomposition* (SVD)
How well does LSI work?

- $k$ too big: won’t capture underlying semantic space
- $k$ too small: too much information lost
- No known analytic method for deriving $k$ — experiment
- Term matching typically improves by 0–30% precision
- Works better in domains with lots of synonymy and ambiguity
- Works better when recall is high

Disadvantages:

- SVD is computationally expensive — only worth it for huge document collections
- inverted index is not feasible