

Using Wikipedia in NLP Research



AI Tutorial

下午第三、四場

陳至杰・杜海倫

清華大學資工系・自然語言研究小組

馬偉雲

中央研究院・資訊科學研究所・
詞庫小組

張俊盛

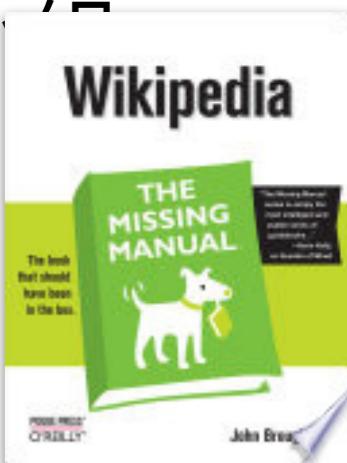
清華大學資工系・自然語言研究小組

National Taipei University (NTPU)
New Taipei, Taiwan, October 5, 2019
1:30 - 3:00, 3:20 - 1650

用維基百科做研究

開發自然語言處理系統

張俊盛 · 國立清華大學資訊工程系 · 自然語言研究小組



Wikipedia

- 2001年1月Jimmy Wales 和 Larry Sanger 創立
- 現由 Wikimedia Foundation 主導
- Alexa 統計 Wikipedia 是流量排名第九的網站
(Google 大力贊助)
- Wikipedia 有什麼不同？
 - 自由的百科全書
 - 全球志工協力創作、編輯
 - 多語言（簡英、台語、粵語、吳語都有）
 - 精確 according to Giles (2005) in *Nature*
 - 廣域、規模
 - 即時、活體
 - 半結構化
 - 詮釋資料





WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item
Cite this page

In other projects

Article Talk

Read

View source

View history

Search Wikipedia



Not logged in Talk Contributions Create account Log in

Wikipedia

From Wikipedia, the free encyclopedia



This article is about this online encyclopedia itself. For Wikipedia's home page, see [Main Page](#). For Wikipedia's visitor introduction, see [Wikipedia:About](#). For other uses, see [Wikipedia \(disambiguation\)](#).

"The Free Encyclopedia" redirects here. For other encyclopedias, see [Lists of encyclopedias](#).

Wikipedia (/wɪkɪ'piːdiə/ (audio) *wik-iH-PEE-dee-ə* or /wɪkɪ'piːdiə/ (audio) *wik-ee-PEE-dee-ə*) is a multilingual [online encyclopedia](#) created and maintained as an [open collaboration project](#)^[3] using a [wiki](#)-based editing system.^[4] It is the largest and most popular general reference work on the [World Wide Web](#),^{[5][6][7]} and is one of the most popular websites ranked by [Alexa](#) as of September 2019.^[8] It features exclusively [free content](#) and no commercial [ads](#), and is owned and supported by the [Wikimedia Foundation](#), a non-profit organization funded primarily through [donations](#).^{[9][10][11][12]}

Wikipedia was launched on January 15, 2001, by [Jimmy Wales](#) and [Larry Sanger](#).^[13] Sanger coined its name,^{[14][15]} as a [portmanteau](#) of "wiki" (the Hawaiian word for "quick"^[16]) and "encyclopedia". Initially an English-language encyclopedia, versions in other languages were quickly developed. With at least 5,941,619 articles,^[note 3] the [English Wikipedia](#) is the largest of the more than 290 Wikipedia encyclopedias. Overall, Wikipedia comprises more than 40 million articles in 301 different languages^[17] and by February 2014 it had reached 18 billion page views and nearly 500 million unique visitors per month.^[18]

In 2005, [Nature](#) published a peer review comparing 42 hard science articles from [Encyclopædia Britannica](#) and Wikipedia and found that Wikipedia's level of accuracy approached that of *Britannica*,^[19] although critics suggested that it might not have fared so well in a similar study of a [random sampling](#) of all articles or one focused on social science or contentious social issues.^{[20][21]} The following year, [Time](#) magazine stated that the open-door

Wikipedia



WIKIPEDIA

The [logo of Wikipedia](#), a globe featuring [glyphs](#) from several [writing systems](#)

Screenshot	[show]
Type of site	Online encyclopedia
Available in	303 languages
Owner	Wikimedia Foundation
Created by	Jimmy Wales Larry Sanger ^[1]
Website	www.wikipedia.org
Alexa rank	▲ 9

WIKIPEDIA

自由的百科全書

中文

1 074 000+ 條目

English

5 935 000+ articles



Español

1 546 000+ artículos

Deutsch

2 345 000+ Artikel

Français

2 141 000+ articles

Português

1 014 000+ artigos

日本語

1 169 000+ 記事

Русский

1 569 000+ статей

Italiano

1 554 000+ voci

Polski

1 360 000+ haset

party

EN ▾



文 A Read Wikipedia in your language ^

1 000 000+ 條目

Deutsch

English

Español

Français

Italiano

Nederlands

日本語

Polski

Português

Русский

Sinugboanong

Binisaya

Svenska

Tiếng Việt

Winaray

中文

100 000+ 條目

العربية

Azərbaycanca

Български

Bân-lâm-gú /

Hō-ló-oē

Беларуская

(Акадэмічнай)

Català

Čeština

Dansk

Eesti

Ελληνικά

Esperanto

Euskara

فارسی

Galego

Հայերեն

हिन्दी

Hrvatski

Bahasa

Indonesia

עברית

ქართული

Latina

Lietuvių

Magyar

Македонски

Bahasa Melayu

Bahaso

Minangkabau

Norsk (Bokmål ·

Nynorsk)

Нохчийн

O'zbekcha /

Ўзбекча

Қазақша /

Qazaqsha /

Română

Cymraeg

Simple English

Slovenčina

Slovenščina

Српски / Srpski

Srpskohrvatski /

Українська

Српскохрватски

Volapük

தமிழ்

ไทย

한국어

• 百萬俱樂部

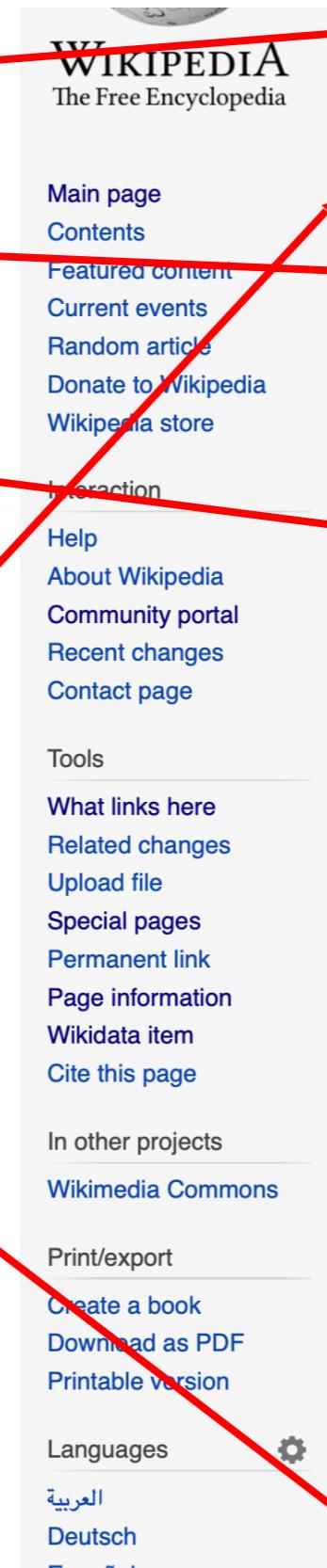
- 英文、日文
- 德文、中文
- 法文、波蘭文
- 義大利文
- 葡萄牙文
- 西班牙文
- 波蘭文
- 荷蘭文

• 十萬

- 簡單英文

Party

- 標題
- 摘要、本文
- 影像
- 資訊框
- 地理座標
- 主題分類
- 連結
 - 重新定向
 - 其他維基
 - 其他語言
 - 解歧義



Party

From Wikipedia, the free encyclopedia

For other uses, see [Party \(disambiguation\)](#).

"Pool party" redirects here. For other uses, see [Pool party \(disambiguation\)](#).

A **party** is a gathering of people who have been invited by a **host** for the purposes of **socializing**, **conversation**, **recreation**, or as part of a **festival** or other commemoration of a special occasion. A party will typically feature **food** and **beverages**, and often **music** and **dancing** or other forms of **entertainment**. In many Western countries, parties for teens and adults are associated with drinking alcohol such as **beer**, **wine**, or **distilled spirits**.

Contents [hide]

- 1 Description
- 2 Types
 - 2.1 Birthday party
 - 2.2 Surprise party
 - 2.3 Dinner party
 - 2.4 Garden party
 - 2.5 Cocktail party
 - 2.6 Tea party
 - 2.7 Reception
 - 2.8 Soirées
 - 2.9 Dances and balls
 - 2.10 Block party
 - 2.11 Costume or fancy dress party
 - 2.12 Christmas caroling party
 - 2.13 Parties for teenagers and young adults



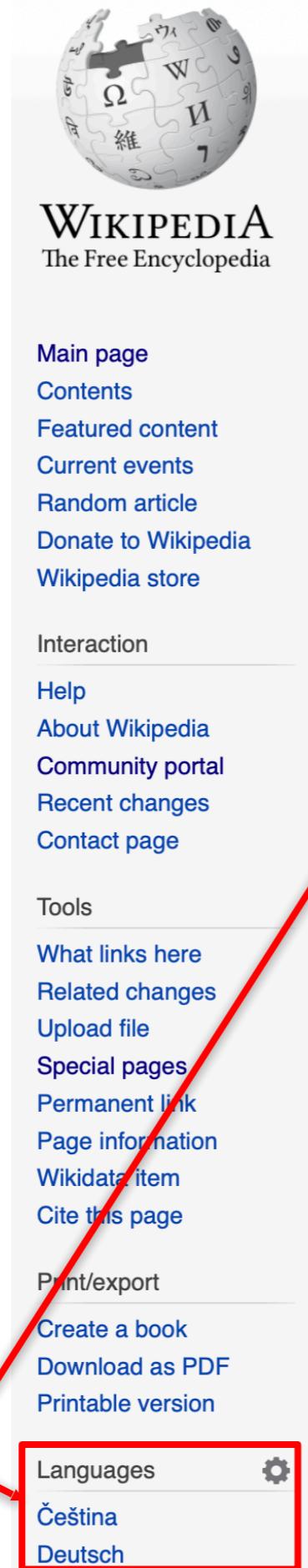
Hip, Hip, Hurrah! (1888) by Peder Severin Krøyer, a painting portraying an artists' party in 19th century Denmark



A September 1945 [street party](#) in Selborne Road, Kent, celebrating victory in the [Second World War](#)

解歧義

- 標題
- 摘要
- 影像
- 資訊框
- 地理座標
- 主題分類
- 連結
 - 重新定向
 - 其他維基
 - 其他語言
 - 解歧義



Not logged in Talk Contributions Create account Log in

Article Talk Read Edit View history Search Wikipedia

DPP

From Wikipedia, the free encyclopedia

DPP may stand for:

Business [edit]

- Digital Production Partnership, of UK public service broadcasters
- Direct Participation Program, a financial security
- Discounted payback period

Photography [edit]

- Digital Photo Professional, Canon software

Law enforcement [edit]

- Director of Public Prosecutions, responsible for criminal prosecutions

Politics [edit]

- Danish People's Party, a Danish political party
- Democratic Progressive Party, a political party of the Republic of China (Taiwan)
- Democratic Progressive Party (Malawi)
- Democratic Progressive Party (Singapore)
- Democratic Party for the People, a Japanese political party

Medicine and science [edit]

- Diabetes Prevention Program
- Decapentaplegic, a morphogen involved in development
- Diketopyrrolopyrrole dye, a class of organic dyes and pigments
- Dipeptidyl peptidases 3-10: DPP3, DPP4, DPP6, DPP7, DPP8, DPP9, DPP10
- Differential Pulse Polarography, a type of electrochemical scan

Look up [dpp](#) in Wiktionary, the free dictionary.

