

Lexical Knowledge Bases

Natural Language Processing

Lexical Knowledge Bases

- A *lexicon* goes beyond a “dictionary” in that it is machine-readable and carries the information needed to perform major NLP functions, such as:
 - Parts of speech
 - Inflections
 - Transitive vs. intransitive verbs

Lexical Knowledge Bases

- A *lexical knowledge base* (“*lexical KB*”) goes even further, breaking words into senses, and linking senses to senses via relations such as:
 - Synonym/antonym
 - Hyponym/hypernym
 - Holonom/meronym

WordNet

Our “go-to” lexical knowledge base (KB)

WordNet Search - 3.1
- [WordNet home page](#) - [Glossary](#) - [Help](#)

Word to search for:

Display Options: ▾

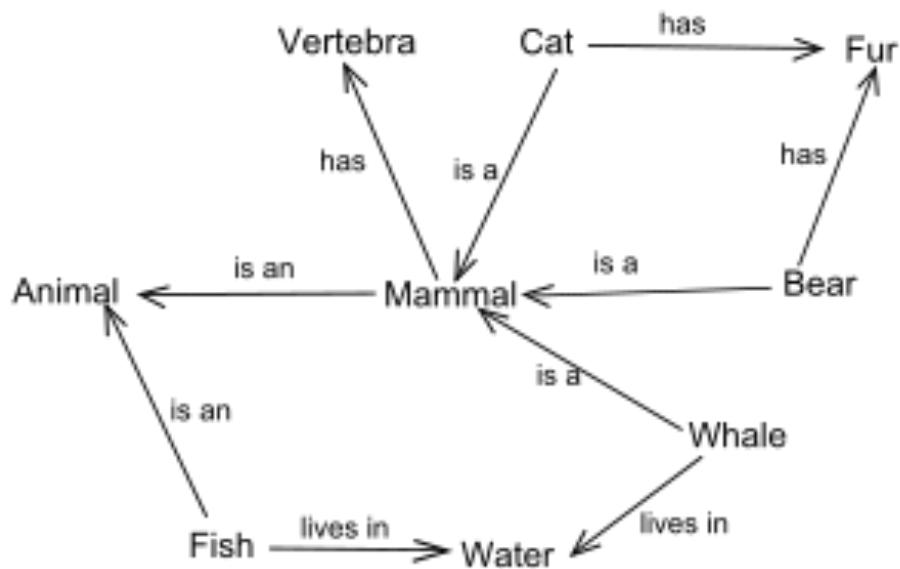
Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations
Display options for sense: (gloss) "an example sentence"

Noun

- S: (n) wordnet (any of the machine-readable lexical databases modeled after the Princeton WordNet)
- S: (n) WordNet, [Princeton WordNet](#) (a machine-readable lexical database organized by meanings; developed at Princeton University)

WordNet

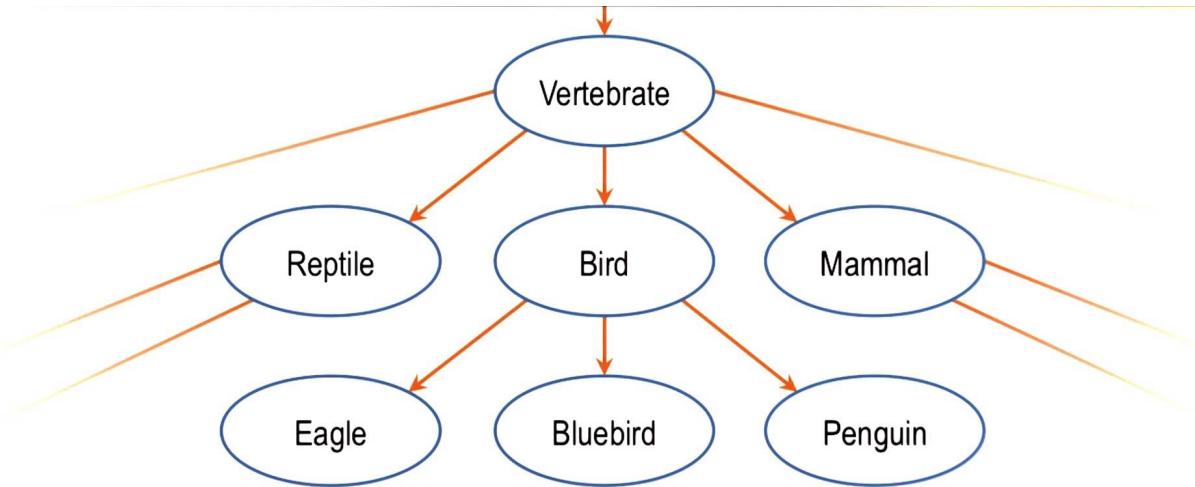
WordNet is an example (in limited form) of a *semantic network*, which means it can tell you relations similar to these:



But WordNet is limited largely to just the “is a” and “has part / is part” relations.

Concept Hierarchies

It's possible
to put a
semantic
network into
a strict
hierarchy like
this, using
just the “is-a”
relation.



Improving Feature Extraction

- We can use a lexical KB's concept hierarchy; we can improve our low-level feature extraction.
- Default ontological distance (number of nodes “travelled” through the hierarchy to get from node A to node B):
 - Compute default ontological distance of any two headwords in WordNet.
 - Without knowing which word sense, just check all senses and find the closest/longest distance via hyponym trees.
 - This gives us a rough sense of distance in “lexical space” between terms.

**Can you guess the ontological distance
in WordNet of “chair” and “table”?**

Example of Lexical Distance

Can you guess the ontological distance in WordNet of “chair” and “table”?

Comparing noun sense 1 of “chair” with noun sense 2 of “table”

WordNet 2.1 Browser

File History Options Help

Search Word: chair

Redisplay Overview

Searches for chair: Noun Verb

Senses: []

4 senses of chair

Sense 1

chair -- (a seat for one person, with a support for the back; "he put his coat over the back of the chair and sat down")
=> seat -- (furniture that is designed for sitting on; "there were not enough seats for all the guests")
=> furniture, piece of furniture, article of furniture -- (furnishings that make a room or other area ready for occupancy; "they had too much furniture for the small apartment"; "there was only one piece of furniture in the room")
=> furnishing -- ((usually plural) the instrumentalities (furniture and appliances and other movable accessories including curtains and rugs) that make a home (or other area) livable)
=> instrumentality, instrumentation -- (an artifact (or system of artifacts) that is instrumental in accomplishing some end)
=> artifact, artefact -- (a man-made object taken as a whole)
=> 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")
=> object, physical object -- (a tangible and visible entity; an entity that can cast a shadow; "it was full of rackets, balls and other objects")

"Hypernyms (this is a kind of...)" search for noun "chair"

WordNet 2.1 Browser

File History Options Help

Search Word: table

Redisplay Overview

Searches for table: Noun Verb

Senses: []

Sense 2

table -- (a piece of furniture having a smooth flat top that is usually supported by one or more vertical legs; "it was a sturdy table")
=> furniture, piece of furniture, article of furniture -- (furnishings that make a room or other area ready for occupancy; "they had too much furniture for the small apartment"; "there was only one piece of furniture in the room")
=> furnishing -- ((usually plural) the instrumentalities (furniture and appliances and other movable accessories including curtains and rugs) that make a home (or other area) livable)
=> instrumentality, instrumentation -- (an artifact (or system of artifacts) that is instrumental in accomplishing some end)
=> artifact, artefact -- (a man-made object taken as a whole)
=> 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")
=> object, physical object -- (a tangible and visible entity; an entity that can cast a shadow; "it was full of rackets, balls and other objects")
=> physical entity -- (an entity that has physical existence)
=> entity -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

"Hypernyms (this is a kind of...)" search for noun "table"

Distance = 3

Example of Lexical Distance

Can you guess the ontological distance in WordNet of “chair” and “table”?

Comparing noun sense 1 of “chair” with noun sense 2 of “table”

chair -- (a seat for one person, with a support for the back; "he got up from the chair and sat down")

=> seat -- (furniture that is designed for sitting on; "there were not enough seats for all the guests")

=> furniture, piece of furniture, article of furniture -- (furnishings that make a room or other area ready for occupancy; "they had too much furniture for the small apartment"; "there was only one piece of furniture in the room")

=> seat -- (furniture that is designed for sitting on; "there were not enough seats for all the guests")

=> furniture, piece of furniture, article of furniture -- (furnishings that make a room or other area ready for occupancy; "they had too much furniture for the small apartment"; "there was only one piece of furniture in the room")

=> furnishing -- ((usually plural) the instrumentalities (furniture and appliances and other movable accessories including curtains and rugs) that make a home (or other area) livable)

=> instrumentality, instrumentation -- (an artifact (or system of artifacts) that is instrumental in accomplishing some end)

=> artifact, artefact -- (a man-made object taken as a whole)

=> 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")

=> object, physical object -- (a tangible and visible entity; an entity that can be seen and touched)

"Hypernyms (this is a kind of...)" search for noun "chair"

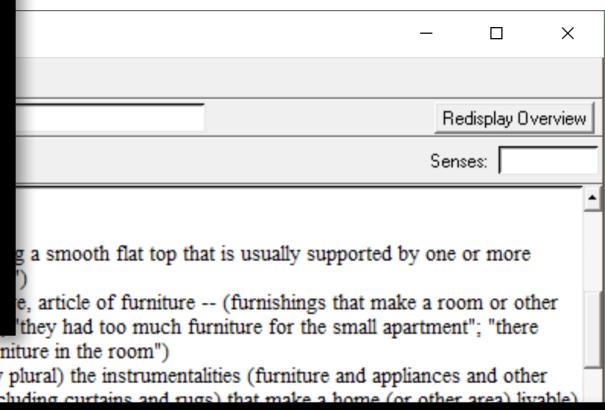


table -- (a piece of furniture having a smooth flat top supported by vertical legs; "it was a sturdy table")

=> furniture, piece of furniture, article of furniture -- (furnishings that make a room or other area ready for occupancy; "they had too much furniture for the small apartment"; "there was only one piece of furniture in the room")

Distance = 3

DataScience@SMU

Lexical Knowledge Bases

Natural Language Processing

Comparative Term Distance

- So, in line with our intuitions, “chair” is closer to “table” than it is to “golf ball.”



If you were to check this distance between the most frequent terms across two different documents, you could add up the cumulative distance, and *voila!* You'd have an automated measure of the general *document distance*.

Limits of This Approach

- We don't know which sense of "chair" was intended in each occurrence.
- We would have to disambiguate word senses, which we'll talk about later.



Furniture
to sit on



Leader of an
academic
department



Instrument of
capital
punishment

Polysemy vs. Monosemy

- Monosemy = having only one sense
- Polysemy = having more than one sense

Unfortunately, many words in English are polysemous.