Contents [hide]

- 1 Business
- 2 Photography
- 3 Law enforcement
- 4 Politics
- 5 Medicine and science
- 6 Other
- 7 See also

Info Box

- 標題

- 摘要

- 影像

- 資訊框

- 地理座標

- 主題分類

- 連結

- 重新定向

- 其他維基

- 其他語言

- 解歧義



WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction

Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools

What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item
Cite this page

In other projects

Wikimedia Commons

Print/export

Create a book
Download as PDF
Printable version

Article Talk

Not logged in Talk Contributions Create account Log in

Read Edit View history

Search Wikipedia

Democratic Progressive Party

From Wikipedia, the free encyclopedia

For other uses, see [Democratic Progressive Party \(disambiguation\)](#).



This article **needs additional citations for verification**. Please help [improve this article](#) by adding citations to reliable sources. Unsourced material may be challenged and removed.
Find sources: "Democratic Progressive Party" – news · newspapers · books · scholar · JSTOR (December 2010) (Learn how and when to remove this template message)

The Democratic Progressive Party (DPP) is a Taiwanese nationalist and centre-left political party in Taiwan.^{[5][6][7]} Controlling both the Taiwan Presidency and the unicameral Legislative Yuan, it is the majority ruling party and the dominant party in the Pan-Green Coalition.

Founded in 1986, the DPP is one of two major parties in Taiwan, along with the historically dominant Kuomintang. It has traditionally been associated with strong advocacy of human rights, anti-communism and a distinct Taiwanese identity. The incumbent President and former leader of the DPP, Tsai Ing-wen, is the second member of the DPP to hold the office.^[8]

The DPP is a long-term member of Liberal International and a founding member of the Council of Asian Liberals and Democrats. It represented Taiwan in the Unrepresented Nations and Peoples Organisation. The DPP and its affiliated parties are widely classified as socially liberal because of their strong support for human rights, including support for

Democratic Progressive Party
民主進步黨
Mínzhǔ Jinbù Dǎng



Abbreviation	民進黨
Chairman	Cho Jung-tai
Secretary-General	Luo Wen-jia
Founded	28 September 1986; 33 years ago
Headquarters	Taipei, Taiwan
Think tank	New Frontier Foundation
Membership (2014)	▲ 335,643
Ideology	Liberalism ^{[1][2]} Populism ^[2]

Hyperlinks (WSD, Wikification, IE)

[[link (sense) | anchor (word)]]

Party (law)

- "'party'" bf synonyms
 - 有時候不只一個
 - A **shill**, also called a **plant** or a **stooge**, is a person who ...
- [[legal personality|entity]]
 - <link> = 'legal personality'
 - <anchor> = 'entity'
- [[plaintiff]]
 - <link> = 'plaintiff'
 - <anchor> = 'plaintiff'

From Wikipedia, the free encyclopedia

A **party** is a **person** or group of persons that compose a single **entity** which can be identified as one for the purposes of the **law**. Parties include: **plaintiff** (person filing suit), **defendant** (person sued or charged with a crime), **petitioner** (files a petition asking for a court ruling), **respondent** (usually in opposition to a petition or an appeal), cross-complainant (a defendant who sues someone else in the same lawsuit), or cross-defendant (a person sued by a cross-complainant).^[1] A person who only appears in the case as a **witness** is not considered a party.

Article [Talk](#) [Read](#) Edit source [View history](#) Search Wikipedia 

A ''''party''' is a [[person]] or group of persons that compose a single [[legal personality|entity]] which can be identified as one for the purposes of the [[law]]. Parties include:

[[plaintiff]] (person filing suit), [[defendant]] (person sued or charged with a crime), [[petitioner]] (files a petition asking for a court ruling), [[respondent]] (usually in opposition to a petition or an appeal), cross-complainant (a defendant who sues someone else in the same lawsuit), or cross-defendant (a person sued by a cross-complainant).<ref>{{cite web|url=http://legal-dictionary.thefreedictionary.com/party|title=Party - legal definition}}</ref> A person who only appears in the case as a [[witness]] is not considered a party.

History of revision

- View history and edits
- Before/after: It has been noted that [-**in**-] {+**for**+} the past few decades many naturist and nudist clubs have a noticeable lack of young members in their ranks.

Article Talk Read **Edit** **View history** Search Wikipedia

Issues in social nudity: Difference between revisions

From Wikipedia, the free encyclopedia

Browse history interactively

Revision as of 04:16, 14 October 2014 (edit)
Regulov (talk | contribs)
(→Age: Again....)
← Previous edit

Revision as of 22:43, 17 October 2014 (edit) (undo)
DNBR (talk | contribs)
(→Diversity)
Next edit →

Line 72:

==Age==

[[File:World Naked Bike Ride - Zaragoza.jpg|thumb|250px|right|Participants in the [[World Naked Bike Ride]] in Zaragoza, Spain.]]

It has been noted that **in** the past few decades many naturist and nudist clubs have a noticeable lack of young members in their ranks.<ref>{{cite web | last = Daney | first = Charles | title = Why Don't More Young Adults Try Social

Line 72:

==Age==

[[File:World Naked Bike Ride - Zaragoza.jpg|thumb|250px|right|Participants in the [[World Naked Bike Ride]] in Zaragoza, Spain.]]

It has been noted that **for** the past few decades many naturist and nudist clubs have a noticeable lack of young members in their ranks, **though previously this was not the case**.<ref>{{cite web | last = Daney | first = Charles |

History of revision

Line 58:

Line 96:

==Diversity==

====Body and hair acceptance====

==Controversial activities==

- ===Beauty pageants==

Generally, nudist groups accept people of all [[Race (classification of humans)|races]], ages, sizes, and shapes, including those with [[amputation]]s, [[scar]]s, [[tattoo]]s or trimmed (or absence of) [[pubic hair]]. Recently some "smoothie"<ref><http://www.euronaturist.com/smooth.htm> Euro Naturist - smooth naturists & nudists - Smoothies<!-- Bot generated title --></ref> organizations have expressed a preference for members to completely shave or remove all pubic hair, as some profess to believe that having pubic hair does not leave that body completely nude although head and **perhaps** [[facial hair]] are judged acceptable.

Line 58:

Line 96:

==Diversity==

====Body and hair acceptance====

==Controversial activities==

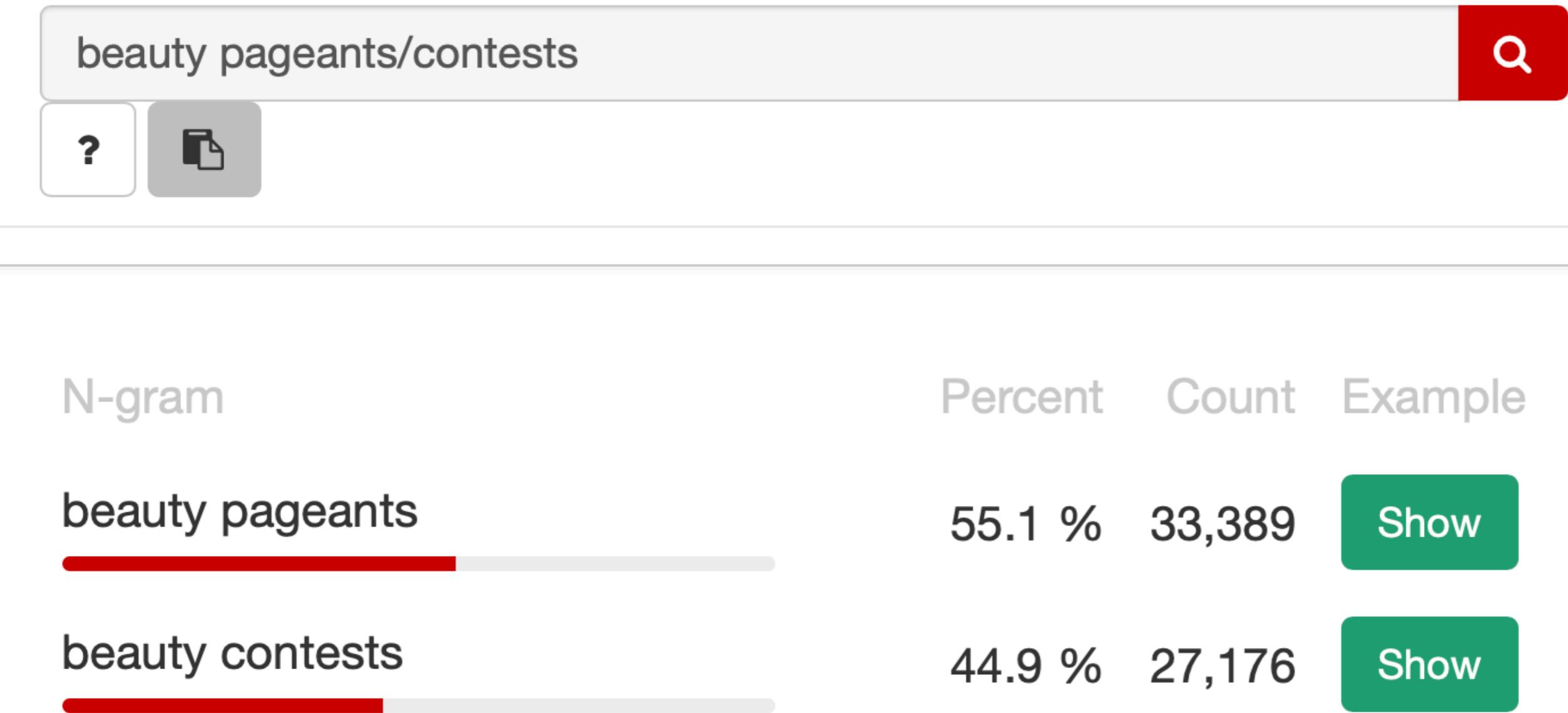
+ ===Beauty contests==

Generally, nudist groups accept people of all [[Race (classification of humans)|races]], ages, sizes, and shapes, including those with [[amputation]]s, [[scar]]s, [[tattoo]]s or trimmed (or absence of) [[pubic hair]]. Recently some "smoothie"<ref><http://www.euronaturist.com/smooth.htm> Euro Naturist - smooth naturists & nudists - Smoothies<!-- Bot generated title --></ref> organizations have expressed a preference for members to completely shave or remove all pubic hair, as some profess to believe that having pubic hair does not leave that body completely nude, although head and **possibly** [[facial hair]] are judged acceptable if not too long. Official mandates are rare, however.

Change pageants to contests, really?

Ask Linggle (<http://linggle.com>)

linggle¹⁰¹²



Change pageants to contests, really?