Polysemy in WordNet 3.0

POS*	Monosemous	Polysemous Words	Polysemous Senses
	Words and Senses		
Noun	101863	15935	44449
Verb	6277	5252	18770
Adjective	16503	4976	14399
Adverb	3748	733	1832
Totals	128391	26896	79450

*part of speech

Polysemy in WordNet 3.0

POS*	Average Polysemy	Average Polysemy
	Including Monosemous Words	Excluding Monosemous Words
Noun	1.24	2.79
Verb	2.17	3.57
Adjective	1.40	2.71
Adverb	1.25	2.50

*part of speech

DataScience@SMU

Building or Extending a Lexical Knowledge Base

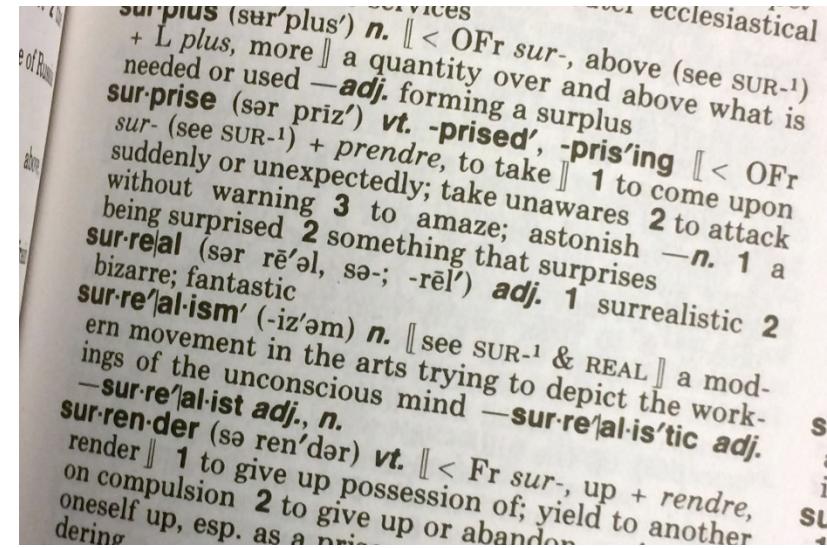
Natural Language Processing

Resources for Creating or Extending Lexical Knowledge Bases

WordNet's a good start, but it's not enough.
How do we expand our lexical KB?

Three types of resources:

- Dictionaries
- Encyclopedias
- Taxonomies



Resources for Creating or Extending Lexical Knowledge Bases

- Main vocabulary resources
 - Baseline: Princeton WordNet
 - Supplementary: Getty Vocabularies, Wiktionary, Urban Dictionary



Can We Utilize Urban Dictionary?

First we want to check their “robots.txt” file to see if it permits crawling the site.
Fortunately, it does!

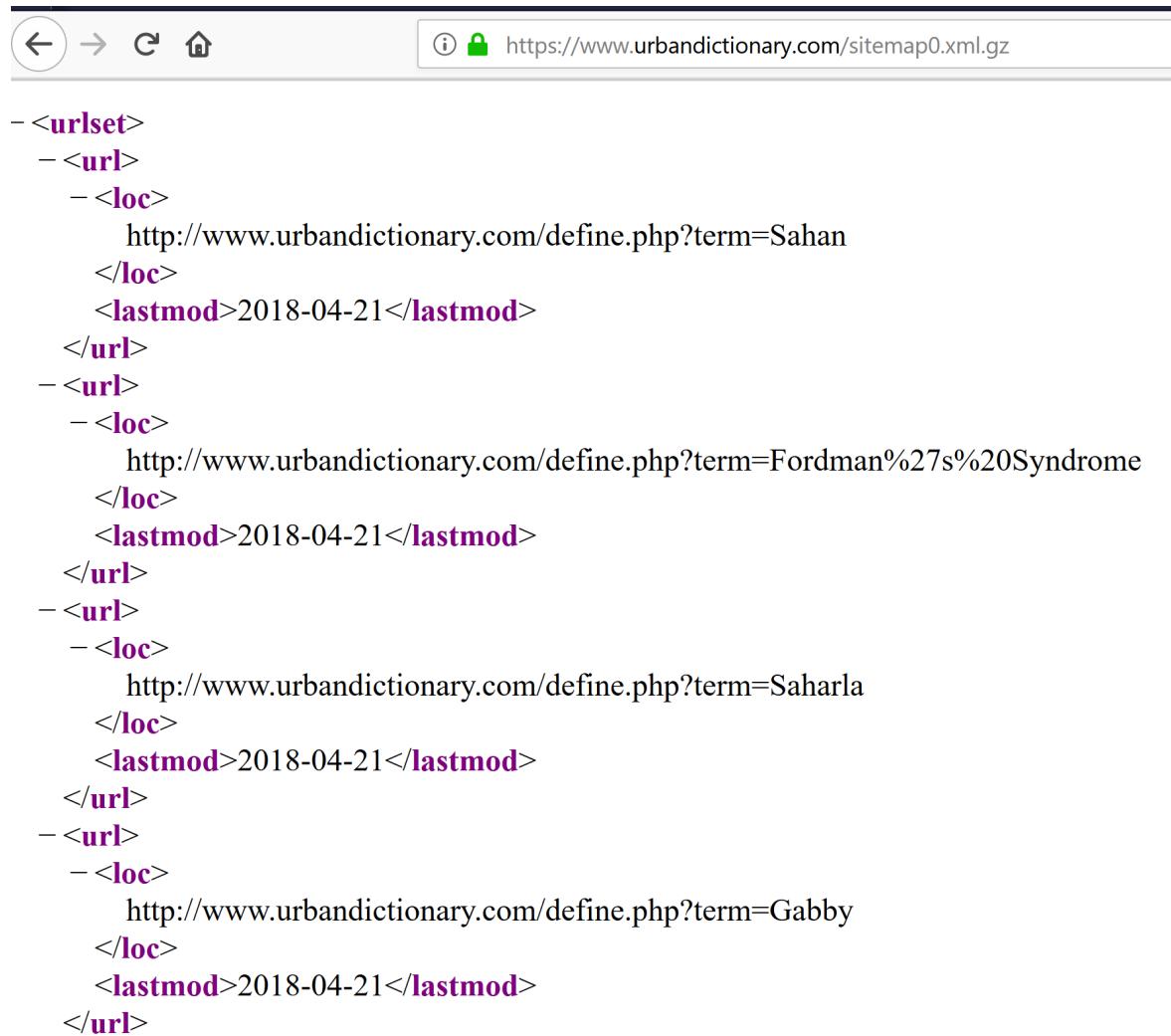


Sitemap: <http://www.urbandictionary.com/sitemap.xml.gz>
Sitemap: <https://www.urbandictionary.com/sitemap-https.xml.gz>

It sends us to an xml sitemap for further crawling.