Ask Linggle (<http://linggle.com>)

linggle¹²¹⁰

? File Search

N-gram	Percent	Count	Example
Beauty contests	51.6 %	2,071	Show
Beauty pageants	48.4 %	1,945	Show

Left Margin Links and The Missing Manual

WIKIPEDIA
The Free Encyclopedia

[Main page](#)
[Contents](#)
[Featured content](#)
[Current events](#)
[Random article](#)
[Donate to Wikipedia](#)
[Wikipedia store](#)

[Interaction](#)

[Help](#)
[About Wikipedia](#)
[Community portal](#)
[Recent changes](#)
[Contact page](#)

Tools

[What links here](#)
[Related changes](#)
[Upload file](#)
[Special pages](#)
[Permanent link](#)
[Page information](#)
[Wikidata item](#)
[Cite this page](#)

Print/export

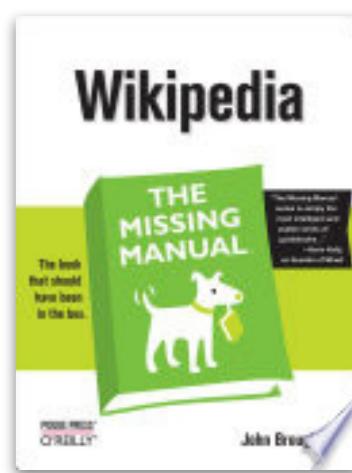
[Create a book](#)
[Download as PDF](#)
[Printable version](#)

Languages 

[Čeština](#)
[Deutsch](#)
[Español](#)

- What links here
 - links from outside pages
- Special pages
 - For management purpose
- Page information
 - Meta data: id, edits, hits, Wikidata item
 - Structured info
- Cite this page
 - Give credit to Wikipedia contributors

[pdfs.semanticscholar.org/
7e13/836478cb971f2fca178
5d85b313f96195012.pdf](https://pdfs.semanticscholar.org/7e13/836478cb971f2fca1785d85b313f96195012.pdf)



Tools What links here (201 in-links vs. 17 out-links)

Tools

[What links here](#)

[Related changes](#)

[Upload file](#)

[Special pages](#)

[Permanent link](#)

[Page information](#)

[Wikidata item](#)

[Cite this page](#)

Print/export

[Create a book](#)

[Download as PDF](#)

[Printable version](#)

Languages



[Čeština](#)

[Deutsch](#)

[Español](#)

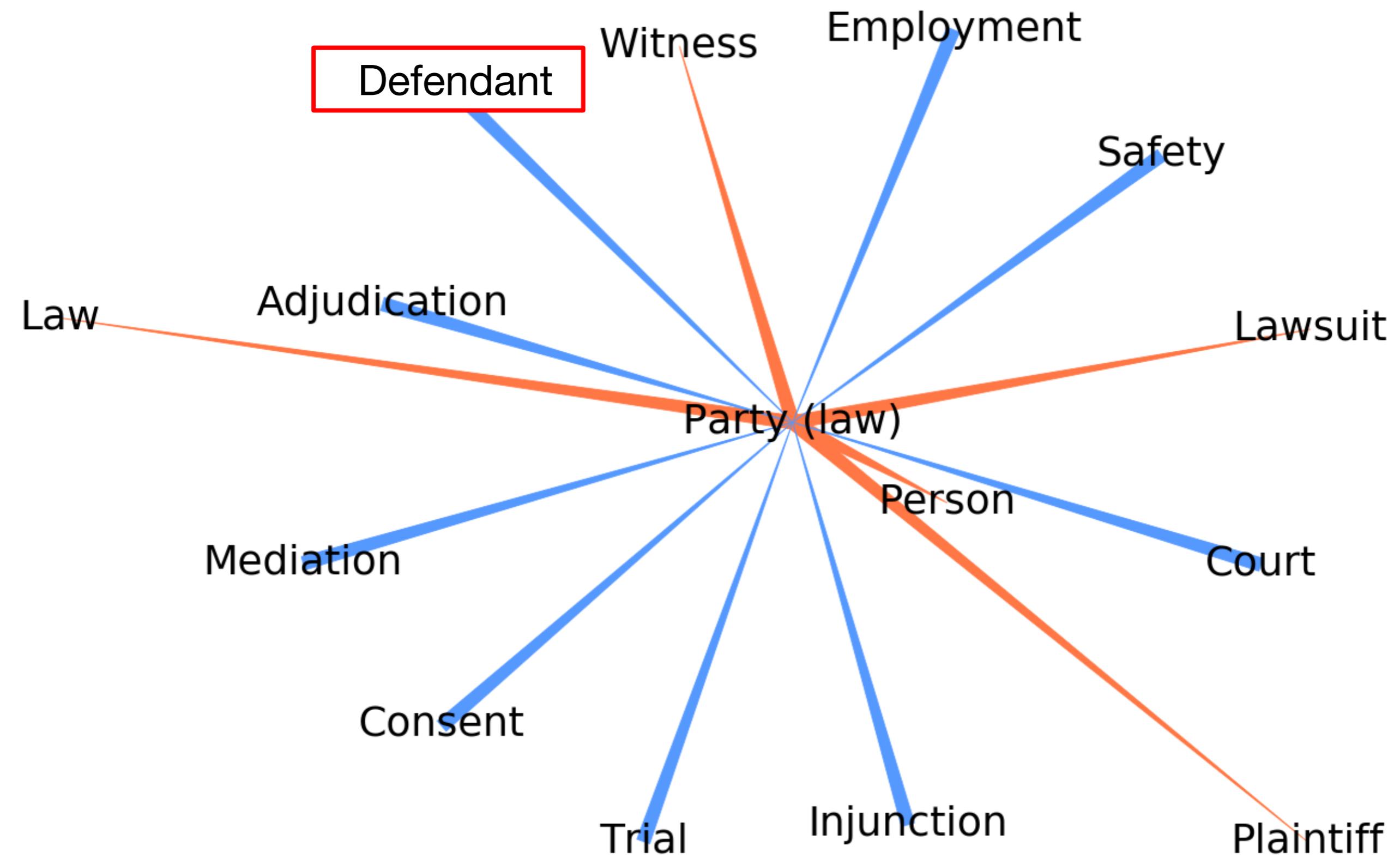
The following pages link to **Party (law)**

External tools: [Show redirects only](#) (grouped by anchor)

[View](#) (previous 50 | [next 50](#)) ([20](#) | [50](#) | [100](#) | [250](#) | [500](#))

- [Answer \(law\)](#) ([links](#) | [edit](#))
- [Class action](#) ([links](#) | [edit](#))
- [Complaint](#) ([links](#) | [edit](#))
- [Deed poll](#) ([links](#) | [edit](#))
- [Defendant](#) ([links](#) | [edit](#))
- [Dispute resolution](#) ([links](#) | [edit](#))
- [Interrogatories](#) ([links](#) | [edit](#))
- [International Court of Justice](#) ([links](#) | [edit](#))
- [Mediation](#) ([links](#) | [edit](#))
- [Non-disclosure agreement](#) ([links](#) | [edit](#))
- [Prima facie](#) ([links](#) | [edit](#))

Wikipedia Link Graph



Tools What links here

Tools

[What links here](#)

[Related changes](#)

[Upload file](#)

[Special pages](#)

[Permanent link](#)

[Page information](#)

[Wikidata item](#)

[Cite this page](#)

Print/export

[Create a book](#)

[Download as PDF](#)

[Printable version](#)

Languages



[Čeština](#)

[Deutsch](#)

[Español](#)

The following pages link to **Party (law)**

External tools: [Show redirects only](#) (grouped by anchor)

[View](#) (previous 50 | [next 50](#)) ([20](#) | [50](#) | [100](#) | [250](#) | [500](#))

- [Answer \(law\)](#) ([links](#) | [edit](#))
- [Class action](#) ([links](#) | [edit](#))
- [Complaint](#) ([links](#) | [edit](#))
- [Deed poll](#) ([links](#) | [edit](#))
- [Defendant](#) ([links](#) | [edit](#))
- [Dispute resolution](#) ([links](#) | [edit](#))
- [Interrogatories](#) ([links](#) | [edit](#))
- [International Court of Justice](#) ([links](#) | [edit](#))
- [Mediation](#) ([links](#) | [edit](#))
- [Non-disclosure agreement](#) ([links](#) | [edit](#))
- [Prima facie](#) ([links](#) | [edit](#))

Tools What links here

Defendant

Tools

[What links here](#)

[Related changes](#)

[Upload file](#)

[Special pages](#)

[Permanent link](#)

[Page information](#)

[Wikidata item](#)

[Cite this page](#)

[Print/export](#)

[Create a book](#)

[Download as PDF](#)

[Printable version](#)

Languages



[Čeština](#)

[Deutsch](#)

[Español](#)

From Wikipedia, the free encyclopedia

A **defendant** is a [person](#) accused of committing a [crime](#) in [criminal prosecution](#) or a person against whom some type of civil relief is being sought in a [civil case](#).

Terminology varies from one [jurisdiction](#) to another. For example, [Scots law](#) does not use the term "defendant"; the terms "accused" or "panel" are used instead in criminal proceedings, and "defender" in civil proceedings.^[1]

Defendants must be present at every stage of the proceedings against them. (There is an exception for very serious crimes.)

A **party** is a person or group of persons that compose a single entity which can be identified as one for the purposes of the law. Parties include: plaintiff, defendant, petitioner, respondent, cross-complainant, or cross-defendant. A person who only appears in the case as a witness is not considered a party.

If more than one person is accused of committing a crime, they are referred to as "co-conspiritor" in English.

In some jurisdictions, the term "plaintiff" is used to describe the party bringing the suit.

[Non-Registered Inten](#)

Civil defendant

In a civil [lawsuit](#), a defendant (or a **respondent**) is also the accused party, although not

TextNet vs. ImageNet

- The “images” of **Party (law)** are sentences from “what links here”
 - *In a civil law suit, the defendant is also a accused **party**, ...*
 - WSD: word = **party**, sense = **Party (law)** or **part.n#5** (WordNet)

Defendants must be present at every stage of the proceedings against them. (There is an exception for violent crimes.)

If more than one person is involved in a crime, they are "co-conspiritor" in English.

In some jurisdictions, there is a category of "Non-Registered Intervenors".

A **party** is a person or group of persons that compose a single entity which can be identified as one for the purposes of the law. Parties include: plaintiff, defendant, petitioner, respondent, cross-complainant, or cross-defendant. A person who only appears in the case as a witness is not considered a party.

Civil defendants

In a civil **lawsuit**, a defendant (or a **respondent**) is also the accused **party**, although not

TextNet: A Large-Scale Sense-Tagged Corpus



<http://textnet.nlplab.cc>
{helen.tu, patina, jason}@nlplab.cc

Hai-Lun Tu, Pei-Chen Ho, Jason S. Chang and Pin-Yuan Chen

TextNet

Sense-Tagged Dateset form Wikipedia

1. Develop Wiki sense inventory ([WikiSI](#))
 - Using hyperlink references in Wikipedia: [[political party|party]], [[party]], [[party (law)|party]]
2. Disambiguate and align Wikipedia hyperlinked articles to WordNet senses
 - [[political party|party]] >= party.n.01
 - [[party]] >= party.n.04
 - [[party (law)|party]] >= party.n.05
3. Extract Wikipedia sentences with anchor | Wiki page | WordNet sense
 - party | [Political_party](#) | party.n.01 (865)
 - party | [Party](#) | party.n.04 (320)
 - party | [Party \(law\)](#) | party.n.05 (123)

♦ Example (see Figure in the right)

- In the [party](#) primary elections and caucuses in each U.S. state,

TextNet ?

Display Options: Select Change Powered by WordNet 3.0, Wiki0601

Party, Political Party (an organization to gain political power)

- Currently, the [party](#) breakdown is as follows: Sula is noted for strong traditions in the field of music. Nils Petter Molvær grew up in Langevåg.
- The first Maoist party, which took the name of the old Communist [Party](#) of Switzerland (CPS), originated in Switzerland in 1963 but later disintegrated.
- Both the members of the cadre [party](#) and sympathizers were expected to make a large time commitment (newspaper sales before factory gates, ideological training) as well as meeting the financial expenses involved in party work.

Party (a group of people gathered together for pleasure)

Party, Company (a band of people associated temporarily in some activity)

Party (an occasion where people can assemble for social interaction and entertainment)

Party (a person involved in legal proceedings)

Comparison with pledged delegates [edit]

Democratic Party rules distinguish pledged and unpledged delegates. Pledged delegates are selected based on their announced preferences in the contest for the presidential nomination. In the [party](#) primary elections and [caucuses](#) in each U.S. state, voters [express the States](#). Pledged delegates are legally required to support the state (Rule 12.J): "Delegates must implement the party's agenda. They fall into three categories: unpledged delegates, and pledged PLECs (Professional Electoral College). At-large pledged delegates are elected by the states' requirements, the party rules require that the presidential candidate shall in all good conscience reflect the sentiments of the state's population."

Alignment Guidelines

- ♦ Align a WN sense with Wiki page if they are:

WordNet lemma is identical to page Identical

bank.n.01 (*WordNet*): sloping land (especially the slope beside a body of water)

Bank (geography) (*Wikipedia*)

Else if a sister term is identical to page title

disk.n.01: something with a round shape resembling a flat circular plate

Ball

Else a daughter term is identical to page title

party.n.05: a person involved in legal proceedings

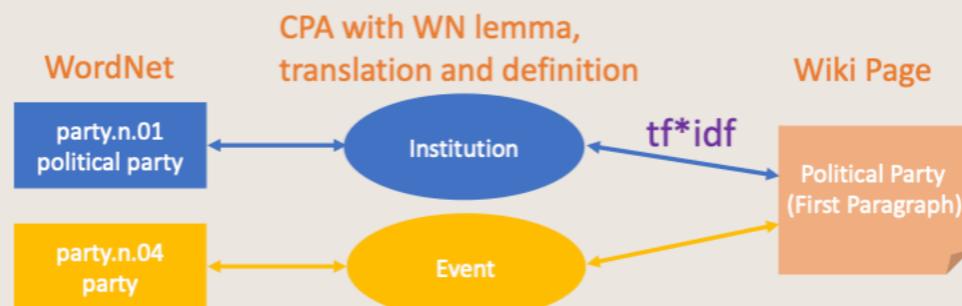
Plaintiff

- ♦ Exclude

- Name entity and modifier
- Nazi **Party** and **Arms** industry

The Methods of Alignment

- ♦ **tf*idf** with CPA
 - The 253 CPA semantic types are mapped to the respective WordNet concepts



Evaluation

- ♦ Development set: 30 words in 158 senses and about 5,000 sentences
- ♦ Full set

	Unique Words	Synsets	Word-Sense Pairs	Ambig. words	(# of Sentences)
Wiki (2019/7/2)	41,615	-	(En) 88,571 (Zh) 22,781	12,159	3,741,348
WordNet	117,798	82,115	146,312	15,935	48,247



Tools Page information

Tools

[What links here](#)
[Related changes](#)
[Upload file](#)
[Special pages](#)
[Permanent link](#)
[Page information](#)
[Wikidata item](#)
[Cite this page](#)

[Print/export](#)

[Create a book](#)

[Download as PDF](#)

[Printable version](#)

Languages



[Čeština](#)

[Deutsch](#)

[Español](#)

Basic information

Display title	Party (law)
Default sort key	Party (Law)
Page length (in bytes)	1,495
Page ID	601627
Page content language	en - English
Page content model	wikitext
Indexing by robots	Allowed
Number of page watchers	59
Number of page watchers who visited recent edits	9
Number of redirects to this page	2
Counted as a content page	Yes
Wikidata item ID	Q1686000
Central description	person or group composing a single entity for the purposes of the law
Page views in the past 30 days	3,557

Tools Page information

Tools

[What links here](#)
[Related changes](#)
[Upload file](#)
[Special pages](#)
[Permanent link](#)
[Page information](#)
[Wikidata item](#)
[Cite this page](#)

[Print/export](#)

[Create a book](#)

[Download as PDF](#)

[Printable version](#)

Languages



[Čeština](#)

[Deutsch](#)

[Español](#)

Edit history

Page creator	Savantpol (talk contribs)
Date of page creation	04:41, 17 April 2004
Latest editor	121.211.9.113 (talk)
Date of latest edit	06:25, 11 July 2019
Total number of edits	62
Recent number of edits (within past 30 days)	0
Recent number of distinct authors	0

Page properties

Hidden categories (8)	<ul style="list-style-type: none">Category:All articles needing additional referencesCategory:All articles that may contain original researchCategory:All stub articlesCategory:Articles needing additional categories from November 2018Category:Articles needing additional references from November 2018Category:Articles that may contain original research from November 2018Category:Articles with multiple maintenance issuesCategory:Flagged dictionary definitions
Transcluded templates (45)	<ul style="list-style-type: none">Party (law) (edit)Template:Ambox (view source) (template protected)Template:Asbox (view source) (template protected)Template:Cite web (view source) (protected)



WIKIPEDIA

The Free Encyclopedia

Main page

Contents

Featured content

Current events

Random article

Donate to Wikipedia

Wikipedia store

Interaction

Help

About Wikipedia

Community portal

Recent changes

Contact page

Tools

What links here

Related changes

Upload file

Article Talk

Not logged in Talk Contributions Create account Log in

Read Edit View history

Search Wikipedia



Wikidata

From Wikipedia, the free encyclopedia

For Wikipedia's information page on Wikidata, see [Wikipedia:Wikidata](#).

Wikidata is a [collaboratively edited knowledge base](#) hosted by the [Wikimedia Foundation](#). It is a common source of [open data](#) that Wikimedia projects such as [Wikipedia](#) can use,^{[5][6]} and anyone else, under a [public domain license](#).

This is similar to the way [Wikimedia Commons](#) provides storage for media files and access to those files for all Wikimedia projects, and which are also freely available for reuse. Wikidata is powered by the software [Wikibase](#).^[7]

Contents [hide]

1 Concepts

1.1 Items

1.2 Statements

1.3 Lexemes

Wikidata
the free knowledge base that anyone can edit

Wikidata is a free knowledge base that can be read and edited by humans and machines alike. It is for data what Wikimedia Commons is for media files: it centralizes access and management of structured data, such as interwiki references and statistical information. Wikidata contains data in all languages for which there are Wikimedia projects.

You can read the [introduction](#) to learn more about Wikidata.

Project page Discussion Read View source View history English Create account Log in Statistics

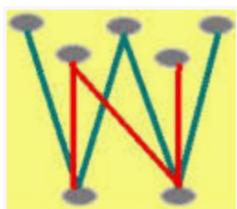
Navigation Main page Community portal Project chat Recent changes Random page Help Donate Toolbox What links here Related changes Special pages Printable version



Wikidata

- 2012年 Allen Institute for AI, Moore Foundation, Google 捐助 1百30萬歐元創辦 (the new FreeBase?)
- 人工編輯的固定格式知識庫
 - 描述、語言連結、陳述
 - predicate
 - e.g., (Douglas_Adams, educated_at, St._John_College)
 - ontological relation
 - e.g., (party (Q100), subclass-of, **person**)
 - e.g., (party (Q100), instance-of, **legal-concept**)
 - e.g., (party (Q200), subclass-of, **event**)
 - These align well with WordNet **supersense** and **domain**

Aligning WordNet with Wikipedia



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: <lexical filename> (gloss) "an example sentence"

Display options for word: word#sense number

Noun

- <noun.group> S: (n) **party#1, political party#1** (an organization to gain political power) *"in 1992 Perot tried to organize a third party at the national level"*
- <noun.group> S: (n) **party#2** (a group of people gathered together for pleasure) *"she joined the party after dinner"*
- <noun.group> S: (n) **party#3, company#7** (a band of people associated temporarily in some activity) *"they organized a party to search for food"; "the company of cooks walked into the kitchen"*
- <noun.event> S: (n) **party#4** (an occasion on which people can assemble for social interaction and entertainment) *"he planned a party to celebrate Bastille Day"*
- <noun.person> S: (n) **party#5** (a person involved in legal proceedings) *"the party of the first part"*

political party (Q7278)

organization

?

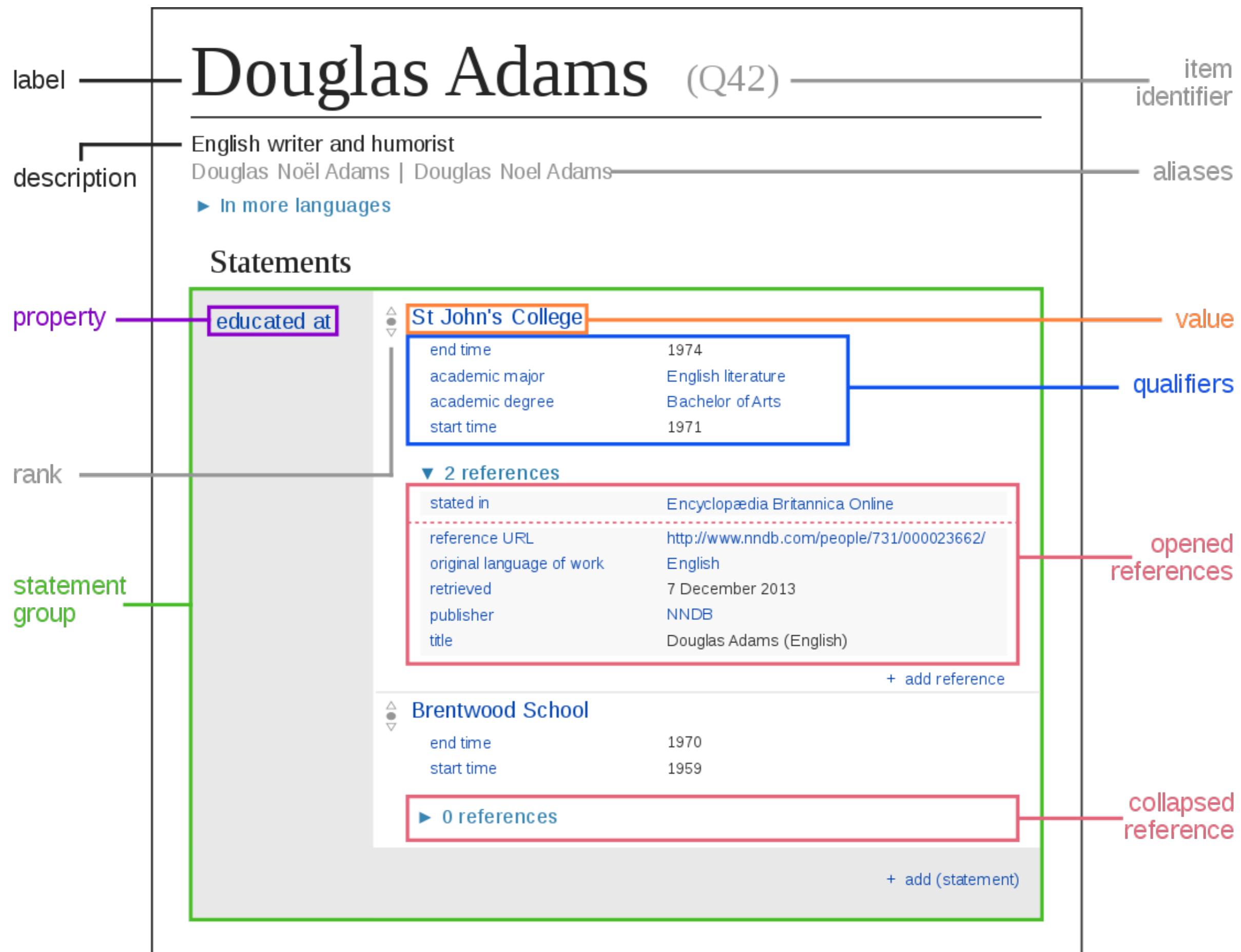
?

party (Q200538)

event

party (Q1686000)

person



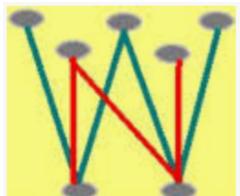
Tools Wikidata item



Tools
[What links here](#)
[Related changes](#)
[Upload file](#)
[Special pages](#)
[Permanent link](#)
[Page information](#)
[Wikidata item](#)
[Cite this page](#)

Print/export
[Create a book](#)
[Download as PDF](#)
[Printable version](#)

Languages [Čeština](#) [Deutsch](#) [Español](#)



party (Q1686000)

person or group composing a single entity for the purposes of the law

edit

▼ In more languages Configure

Language	Label	Description	Also known as
English	party	person or group composing a single entity for the purposes of the law	
Chinese	No label defined	No description defined	
Traditional Chinese	No label defined	No description defined	
Sediq Taroko	No label defined	No description defined	

Statements

instance of

legal concept

edit

▼ 0 references

+ add reference

+ add value

subclass of

person

edit

- <noun.person> [S: \(n\) party#5](#) (a person involved in legal proceedings) "the party of the first part"