Can We Utilize Urban Dictionary?

Looks like it's in no particular order, but it gives us a link to every single word in the urban dictionary.



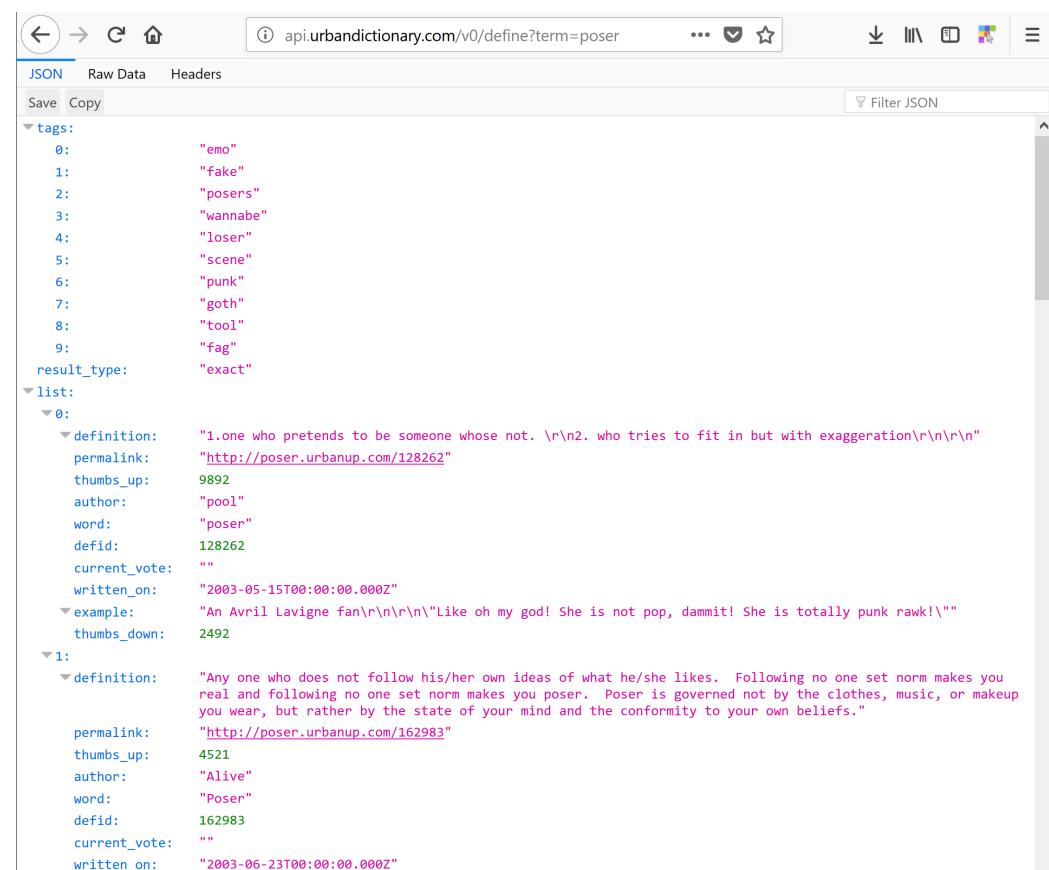
The screenshot shows a browser window with the URL <https://www.urbandictionary.com/sitemap0.xml.gz>. The page content is an XML sitemap listing several URLs for word definitions, each with its last modification date. The XML code is as follows:

```
- <urlset>
- <url>
- <loc>
    http://www.urbandictionary.com/define.php?term=Sahan
</loc>
<lastmod>2018-04-21</lastmod>
</url>
- <url>
- <loc>
    http://www.urbandictionary.com/define.php?term=Fordman%27s%20Syndrome
</loc>
<lastmod>2018-04-21</lastmod>
</url>
- <url>
- <loc>
    http://www.urbandictionary.com/define.php?term=Saharla
</loc>
<lastmod>2018-04-21</lastmod>
</url>
- <url>
- <loc>
    http://www.urbandictionary.com/define.php?term=Gabby
</loc>
<lastmod>2018-04-21</lastmod>
</url>
```

How Else Can We Access These?

You must be prepared to tackle a variety of access methods.

- Getty is available as JSON, JSONLD, or RDF.
- Wiktionary can be downloaded as XML.
- Urban Dictionary has an undocumented API.



The screenshot shows a browser window displaying the JSON response from the Urban Dictionary API. The URL in the address bar is `api.urbandictionary.com/v0/define?term=poser`. The response is a JSON object with the following structure:

```
tags: [emo, fake, posers, wannabe, loser, scene, punk, goth, tool, fag]
result_type: exact
list:
  0:
    definition: "1. one who pretends to be someone whose not. \r\n2. who tries to fit in but with exaggeration\r\n\r\n"
    permalink: "http://poser.urbanup.com/128262"
    thumbs_up: 9892
    author: "pool"
    word: "poser"
    defid: 128262
    current_vote: ""
    written_on: "2003-05-15T00:00:00.000Z"
    example: "An Avril Lavigne fan\r\n\r\n\"Like oh my god! She is not pop, dammit! She is totally punk rawk!\""
    thumbs_down: 2492
  1:
    definition: "Any one who does not follow his/her own ideas of what he/she likes. Following no one set norm makes you real and following no one set norm makes you poser. Poser is governed not by the clothes, music, or makeup you wear, but rather by the state of your mind and the conformity to your own beliefs."
    permalink: "http://poser.urbanup.com/162983"
    thumbs_up: 4521
    author: "Alive"
    word: "Poser"
    defid: 162983
    current_vote: ""
    written_on: "2003-06-23T00:00:00.000Z"
```

How Do We Access These?

Here is a call to look up the word “poser”:

- Note that this will give you tags, and then you can look up all headwords for any given tag, for example:

[https://www.urbandictionary.com/
tags.php?tag=emo](https://www.urbandictionary.com/tags.php?tag=emo)

You could then use a parser like BeautifulSoup to find all the words sharing the “emo” tag where

```
class = "word"
```

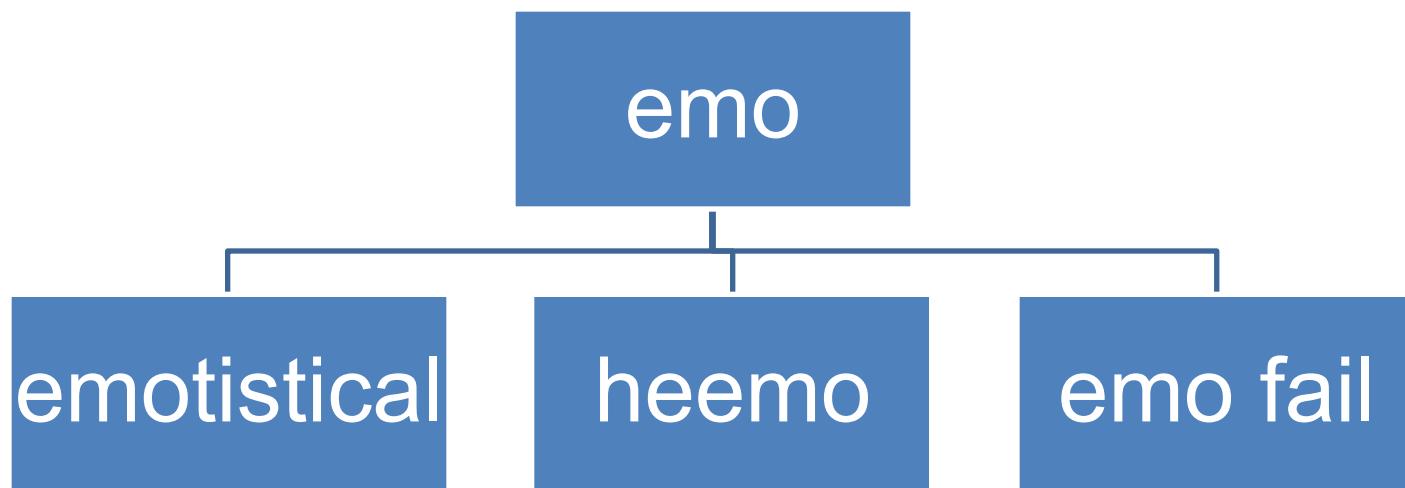
The screenshot shows a JSON viewer interface with the following details:

- Header buttons: Back, Forward, Refresh, Home, and an info icon.
- Tab selection: The "JSON" tab is selected, while "Raw Data" and "Headers" are also visible.
- Action buttons: "Save" and "Copy".
- JSON data structure:
 - A collapsed section labeled "tags:".
 - An expanded list of 9 items, indexed from 0 to 8, each consisting of a blue-numbered key and a pink-colored value.

Index	Value
0:	"emo"
1:	"fake"
2:	"posers"
3:	"wannabe"
4:	"loser"
5:	"scene"
6:	"punk"
7:	"goth"
8:	"tool"

How Could We Use This?

In this manner, you could build up families of slang terms with a common “parent” tag:



And *that's* starting to look like WordNet!

Word of Warning

I'm not a lawyer!

Be sure to consult a site's ToS ("Terms of Service") to see how you are allowed to use their site data—especially if you intend to commercially exploit it.

The screenshot shows the Urban Dictionary homepage. At the top, there is a dark header bar with the "urban DICTIONARY" logo on the left and links for "Terms of Service", "Privacy Policy", and "DMCA" on the right. Below the header is a search bar containing the placeholder text "Type any word here...". The main content area features a large, bold title "Terms of Service". Below the title, there is a paragraph of text describing the terms of service for the website. The text includes a link to the "Terms of Service" page and an email address for legal inquiries.

urban DICTIONARY

Terms of Service Privacy Policy DMCA

Type any word here...

Terms of Service

Urban Dictionary LLC (the "Company") offers UrbanDictionary.com (the "Website") according to the following terms of service. The Company reserves the right to revise these terms from time to time. We will post a notice of such revisions on this page. Your continued usage of the Website constitutes your acceptance of these terms, available at <http://www.urbandictionary.com/tos.html>. Questions about the Terms of Service may be sent to this address: legal@urbandictionary.com

Resources for Creating or Extending Lexical Knowledge Bases

Encyclopedic resources

- Baseline: Wikipedia
- Supplementary: IMDB, DotDash (formerly About.com), more!



≡ ThoughtCo. | LIFELONG LEARNING

The image shows a screenshot of the ThoughtCo. website. At the top, there is a navigation bar with the site's name and a "LIFELONG LEARNING" tag. Below the navigation, there are three main categories represented by icons and text: "SCIENCE, TECH, MATH" (with an icon of a deer and a flask), "HUMANITIES" (with an icon of a brain and a quill pen), and "ARTS, MUSIC, RECREATION" (with an icon of a smiling mask).

Resources for Creating or Extending Lexical Knowledge Bases

Encyclopedic resources



Resources for Creating or Extending Lexical Knowledge Bases

Taxonomical organizations of web pages

- Baseline: Curie (formerly DMOZ)
- Supplementary: Sitemaps (or RSS taxonomies) for sites like CNN Money, LA Times, Vogue, SFGate, etc.



Los Angeles Times

SFGATE

And Don't Forget Amazon...

Books are published on everything, and
Amazon has all the books...
...organized into a lot of categories.

Books at Amazon

Shop by Category



Arts &
Photography



Biographies &
Memoirs



Business &
Investing



Children's Books



Cookbooks, Food
& Wine



History



Literature &
Fiction



Mystery &
Suspense



Romance



Sci-Fi & Fantasy



Teens & Young
Adult

And Don't Forget Amazon...