Tools Wikidata item



party (Q200538)

gathering of invited guests

edit

In more languages Configure

Language	Label	Description	Also known as
English	party	gathering of invited guests	
Chinese	派對	No description defined	派对 聚會
Traditional Chinese	聚會	No description defined	
Sediq Taroko	No label defined	No description defined	

All entered languages

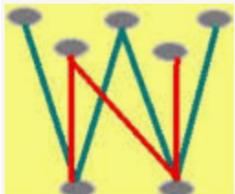
Statements

subclass of

event

edit

- <noun.event> S: (n) party#4 (an occasion on which people can assemble for social interaction and entertainment) "he planned a party to celebrate Bastille Day"



Tools

[What links here](#)
[Related changes](#)
[Upload file](#)
[Special pages](#)
[Permanent link](#)
[Page information](#)
[Wikidata item](#)
[Cite this page](#)

[Print/export](#)

[Create a book](#)
[Download as PDF](#)
[Printable version](#)

Languages

[Čeština](#)
[Deutsch](#)
[Español](#)

Tools Wikidata item



political party (Q7278)



organization that seeks to influence government policy and actions

▼ In more languages Configure

Language	Label	Description	Also known as
English	political party	organization that seeks to influence government policy and actions	
Chinese	政党	以影响政府政策和运作为目标的组织	
Traditional Chinese	政黨	以影響政府政策和運作為目標的組織	
Sediq Taroko	No label defined	No description defined	

All entered languages

Tools

- [What links here](#)
- [Related changes](#)
- [Upload file](#)
- [Special pages](#)
- [Permanent link](#)
- [Page information](#)
- [Wikidata item](#)
- [Cite this page](#)

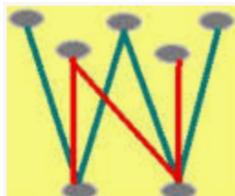
Print/export

- [Create a book](#)
- [Download as PDF](#)
- [Printable version](#)

Languages



- [Čeština](#)
- [Deutsch](#)
- [Español](#)



Statements

subclass of

- non-governmental organization
- political organisation
- juridical person ?

- <noun.group> **S: (n) party#1, political party#1** (an organization to gain political power) "in 1992 Perot tried to organize a third party at the national level"

Languages 中文

Tools

[What links here](#)

[Related changes](#)

[Upload file](#)

[Special pages](#)

[Permanent link](#)

[Page information](#)

[Wikidata item](#)

[Cite this page](#)

[In other projects](#)

[Wikimedia Commons](#)

[Print/export](#)

[Create a book](#)

[Download as PDF](#)

[Printable version](#)

[Languages](#)



[العربية](#)

[Deutsch](#)

[Español](#)

[Français](#)

[हिन्दी](#)

[Bahasa Indonesia](#)

[Português](#)

[Русский](#)

[中文](#)

[44 more](#)

Party



From Wikipedia, the free encyclopedia

For other uses, see [Party \(disambiguation\)](#).

"Pool party" redirects here. For other uses, see [Pool party \(disambiguation\)](#).

A **party** is a gathering of people who have been invited by a **host** for the purposes of socializing, conversation, recreation, or as part of a **festival** or other commemoration of a special occasion. A party will typically feature **food** and **beverages**, and often **music** and **dancing** or other forms of **entertainment**. In many Western countries, parties for teens and adults are associated with drinking alcohol such as **beer**, **wine**, or **distilled spirits**.



Hip, Hip, Hurrah! (1888) by Peder Severin Krøyer, a painting portraying an artists' party in 19th century Denmark



A September 1945 [street party](#) in Selborne Road, Kent, celebrating victory in the [Second World War](#)

派對

[[編輯](#)]

維基百科，自由的百科全書

派對源自英文**party**，又譯趴踢、或稱聯歡會，通常是社會中的人們聚在一起打算主要用於慶祝和休閒的一種方式。有一些派對是因為**宗教**和**季節性**的原因，有些派對則是由個人舉辦的。派對舉辦的場地並無地點限制，但通常必須是安全並且令人愉快的。

人們派對可以放鬆自己的身心，和同伴一起盡情的玩耍，快樂就是人們在派對中的感受。同時派對為各種各樣的人在社會的生活提供了許多機會。人們在派對上可以結交不同的朋友，通常這些朋友都有著不同的**文化**，通過這種方式你可以了解更多的來自不同地區的人的文化和生活習俗。



較為正式禮儀的派對



霹靂舞派對

External Tools – The Missing Manual



- <http://tools.wikimedia.de/~daniel/WikiSense/CategoryIntersect.php>
- <https://petscan.wmflabs.org>

PetScan Manual Issues [English](#)  Examples List

This is a new version of PetScan. The previous version is here.

Categories Page properties Templates&links Other sources Wikidata Output

Language en Commons Wikispecies Wikidata

Project wikipedia

Depth 0 (depth 0 means no subcategories; appending '|' and a number to a category in the fields below will override it for that category)

Categories

Combination Subset Union

Negative categories

Wikipedia in NLP Research

學術搜尋約有 37,700 項結果

- Wikipedia category graph
- WikiRelate!
- **Information Extraction**
- Word Sense Disambiguation
- Wikipedia taxonomy / thesaurus
- Wikipedia categories
 - decode, simplify ([Yago3](#))
- Grammatical error correction
- Word sense inventory (WSI)
- Word sense disambiguation (WSD)
- Ontology alignment (Wiki↔WordNet)

([BabelNet](#), [TextNet](#))

- Document classification
- Automated articles/biography writing

nlp wikipedia

學術搜尋 約有 37,700 項結果 (0.08 秒) 年份 ▾

[PDF] [Analysis of the Wikipedia category graph for NLP applications](#) [PDF] [aclweb.org](#)
[T Zesch, I Gurevych](#) - Proceedings of the Second Workshop on ..., 2007 - aclweb.org
In this paper, we discuss two graphs in Wikipedia (i) the article graph, and (ii) the category graph. We perform a graphtheoretic analysis of the category graph, and show that it is a scale-free, small world graph like other well-known lexical semantic networks. We ...
☆ ↗ 被引用 184 次 相關文章 全部共 15 個版本 ▾

[PDF] [WikiRelate! Computing semantic relatedness using Wikipedia](#) [PDF] [aaai.org](#)
[M Strube, SP Ponzetto](#) - AAAI, 2006 - new.aaai.org
... We also show that including Wikipedia improves the performance of an NLP application processing naturally occurring texts ... More interestingly, it indicates that Wikipedia can indeed be used as a resource for large-scale NLP applications ...
☆ ↗ 被引用 1012 次 相關文章 全部共 17 個版本 ▾

[PDF] [Extracting Lexical Semantic Knowledge from Wikipedia and Wiktionary.](#) [PDF] [semanticscholar.org](#)
[T Zesch, C Müller, I Gurevych](#) - LREC, 2008 - pdfs.semanticscholar.org
... 5. Example Usage in NLP The APIs for access to Wikipedia and Wiktionary proposed in this paper have already been put into service for large-scale NLP research, such as analyzing and accessing the structure of the Wikipedia category graph (Zesch and Gurevych, 2007 ...
☆ ↗ 被引用 391 次 相關文章 全部共 11 個版本 ▾

[Open information extraction using Wikipedia](#) [PDF] [aclweb.org](#)
[F Wu, DS Weld](#) - Proceedings of the 48th annual meeting of the ..., 2010 - dl.acm.org
... 3.1 Preprocessor The preprocessor converts the raw Wikipedia text into a sequence of sentences, attaches NLP annotations, and builds synonym sets for key entities. The resulting data is fed to the matcher, described in Section 3.2, which generates the training set ...
☆ ↗ 被引用 620 次 相關文章 全部共 15 個版本 ▾

[PDF] [Analyzing and accessing Wikipedia as a lexical semantic resource](#) [PDF] [psu.edu](#)
[T Zesch, I Gurevych, M Mühlhäuser](#) - Data Structures for Linguistic ..., 2007 - Citeseer
... If Wikipedia is to be used in large-scale NLP tasks, efficient programmatic access to these knowledge sources is required ... In order to foster the research on using Wikipedia in NLP, we made the Wikipedia API freely available to the research community ...
☆ ↗ 被引用 119 次 相關文章 全部共 7 個版本 ▾

[PDF] [Deriving a large scale taxonomy from Wikipedia](#) [PDF] [aaai.org](#)
[SP Ponzetto, M Strube](#) - AAAI, 2007 - new.aaai.org
... Simone Paolo Ponzetto and Michael Strube EML Research gGmbH Schloss-Wolfsbrunnenweg 33 69118 Heidelberg, Germany http://www.eml-research.de/nlp Abstract We take the category system in Wikipedia as a conceptual network ...
☆ ↗ 被引用 517 次 相關文章 全部共 19 個版本 ▾

[Knowledge derived from Wikipedia for computing semantic relatedness](#) [PDF] [jair.org](#)
[SP Ponzetto, M Strube](#) - Journal of Artificial Intelligence Research, 2007 - jair.org

Wikipedia in QA/IE Research

學術搜尋約有 67,400 項結果

- Reading comprehension
- Mining Wikipedia for QA
- Wikiqa
- Open domain QA
- TREC QA Treck
- Multilingual QA
- Finding similar sentences across languages
- Machine reading (Wikireading)
- Named Entity Relation (NER)

wikipedia question answering

學術搜尋 約有 67,400 項結果 (0.09 秒)

- [\[PDF\] arxiv.org](#) [\[PDF\] researchgate.net](#) [\[PDF\] aclweb.org](#) [\[PDF\] psu.edu](#)
- Reading wikipedia to answer open-domain questions**
D Chen, A Fisch, J Weston, A Bordes - arXiv preprint arXiv:1704.00051, 2017 - arxiv.org
... SQuAD **questions** have been collected via a process involving showing a paragraph to each human annotator and asking them to write a **question** ... QA that have been constructed in different ways (not necessarily in the context of **answering** from Wikipedia) ...
☆ ↗ 被引用 347 次 相關文章 全部共 14 個版本 ➔
- [PDF] Mining knowledge from wikipedia for the question answering task** [\[PDF\] researchgate.net](#)
D Buscaldi, P Rosso - Proceedings of the International Conference ..., 2006 - researchgate.net
Although significant advances have been made recently in the **Question Answering** technology, more steps have to be undertaken in order to obtain better results. Moreover, the best systems at the CLEF and TREC evaluation exercises are very complex systems based ...
☆ ↗ 被引用 69 次 相關文章 全部共 14 個版本 ➔
- [PDF] Wikiqa: A challenge dataset for open-domain question answering** [\[PDF\] aclweb.org](#)
Y Yang, W Yih, C Meek - Proceedings of the 2015 Conference on ..., 2015 - aclweb.org
... We present WIKIQA, a dataset for open- domain **question answering**.2 The dataset contains 3,047 **questions** originally sampled from Bing query logs. Based on the user clicks, each **question** is associated with a **Wikipedia** page presumed to be the topic of the **question** ...
☆ ↗ 被引用 292 次 相關文章 全部共 11 個版本 ➔
- Open domain question answering using Wikipedia-based knowledge model** [\[PDF\] researchgate.net](#)
PM Ryu, MG Jang, HK Kim - Information Processing & Management, 2014 - Elsevier
This paper describes the use of **Wikipedia** as a rich knowledge source for a **question answering** (QA) system. We suggest multiple answer matching modules based on different types of semi-structured knowledge sources of **Wikipedia**, including article content ...
☆ ↗ 被引用 39 次 相關文章 全部共 5 個版本 ➔
- 相關搜尋**
- question answering **system** [open domain question answering](#)
question answering **task** [conversational agents question answering](#)
- [PDF] Using Wikipedia at the TREC QA Track.** [\[PDF\] psu.edu](#)
D Ahn, V Jijkoun, G Mishne, K Müller, M de Rijke... - TREC, 2004 - Citeseer
... This year, our work for the **Question Answering** track was largely motivated by the wish to extend our QA system to handle the new, more complex **question** presentation, and to exploit **Wikipedia** at various stages of our QA architecture, for factoids and for “other”

Wikipedia in Multi-hop RC

學術搜尋約有 341 項結果

- HotpotQA (Yang, Peng, Zhang 2018)
 - <https://hotpotqa.github.io>
 - <https://arxiv.org/pdf/1809.09600.pdf>
- Wiki-Hop, QAngaroo, FEVER (Amazon)

Claim: The Rodney King riots took place in the most populous county in the USA.

[[wiki/Los Angeles Riots](#)]

The 1992 Los Angeles riots, also known as the Rodney King riots were a series of riots, lootings, arsons, and civil disturbances that occurred in Los Angeles County, California in April and May 1992.

[[wiki/Los Angeles County](#)]

Los Angeles County, officially the County of Los Angeles, is the most populous county in the USA.

Verdict: Supported

- <https://www.aclweb.org/anthology/>
N19-1405/

wikipedia AND "multi-hop" QA

學術搜尋 約有 341 項結果 (0.05 秒)

年份 ▾



Hotpotqa: A dataset for diverse, explainable multi-hop question answering [PDF] arxiv.org

Z Yang, P Qi, S Zhang, Y Bengio, WW Cohen... - arXiv preprint arXiv ..., 2018 - arxiv.org

... the model's ability to locate relevant facts as well as reasoning about them by requiring it to answer the question given the first paragraphs of all Wikipedia articles without ... This full wiki setting truly tests the performance of the systems' ability at multi-hop reasoning in ...

☆ ↗ 被引用 43 次 相關文章 全部共 5 個版本 ↘

Constructing datasets for multi-hop reading comprehension across documents [PDF] mitpressjournals...

J Welbl, P Stenetorp, S Riedel - Transactions of the Association for ..., 2018 - MIT Press

... The promise of multi-hop methods is finding and combining individual observations that can suggest previously unobserved DDIIs, aiding ... and WIKIPEDIA (8%), eg when the date of birth for a person differs between WIKI- DATA and what is stated in the WIKIPEDIA article, or ...

☆ ↗ 被引用 57 次 相關文章 全部共 8 個版本 ↘

Exploring graph-structured passage representation for multi-hop reading comprehension with graph neural networks [PDF] arxiv.org

L Song, Z Wang, M Yu, Y Zhang, R Florian... - arXiv preprint arXiv ..., 2018 - arxiv.org

... to correctly answer a question "What causes precipitation to fall?", a QA system only ... Note that the current multi-hop reading comprehension datasets all focus on the situation where the ... To adapt their model for the Wiki- Hop benchmark, where an answer candidate can con ...

☆ ↗ 被引用 15 次 相關文章 全部共 2 個版本 ↘

Cognitive Graph for Multi-Hop Reading Comprehension at Scale [PDF] arxiv.org

M Ding, C Zhou, Q Chen, H Yang, J Tang - arXiv preprint arXiv ..., 2019 - arxiv.org

... Intuitively, we adopt a directed graph structure for step-by-step deduction and exploration in cognitive process of multi-hop QA ... Algorithm 1: Cognitive Graph QA Input: System 1 model S1, System 2 model S2, Question Q, Predictor F, Wiki Database W 1 Initialize ...

☆ ↗ 被引用 4 次 相關文章 全部共 4 個版本 ↘

Multi-hop Inference for Sentence-level TextGraphs: How Challenging is Meaningfully Combining Information for Science Question Answering? [PDF] arxiv.org

P Jansen - arXiv preprint arXiv:1805.11267, 2018 - arxiv.org

... manually-annotated judgements across knowledge graphs built from three free-text corpora (including study guides and Simple Wikipedia) ... and methods for aggregation, and is leading to both the development of datasets specifically designed for multi-hop QA (Jansen et al ...

☆ ↗ 被引用 1 次 相關文章 全部共 5 個版本 ↘

Wikipedia in

Grammatical Error Correction

學術搜尋約有 301 項結果

- Wikipedia revisions (prepositions)
- WikEd Error Corpus (WEEC)
- Wiki Edits
 - GitHub
- Syntactic ngram language model
- Human-level GEC (shame on Singapore)
- Human-level GEC (shame on Microsoft)

wikipedia "grammatical error correction"



約有 301 項結果 (0.08 秒)

[PDF] Robust systems for preposition error correction using wikipedia revisions

[A Cahill, N Madnani, J Tetreault...](#) - Proceedings of the 2013 ..., 2013 - aclweb.org

... To evaluate the usefulness of Wikipedia revision history for grammatical error correction, we address the task of correcting errors in preposition selection (ie, where the context licenses the use of a prepo- sition, but ... Wikipedia Snapshot Artificial Errors (Wiki) 31.17 * 24.52 ...

☆ ↗ 被引用 34 次 相關文章 全部共 7 個版本 »»

The WikEd error corpus: A corpus of corrective Wikipedia edits and its application to grammatical error correction

[R Grundkiewicz, M Junczys-Dowmunt](#) - International Conference on ..., 2014 - Springer

This paper introduces the freely available WikEd Error Corpus. We describe the data mining process from Wikipedia revision histories, corpus content and format. The corpus consists of more than 12 million sentences with a total of 14 million edits of various types. As one ...

☆ ↗ 被引用 15 次 相關文章 全部共 8 個版本

A multilayer convolutional encoder-decoder neural network for grammatical error correction

[S Chollampatt, HT Ng](#) - Thirty-Second AAAI Conference on Artificial ..., 2018 - aaai.org

... Grammatical error correction (GEC) is a well-established natural language processing (NLP) task that deals with ... We also make use of the larger English corpora from Wikipedia (1.78B ... This work (using additional corpora) MLConverbed L8, NUCLE Yes Wiki 60.90 23.74 46.38 ...

☆ ↗ 被引用 35 次 相關文章 全部共 5 個版本 »»

[PDF] Generating artificial errors for grammatical error correction

[M Felice, Z Yuan](#) - Proceedings of the Student Research Workshop at the ..., 2014 - aclweb.org

... new approach based on linguistic information and evaluate it using the test data provided for the CoNLL 2013 shared task on grammatical error correction (Ng et ... In particular, they compile a set of sentences from the English Wikipedia and apply the following generation methods ...

☆ ↗ 被引用 33 次 相關文章 全部共 6 個版本 »»

Connecting the dots: Towards human-level grammatical error correction

[S Chollampatt, HT Ng](#) - Proceedings of the 12th Workshop on Innovative ..., 2017 - aclweb.org

... Grammatical error correction (GEC) is the task of correcting various textual errors includ- ing spelling, grammar, and collocation errors ... use two 5-gram LMs, one of them trained on the target side of the parallel training data and the other trained on Wikipedia texts (Wiki LM) ...

☆ ↗ 被引用 20 次 相關文章 全部共 4 個版本 »»

[PDF] CoNLL 2014 shared task: Grammatical error correction with a syntactic n-gram language model from a big corpora

[SD Hdez, H Calvo](#) - Proceedings of the Eighteenth Conference on ..., 2014 - aclweb.org

... Daniel Dahlmeier and Hwee Tou Ng 2012. Better Evaluation for Grammatical Error Correction ... 2012. Multithread-Wikipedia-Extractor for SMP based architectures, Ver- sion: 1.0 (October 15,

從哪裡開始？

- dumps.wikimedia.org/enwiki/latest/
- [enwiki-latest-pages-articles.xml.bz2](https://dumps.wikimedia.org/enwiki/latest/enwiki-latest-pages-articles.xml.bz2)
16.5 GB (台北大學2MB/秒 <200分鐘)
- [simplewiki-latest-pages-articles.xml.bz2](https://dumps.wikimedia.org/enwiki/latest/simplewiki-latest-pages-articles.xml.bz2) 158 MB (<2 分鐘) 解壓縮 730 MB
- 寫程式嗎？我們儘量跑 Unix
 - bzgrep -e 或 bzgrep -e -o
 - 一行可生出 graph adjacency list
- 自己寫 code，不如 call **networks**
import networkx **as** nx
G = nx.read_adjlist('en.link.adjlist.txt',
create_using = nx.DiGraph(),
nodetype = int)

Index of /enwiki/

..		
20190701/	21-Aug-2019 01:36	
20190720/	02-Sep-2019 01:29	
20190801/	22-Sep-2019 01:38	
20190820/	02-Oct-2019 01:29	
20190901/	08-Sep-2019 17:37	
20190920/	23-Sep-2019 13:47	
20191001/	03-Oct-2019 07:40	
latest/	03-Oct-2019 07:40	

Index of /simplewiki/latest/

simplewiki-latest-pages-articles.xml.bz2	02-Oct-2019 15:03	158798130
simplewiki-latest-pages-articles.xml.bz2-rss.xml	03-Oct-2019 04:29	809
simplewiki-latest-pages-logging.xml.gz	24-Sep-2019 12:23	46779990
simplewiki-latest-pages-logging.xml.gz-rss.xml	24-Sep-2019 12:23	803
simplewiki-latest-pages-meta-current.xml.bz2	24-Sep-2019 03:21	227252903
simplewiki-latest-pages-meta-current.xml.bz2-rs..>	24-Sep-2019 12:16	821
simplewiki-latest-protected_titles.sql.gz	01-Oct-2019 23:23	14752
simplewiki-latest-protected_titles.sql.gz-rss.xml	03-Oct-2019 04:24	812
simplewiki-latest-redirect.sql.gz	01-Oct-2019 23:25	734267
simplewiki-latest-redirect.sql.gz-rss.xml	03-Oct-2019 04:20	788

- 我們提供 Jupyter notebook 有完整流程
- 資訊框、摘要、本文，通常是 XML 格式
 - pages.articles.xml 有最完整的資訊
- 標題、主題分類、連結，通常是 SQL 格式
 - e.g., redirect.sql.gz
- 一般文章、分類、解歧義都是 pages