Robots.txt has
a lot of
Disallow
statements, but
“bestsellers” is
not among
them, so...



```
User-agent: *
Disallow: /exec/obidos/account-access-login
Disallow: /exec/obidos/change-style
Disallow: /exec/obidos/flex-sign-in
Disallow: /exec/obidos/handle-buy-box
Disallow: /exec/obidos/tg/cm/member/
Disallow: /gp/aw/help/id=sss
Disallow: /gp/cart
Disallow: /gp/flex
Disallow: /gp/product/e-mail-friend
Disallow: /gp/product/product-availability
Disallow: /gp/product/rate-this-item
Disallow: /gp/sign-in
Disallow: /gp/reader
Disallow: /gp/sitbv3/reader
Disallow: /gp/richpub/syltguides/create
Disallow: /gp/gfix
Disallow: /gp/associations/wizard.html
Disallow: /gp/dmusic/order
Disallow: /gp/legacy-handle-buy-box.html
Disallow: /gp/aws/ssop
Disallow: /gp/yourstore
Disallow: /gp/gift-central/organizer/add-wishlist
Disallow: /gp/vote
Disallow: /gp/voting/
Disallow: /gp/music/wma-pop-up
```

And Don't Forget Amazon...

There are
really a lot of
bestseller
lists...

[Any Department](#)

Books

- [Arts & Photography](#)
- [Audible Audiobooks](#)
- [Biographies & Memoirs](#)
- [Books on CD](#)
- [Business & Money](#)
- [Calendars](#)
- [Children's Books](#)
- [Christian Books & Bibles](#)
- [Comics & Graphic Novels](#)
- [Computers & Technology](#)
- [Cookbooks, Food & Wine](#)
- [Crafts, Hobbies & Home](#)
- [Deals in Books](#)
- [Education & Teaching](#)
- [Engineering & Transportation](#)
- [Gay & Lesbian](#)
- [Health, Fitness & Dieting](#)
- [History](#)
- [Humor & Entertainment](#)
- [Law](#)
- [Libros en español](#)
- [Literature & Fiction](#)
- [Medical Books](#)
- [Mystery, Thriller & Suspense](#)
- [Parenting & Relationships](#)
- [Politics & Social Sciences](#)
- [Reference](#)
- [Religion & Spirituality](#)
- [Romance](#)
- [Science & Math](#)
- [Science Fiction & Fantasy](#)
- [Self-Help](#)
- [Sports & Outdoors](#)
- [Teens](#)
- [Test Preparation](#)
- [Textbooks](#)
- [Travel](#)

And Don't Forget Amazon...

There are
really a lot of
bestseller
lists...

Any Department

Books

- Arts & Photography
- Audible Audiobooks
- Biographies & Memoirs
- Books on CD
- Business & Money
- Calendars
- Children's Books
- Christian Books & Bibles
- Comics & Graphic Novels
- Computers & Technology
- Cookbooks, Food & Wine
- Crafts, Hobbies & Home
- Deals in Books
- Education & Teaching
- Engineering & Transportation

- Any Department
- Books
 - Arts & Photography
 - Audible Audiobooks
 - Biographies & Memoirs
 - Books on CD
 - Business & Money
 - Calendars
 - Children's Books
 - Christian Books & Bibles
 - Comics & Graphic Novels
 - Computers & Technology
 - Cookbooks, Food & Wine
 - Crafts, Hobbies & Home
 - Deals in Books
 - Education & Teaching
 - Engineering & Transportation
 - Gay & Lesbian
 - Health, Fitness & Dieting
 - History
 - Humor & Entertainment
 - Law
 - Libros en español
 - Literature & Fiction
 - Medical Books
 - Mystery, Thriller & Suspense
 - Parenting & Relationships
 - Politics & Social Sciences
 - Reference
 - Religion & Spirituality
 - Romance
 - Science & Math
 - Science Fiction & Fantasy
 - Self-Help
 - Sports & Outdoors
 - Teens
 - Test Preparation
 - Textbooks
 - Travel

And Don't Forget Amazon...

There are
really a lot of
bestseller
lists...

Any Department

Books

- Arts & Photography
- Audible Audiobooks
- Biographies & Memoirs
- Books on CD
- Business & Money
- Calendars
- Children's Books
- Christian Books & Bibles
- Comics & Graphic Novels
- Computers & Technology
- Cookbooks, Food & Wine
- Crafts, Hobbies & Home
- Deals in Books
- Education & Teaching
- Engineering & Transportation

Any Department

Books

- Arts & Photography
- Architecture
- Business of Art
- Collections, Catalogs & Exhibitions
- Decorative Arts & Design
- Drawing
- Fashion
- Graphic Design
- History & Criticism
- Individual Artists
- Music
- Other Media
- Painting
- Performing Arts
- Photography & Video
- Religious
- Sculpture
- Study & Teaching
- Vehicle Pictorials

Benefits

Amazon Best Sellers

Our most popular products based on sales. Updated hourly.