從哪裡開始？

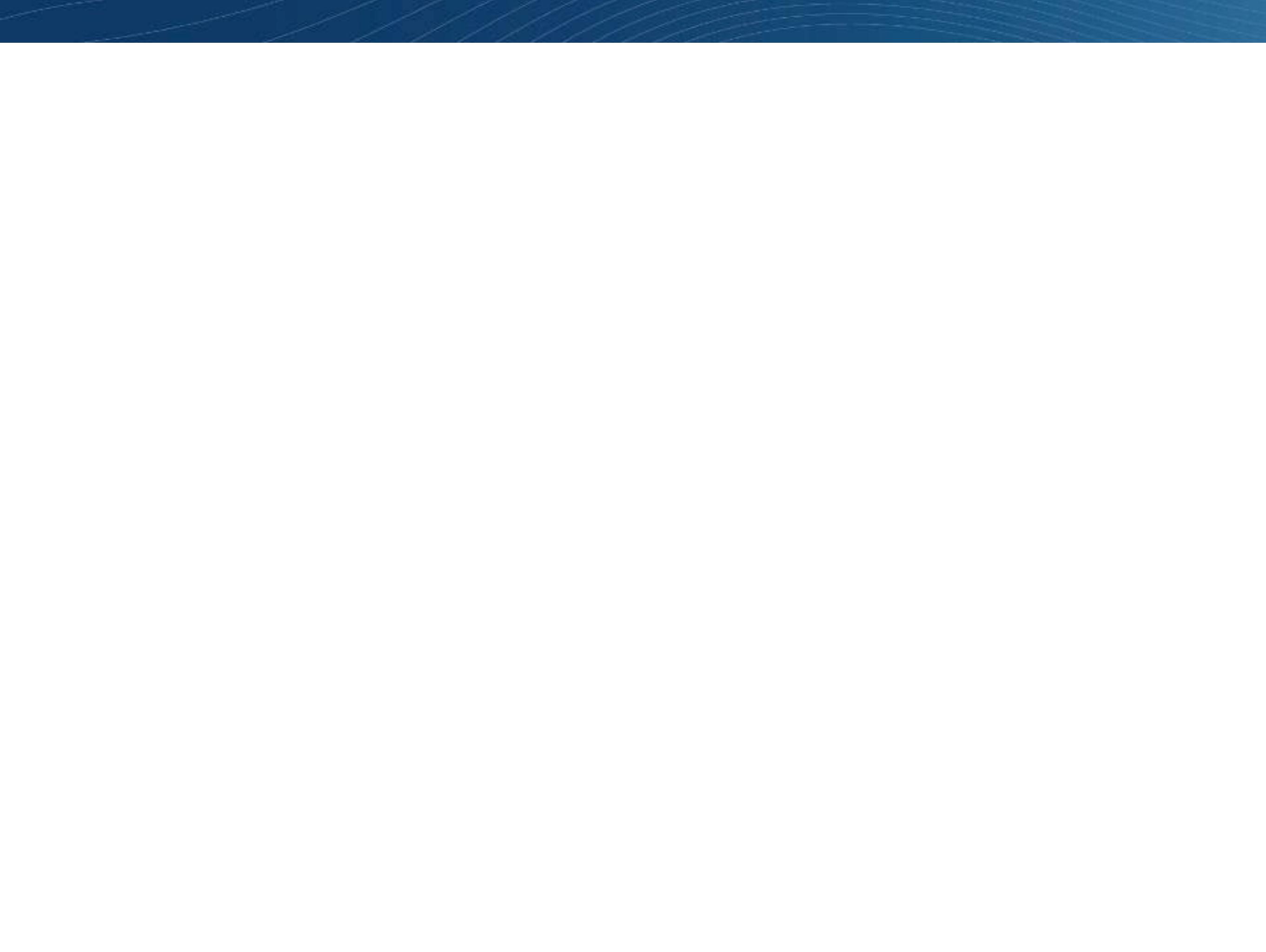
- 我們提供 Jupyter notebook 有完整流程
- 資訊框、摘要、本文，通常是 XML 格式
 - pages.articles.xml 有最完整的資訊
- 標題、主題分類、連結，通常是 SQL 格式
 - e.g., redirect.sql.gz
- 一般文章、分類、解歧義都是 pages

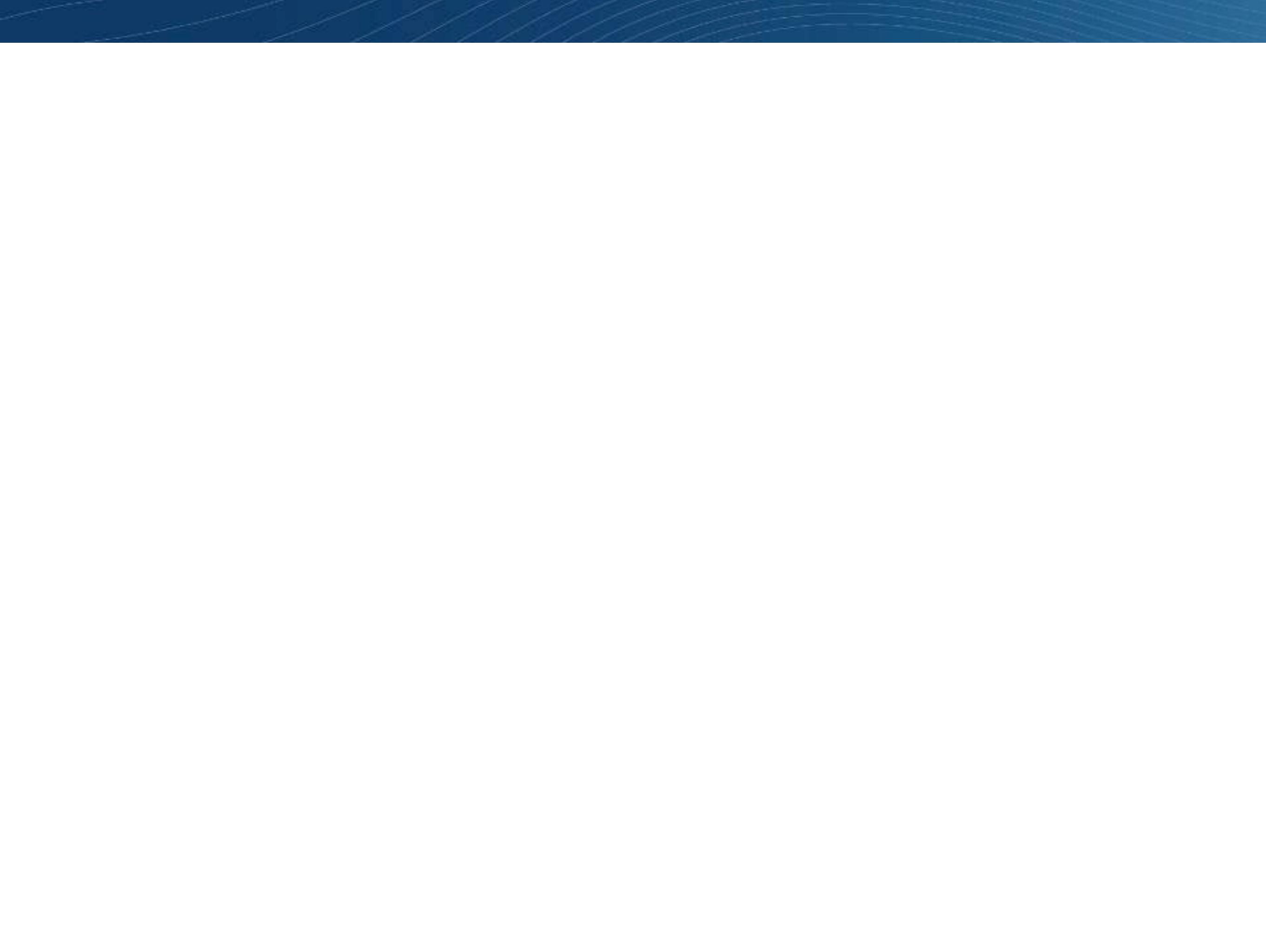
Index of /enwiki/

.. /		
20190701 /	21-Aug-2019 01:36	
20190720 /	02-Sep-2019 01:29	
20190801 /	22-Sep-2019 01:38	
20190820 /	02-Oct-2019 01:29	
20190901 /	08-Sep-2019 17:37	
20190920 /	23-Sep-2019 13:47	
20191001 /	03-Oct-2019 07:40	
latest /	03-Oct-2019 07:40	

Index of /simplewiki/latest/

simplewiki-latest-pages-articles.xml.bz2	02-Oct-2019 15:03	158798130
simplewiki-latest-pages-articles.xml.bz2-rss.xml	03-Oct-2019 04:29	809
simplewiki-latest-pages-logging.xml.gz	24-Sep-2019 12:23	46779990
simplewiki-latest-pages-logging.xml.gz-rss.xml	24-Sep-2019 12:23	803
simplewiki-latest-pages-meta-current.xml.bz2	24-Sep-2019 03:21	227252903
simplewiki-latest-pages-meta-current.xml.bz2-rs..>	24-Sep-2019 12:16	821
simplewiki-latest-protected_titles.sql.gz	01-Oct-2019 23:23	14752
simplewiki-latest-protected_titles.sql.gz-rss.xml	03-Oct-2019 04:24	812
simplewiki-latest-redirect.sql.gz	01-Oct-2019 23:25	734267
simplewiki-latest-redirect.sql.gz-rss.xml	03-Oct-2019 04:20	788





Generating a

Wikipedia Roger's Thesaurus (WRT)

by Aligning Category Graphs

Wikipedia

- Wikipedia holds large amount of data
- The Category system and basic structure of website holds a tree structure
- However, Wikipedia category system is highly unorganized
- We decided to use an organized dataset to help re-organize Wikipedia

Roget's Thesaurus

- Roget's Thesaurus has a four layers structure
 - 24 Main Categories
 - 856 Categories
 - ~10,000 Headwords
 - ~ 100,000 Words / Phrases