Any Department
Books
Arts & Photography
Architecture
Buildings
Criticism
Decoration & Ornament
Drafting & Presentation
Historic Preservation
History
Individual Architects & Firms
Interior Design
Landscape
Project Planning & Management
Regional
Security Design
Sustainability & Green Design
Urban & Land Use Planning
Vernacular

The Home Office
Shop office accessories
Learn more •



Ad feedback

Best Sellers in Architecture

#1	The Dot Peter H. Reynolds ★★★★★ 427 Hardcover \$10.19	#2	The Last Castle: The Epic Story of Love, Loss, and... Denise Kiernan ★★★★★ 126 Kindle Edition \$12.99	#3	2018 - 2019 Weekly & Monthly Planner: 2018... Nicole Planner ★★★★★ 3 Paperback \$8.99	#4	THE FINER THINGS: TIMELESS FURNITURE, TEXTILES,... Christine Lemieux ★★★★★ 34 Hardcover \$54.53	#5	THE BRIDGE: HOW THE ROEBLING'S CONNECTED... Peter J. Tomasi Hardcover \$18.81	#6	Styled: Secrets for Arranging Rooms, from... Emily Henderson ★★★★★ 232 Hardcover \$22.70	#7	Essential Perennials: The Complete Reference to... Ruth Rogers Clausen ★★★★★ 30 Kindle Edition \$11.20	#8	Structures: Or Why Things Don't Fall Down J. E. Gordon ★★★★★ 96 Paperback \$16.44
#9	The Last Castle Denise Kiernan ★★★★★ 126 Hardcover \$17.17	#10	The Last Castle: The Epic Story of Love, Loss, and... Denise Kiernan ★★★★★ 126 Hardcover \$17.17	#11	India Hicks: A Slice of England India Hicks ★★★★★ 3 Hardcover \$34.25	#12	BROADWAY: A HISTORY OF NEW YORK CITY IN THIRTEEN... Fran Lebowitz Hardcover \$24.20	#13	A GUIDE TO THE GOOD LIFE: THE ANCIENT ART OF... William B. Irvine Hardcover \$17.75	#14	A Pattern Language: Towns, Buildings,... Christopher Alexander ★★★★★ 211 Hardcover \$44.02	#15	RAY BOOTH: EVOCATIVE INTERIORS Ray Booth ★★★★★ 5 Hardcover \$37.00	#16	The Well-Tended Perennial Garden: The Essential... Tracy DiSabato-Aust ★★★★★ 25 Kindle Edition \$1.99

New Releases In ARCHITECTURE

#17	Rachel Ashwell: My Floral Affair: Whimsical... Rachel Ashwell ★★★★★ 17 Hardcover \$25.50	#18	Veranda Decorating Mario López-Cordero ★★★★★ 4 Hardcover \$28.12	#19	Rosa's Castle Deanna Edens ★★★★★ 27 Kindle Edition \$0.99		
#20	THE DEATH AND LIFE OF GREAT AMERICAN CITIES Jane Jacobs ★★★★★ 196 Paperback \$11.52	#21	A FIELD GUIDE TO AMERICAN HOUSES (REVISED): THE DEFINITIVE GUIDE TO RECOGNIZING AND UNDERSTANDING AMERICA'S DOMESTIC ARCHITECTURE Virginia Savage McAlester ★★★★★ 278 Paperback \$20.56	#22	HENRI SAMUEL: MASTER OF THE FRENCH INTERIOR Thomas Rainer ★★★★★ 6 Hardcover \$46.59	#23	PLANTING IN A POST-WILD WORLD: DESIGNING PLANT... Thomas Rainer ★★★★★ 70 Kindle Edition \$1.13

Most Wished For in ARCHITECTURE

#24	GARDEN REVOLUTION: HOW OUR LANDSCAPES CAN BE... Larry Weinger ★★★★★ 32 Kindle Edition \$1.20
-----	--

#25	The Last Castle Denise Kiernan ★★★★★ 126 Hardcover \$17.17	#26	India Hicks: A Slice of England India Hicks ★★★★★ 3 Hardcover \$34.25	#27	Rosa's Castle Deanna Edens ★★★★★ 27 Kindle Edition \$0.99	#28	THE DEATH AND LIFE OF GREAT AMERICAN CITIES Jane Jacobs ★★★★★ 196 Paperback \$11.52	#29	A FIELD GUIDE TO AMERICAN HOUSES (REVISED): THE DEFINITIVE GUIDE TO RECOGNIZING AND UNDERSTANDING AMERICA'S DOMESTIC ARCHITECTURE Virginia Savage McAlester ★★★★★ 278 Paperback \$20.56	#30	HENRI SAMUEL: MASTER OF THE FRENCH INTERIOR Thomas Rainer ★★★★★ 6 Hardcover \$46.59	#31	PLANTING IN A POST-WILD WORLD: DESIGNING PLANT... Thomas Rainer ★★★★★ 70 Kindle Edition \$1.13	#32	GARDEN REVOLUTION: HOW OUR LANDSCAPES CAN BE... Larry Weinger ★★★★★ 32 Kindle Edition \$1.20
-----	--	-----	---	-----	---	-----	---	-----	---	-----	---	-----	--	-----	--

Benefits of This

Now you know (after a bit of HTML parsing), for just about every subject matter category imaginable:

- Some major subcategories
- Popular book titles
- Popular author names
- Content words from those titles, and descriptions thereof

But Amazon is just an example. You can use any content network that allows you to crawl (and as always, check terms of service!).

DataScience@SMU

Applications of a Lexical Knowledge Base

Natural Language Processing

Applications of Lexical Knowledge Bases

Right away and without any higher-level syntax or semantic analysis, we can already use lexical KBs to improve a variety of applications.

Examples are the following:

- Enhance usability of search engines
- Writing evaluation and advice
- Smarter tag clouds



Enhancements to Search Engines

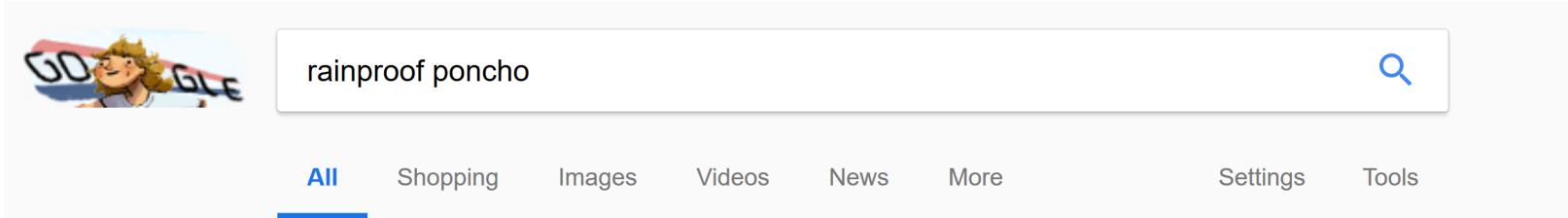
We can use lexical KBs to improve the following:

- Query expansion
- Related searches
- More like this



Query Expansion

- Offer to add synonyms and hyponyms *only* that appear in search result capsules—(pseudorelevance feedback).
- If you dare—automatically show results using expanded terms (like Google does).