Main Topic: Foundations of Knowledge

Topic: Mathematics

4. Number

Number, signed number, directed number, positive number, ... , random number

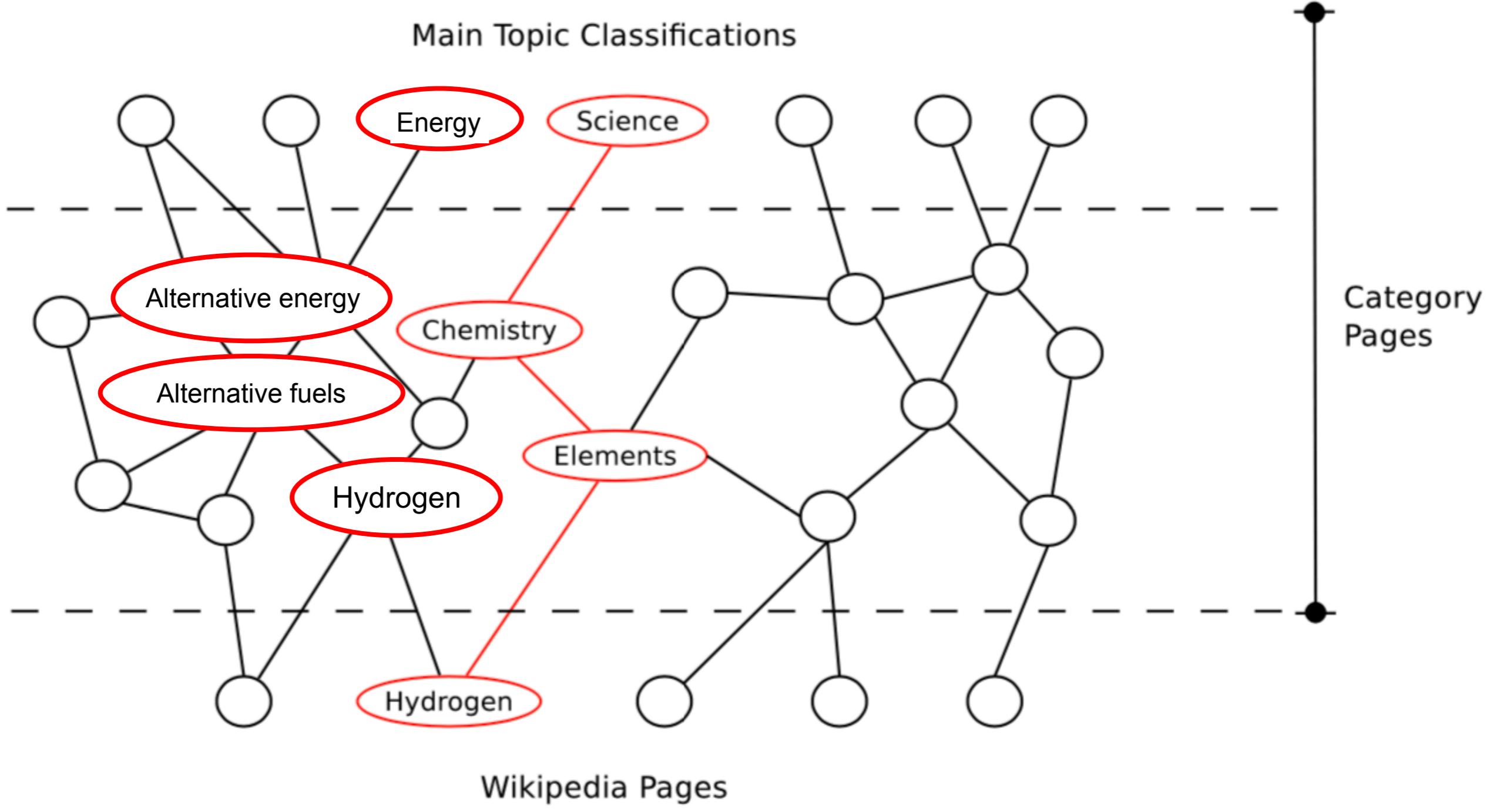
Example

- Roget's

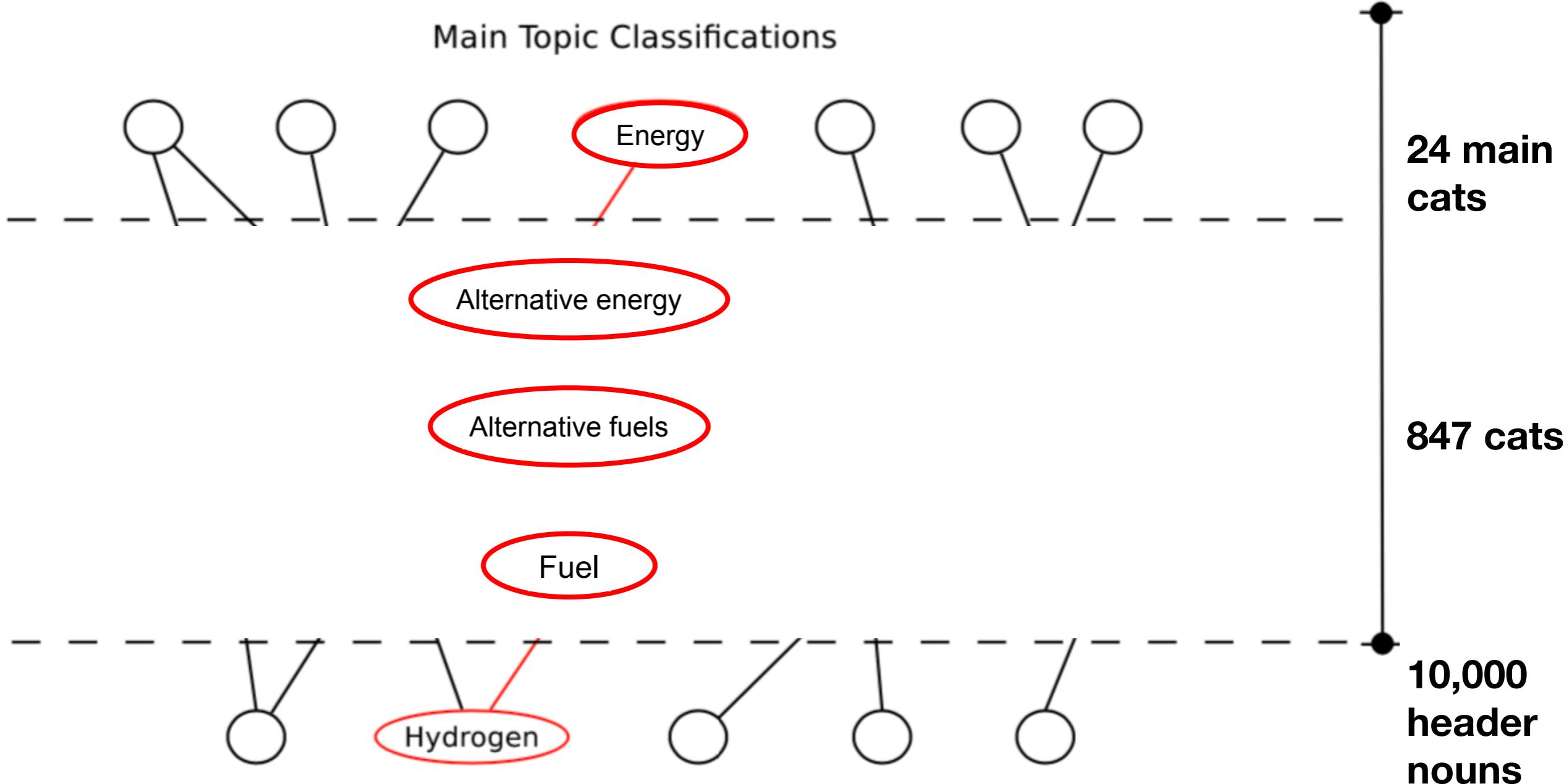
Foundations of Knowledge -> Mathematics -> number

- Wikipedia

Mathematics -> Mathematical concepts -> Mathematical_objects -> Number

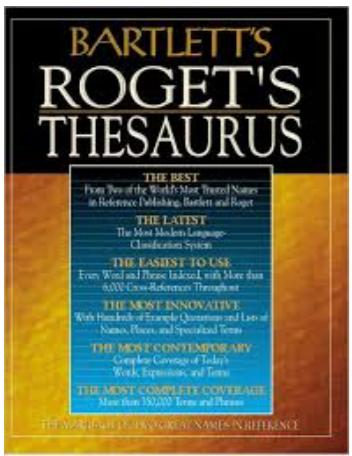


Complicated and Noisy Wikipedia Category Graph



Bartlett's Roget's Thesaurus (BRT)

Aligning BRT-Wiki cat paths



Foundations_of_Knowledge Biochemistry essential_element hydrogen
||| C:Nature C:Matter Atoms C:Chemical_elements Hydrogen
0-0 1-1 1-2 2-3 3-4

BRT path

=====

Foundations_of_Knowledge
Biochemistry
Biochemistry
essential_element
Hydrogen

Wikicat path

=====

C:Nature
C:Matter
C:Atoms
C:Chemical_elements
Hydrogen



WIKIPEDIA
The Free Encyclopedia

Condition Propulsion propellant hydrogen

||| C:Energy C:Alternative_energy C:Alternative_fuels C:Hydrogen Hydrogen
0-0 1-1 1-2 2-3 3-4

BRT path

=====

Condition
Propulsion
Propulsion
propellant
hydrogen

Wikicat path

=====

C:Energy
C:Alternative_energy
C:Alternative_fuels
C:Hydrogen
Hydrogen

WikiSense

– aligning Wikipedia NEs with WordNet Supersenses

Aligning BRT-Wiki cat paths



WIKIPEDIA
The Free Encyclopedia

Foundations_of_Knowledge Biochemistry essential_element hydrogen
||| C:Nature C:Matter Atoms C:Chemical_elements Hydrogen
0-0 1-1 1-2 2-3 3-4

BRT path

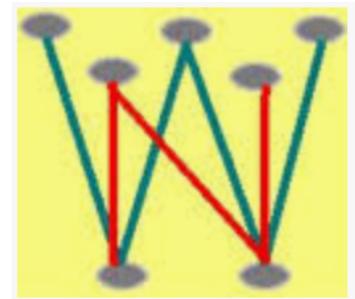
=====

Foundations_of_Knowledge
Biochemistry
Biochemistry
essential_element
Hydrogen

Wikicat path

=====

C:Nature
C:Matter
C:Atoms
C:Chemical_elements
Hydrogen



Condition Propulsion propellant hydrogen
||| C:Energy C:Alternative_energy C:Alternative_fuels C:Hydrogen Hydrogen
0-0 1-1 1-2 2-3 3-4

BRT path

=====

Condition
Propulsion
Propulsion
propellant
hydrogen

Wikicat path

=====

C:Energy
C:Alternative_energy
C:Alternative_fuels
C:Hydrogen
Hydrogen

Using Wikipedia in Multi-hop Reading Comprehension

WikiHop (Wikipedia)

- Challenge: combining multiple documents to infer the correct answer
- qangaroo.cs.ucl.ac.uk

Big Oak Tree State Park is a state - owned nature preserve ... in the Mississippi Alluvial Plain portion of the **Gulf Coastal Plain**.

The **Gulf Coastal Plain** extends around the Gulf of Mexico in the **Southern United States**...

The **Southern United States**, commonly referred to as the American South, Dixie, or simply the South, is a region of the **United States of America**.

Q: (**Big Oak Tree State Park**, located in, ?)
A: **United States of America**

The Hanging Gardens, in **Mumbai**, also known as Pherozeshah Mehta Gardens, are terraced gardens ... They provide sunset views over the **Arabian Sea** ...

Mumbai (also known as Bombay, the official name until 1995) is the capital city of the Indian state of Maharashtra. It is the most populous city in **India** ...

The **Arabian Sea** is a region of the northern Indian Ocean bounded on the north by **Pakistan** and **Iran**, on the west by northeastern **Somalia** and the Arabian Peninsula, and on the east by **India** ...

Q: (Hanging gardens of Mumbai, country, ?)
Options: {Iran, **India**, Pakistan, Somalia, ...}

MedHop (Medline)

- Necessary to combine multiple documents to infer the correct answer
- INPUT: a collection of query-answer pairs and thematically linked documents
- OUTPUT: multi-hop reading comprehension datasets
- qangaroo.cs.ucl.ac.uk

Leuprolide ... elicited a long-lasting potentiation of excitatory postsynaptic currents. ... **GnRH receptor**-induced synaptic potentiation was blocked ... by **Progonadotropin-releasing hormone-1**, a specific **GnRH receptor** antagonist.

Analyses of gene expression demonstrated a dynamic response to the **Progonadotropin-releasing hormone-1** superagonist **Triptorelin**.

Q: (**Leuprolide**, interacts_with, ?)

A: **Triptorelin**

WikiReading (Hewlett et al., 2016)

- Tuples (*item*, *property*, *answer*) in WIKIDATA are aligned with WIKIPEDIA *articles* with *item*
- (*item*, *property*, *answer*) define a task:
 - Given *article* and *property*
 - Predicting *answer*
- Problem: 54.4% of the samples do not state the *answer* explicitly in *article*
- Solution: articles linked from *article has the answer*

MedHop (Medline)

- Necessary to combine multiple documents to infer the correct answer
- INPUT: a collection of query-answer pairs and thematically linked documents
- OUTPUT: multi-hop reading comprehension datasets
- qangaroo.cs.ucl.ac.uk

Leuprolide ... elicited a long-lasting potentiation of excitatory postsynaptic currents. ... **GnRH receptor**-induced synaptic potentiation was blocked ... by **Progonadotropin-releasing hormone-1**, a specific **GnRH receptor** antagonist.

Analyses of gene expression demonstrated a dynamic response to the **Progonadotropin-releasing hormone-1** superagonist **Triptorelin**.

Q: (**Leuprolide**, interacts_with, ?)

A: **Triptorelin**

MedHop (Medline)

- Necessary to combine multiple documents to infer the correct answer
- INPUT: a collection of query-answer pairs and thematically linked documents
- OUTPUT: multi-hop reading comprehension datasets
- qangaroo.cs.ucl.ac.uk

Leuprolide ... elicited a long-lasting potentiation of excitatory postsynaptic currents. ... **GnRH receptor**-induced synaptic potentiation was blocked ... by **Progonadotropin-releasing hormone-1**, a specific **GnRH receptor** antagonist.

Analyses of gene expression demonstrated a dynamic response to the **Progonadotropin-releasing hormone-1** superagonist **Triptorelin**.

Q: (**Leuprolide**, interacts_with, ?)

A: **Triptorelin**

References

- Mihalcea, R. and Csoma, A. Wikify! linking documents to encyclopedic knowledge. *Proceedings of CIKM'07*, Lisbon, Portugal
- V. Nguyen & M. Poesio, 2012. Entity disambiguation and linking over queries using Encyclopedic Knowledge, Proceedings of 6th workshop on Analytics for Noisy Unstructured Text Data
- D. Lungley, M. Trevisan, V. Nguyen, M. Althobaiti, M. Poesio, 2013. GALATEAS D2W: A Multi-lingual Disambiguation to Wikipedia Web Service. Proc. Of ENRICH.
- V. Nastase, and M. Strube, Transforming Wikipedia into a large scale multilingual concept network, *Artificial Intelligence*, 2012
- L. von Ahn and L. Dabbish (2008). Designing games with a purpose. *Communications of the ACM*, v. 51, n.8, 58-67
- Poesio, Chamberlain, Kruschwitz, Robaldo, & Ducceschi, 2013. Phrase Detectives: Utilizing Collective Intelligence for Internet-Scale Language Resource Creation. *ACM Transactions on Intelligent Interactive Systems*