A screenshot of a Google search results page. The search bar at the top contains the query "rainproof poncho". Below the search bar are several navigation links: "All" (which is underlined in blue), "Shopping", "Images", "Videos", "News", and "More". To the right of these are "Settings" and "Tools". Below the search bar, the text "About 5,860,000 results (0.48 seconds)" is displayed.

[The Best Rain Poncho: 5 Waterproof Rain Ponchos Reviewed For ...](#)

<https://carryonguy.com> › Blog › Reviews ▾

A **waterproof rain poncho** is handy for hiking, backpacking or travel but which ones truly keep you dry?
There was only 1 winner. The best rain **poncho** was...

[The Top 10 Best Rain Ponchos In 2018 | Travel Gear Zone](#)

travelgearzone.com/top-10-best-rain-ponchos/ ▾

Jump to **Mil-Tec Men's US Waterproof Ripstop Hooded Nylon Festival Poncho** - The Mil-Tec US **waterproof rain poncho** is meant to function as dexterously as a military **poncho**. Its size is an indicator of how it can serve various purposes. It covers the knees of a 6' tall person. As its long, it is wide as well.

[How to use rain ...](#) · [Which are the best rain ...](#) · [The Top 10 Best Rain ...](#)

Pseudorelevance Feedback

Simple procedure:

1. Retrieve initial search results and assume the top k results are relevant.
2. Utilize the top k results as though they were user input (marked as relevant).
3. Expand or modify initial query and/or initial results accordingly.

In our case, we usually have *too many* synonyms and hyponyms and many of them are *irrelevant*—but we can use the above procedure to narrow the list down to the probably-more-relevant ones.

Note: “patty” is a hyponym of “pie,” but we won’t use it; “currant” is a hyponym of “berry,” but we won’t use it—because there’s no support here.

Results for “berry pie recipes”

Blah blah **pie** blah
strawberry blah

Blah blah blah
raspberry tart blah

Blah blah **cobbler**
blah **blueberry**

Marionberry pie
blah blah blah

Related Searches

- If initially you have *sparse* results, replace one term at a time with a high-frequency member of a *synset* (validated with pseudorelevance feedback)
- Eliminate any zero-result queries
- Suggest in descending weighted order
 - Weightings can include number of results found, median score of top results found, etc.

***Showing results for
“waterproof poncho.”***

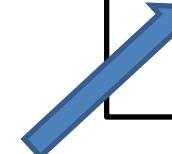
Similar searches:

[rainproof poncho](#)

[water-resistant poncho](#)

[waterproofed poncho](#)

[poncho raincoat](#)



On this one, we swapped the order of the query terms. Can you guess why?

Related Searches

- If initially you have *too many* results, replace one term at a time with a *hyponym* (validated with pseudorelevance feedback).
- Again, eliminate any zero-result queries, and suggest in descending weighted order.

*Showing results for
“dog behavior.”*

Narrow your search:

[puppy behavior](#)

[Dalmatian behavior](#)

[pug behavior](#)

[dog aggression](#)

More Like This

- Which words from the capsule are relevant?
 - Answer: Those related, in our lexical knowledge base, to the query terms
 - Can use those terms to fashion a Boolean search string

Showing results for **dog behavior**
Search instead for **dog behvaior**

Common Dog Behavior Issues | ASPCA

<https://www.aspca.org/pet-care/dog-care/common-dog-behavior-issues> ▾

Find out more here about common dog behavior issues to help you and your pup address some of our canine friends' behaviors and habits. Aggression is the most common and most serious behavior problem in dogs. Different forms of aggression require different treatments. ... Like barking, dogs howl for many reasons.

(dog | canine) & (behavior | aggression | habits)

Writing Advice

Remember the application to writing advice? Now we can give suggested word alternatives when we notice redundancy.

I like Alex because he is fun. When I want to have fun then I have Alex come over and then we have fun together.

You said “fun” three times....
Could you rewrite using words
like “play” or “sport” or “laughs”?



Of the many synonyms and hyponyms of “fun,” these were selected because they were the only ones with high frequency in the relevant corpus (same-grade-level essays).

Writing Advice

- We can also notice a lack of specificity and give suggestions.

About me: I like sports and I like dogs.

Can you say which specific “sports”? For example, rock climbing, track and field, skiing?

Can you say which specific “dogs”? For example, Chihuahua, Maltese, Irish terrier?

We decide *which words* should be more specific by noticing which nouns are *content words* that have a lot of hyponyms. The words “dogs” and “sports” both have a lot of hyponyms at multiple levels.

Filtering those hyponyms by frequency within a reference corpus yields the suggestions shown here.

Smarter Tag Clouds

We can avoid having semiredundant words in a tag cloud by checking synonymy, hypernymy, etc.



Smarter Tag Clouds

What else could you do?

- Color-code the tags according to hypernym tree:

player receiver quarterback halfback wide-out
coach trainer

agreement compliance contract

- Size the tags by hyponym-tree-depth

player receiver tight-end wide-out

Applications, and More Applications

We've just explored a few application areas.

You can brainstorm around these:

- Dating service
 - Match free-form essays prompted by “describe yourself and your interests.”
- Job-seeker service
 - Match resumes to job opening descriptions.
- Taxonomy mapping
 - Match node names from two taxonomies that are not exact word matches.

What would be the overall procedure for each of the above?

What other applications can you think of?

DataScience@SMU