



KDI ● **Knowledge and Data Integration**

Resources

Phase: 1. Introduction &
Representation Diversity

W1.L3.M1.T32

Contents

1 Linguistic Resources

2 Knowledge Resources

3 Data Resources

Contents

1 Linguistic Resources

2 Knowledge Resources

3 Data Resources

Linguistic Resources

A linguistic resource is a dataset which provides data about languages (e.g., meanings, relations between words, ...).

There are two types of mono/multi-lingual resources: (i) online dictionaries and (ii) Wordnet like resources. Wordnets much more useful in data integration as they connect meanings of words in a LKG.

Check the licence (lots of options).

Example

- Global Wordnet Association
- WordNet
- Open Multilingual WordNet
- Datascientia/UKC (forthcoming)

Linguistic Resource Repositories

Global WordNet Association

[Home](#) [About GWA](#) [Home](#) [Resources](#) [Global WordNet Conferences](#) [Contact](#)

Global WordNet Association

**** 10th Conference 2019 ****

A free, public and non-commercial organization that provides a platform for discussing, sharing and connecting wordnets for all languages in the world.

[More info on GWA](#)



**Global
WordNet
Association**

Figure: Global WordNet Association¹

¹<http://globalwordnet.org/>

Linguistic Resource Repositories



What is WordNet

People

News

Use Wordnet Online¹

Download

Citing WordNet

License and Commercial Use

Related Projects

Documentation

Publications

Frequently Asked Questions

What is WordNet?

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the creators of WordNet and do not necessarily reflect the views of any funding agency or Princeton University.

When writing a paper or producing a software application, tool, or interface based on WordNet, it is necessary to properly [cite the source](#). Citation figures are critical to WordNet funding.

About WordNet

WordNet[®] is a large lexical database of English. Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. Synsets are interlinked by means of conceptual-semantic and lexical relations. The resulting network of meaningfully related words and concepts can be navigated with the [browser #](#). WordNet is also freely and publicly available for [download](#). WordNet's structure makes it a useful tool for computational linguistics and natural language processing.

WordNet superficially resembles a thesaurus, in that it groups words together based on their meanings. However, there are some important distinctions. First, WordNet interlinks not just word forms—strings of letters—but specific senses of words. As a result, words that are found in close proximity to one another in the network are semantically disambiguated. Second, WordNet labels the semantic relations among words, whereas the groupings of words in a thesaurus does not follow any explicit pattern other than meaning similarity.

Structure

Note

Due to funding and staffing issues, we are no longer able to accept comment and suggestions.

We get numerous questions regarding topics that are addressed on our [FAQ](#) page. If you have a problem or question regarding something you downloaded from the ["Related projects"](#) page, you must contact the developer directly.

Please note that any changes made to the database are not reflected until a new version of WordNet is publicly released. Due to limited staffing, there are currently no plans for future WordNet releases.

Figure: WordNet Home¹

¹ <https://wordnet.princeton.edu/>

Linguistic Resource Repositories

Open Multilingual Wordnet

This page provides access to open wordnets in a variety of languages, all linked to the [Princeton Wordnet of English](#) (PWN). The goal is to make it easy to use wordnets in multiple languages. The individual wordnets have been made by many different projects and vary greatly in size and accuracy. We have (i) extracted and normalized the data, (ii) linked it to Princeton WordNet 3.0 and (iii) put it in one place. The Open Multilingual Wordnet and its components are [open](#): they can be freely used, modified, and shared by anyone for any purpose. There is a fuller list of wordnets at the Global Wordnet Association's [Wordnets in the World page](#).

If you use these wordnets, please cite the original projects who created them (linked in Table 1), if you got value from this aggregation/normalization, please cite [Bond and Paik \(2012\)](#).

You can access the wordnets through the (python) [Natural Language Tool-Kit wordnet interface \(NLTK\)](#).

We have an [extended version](#) with automatically extracted data for over a 150 languages from [Wiktionary](#) and the [Unicode Common Locale Data Repository](#) ([Bond and Foster, 2013](#)).

[Documentation, News and Updates](#)

Search

We have a [simple search interface](#) (search [the extended wordnet](#)). It uses the SQL database originally developed by the Japanese Wordnet.

34 Open Wordnets Merged

Wordnet	Lang	Synsets	Words	Senses	Core	Licence	Data	Citation
Albanet	als	4,675	5,988	9,599	31%	CC BY 3.0	als.zip (+xml)	cite.als: (.bib)
Arabic WordNet (AWN v2)	arb	9,916	17,785	37,335	47%	CC BY SA 3.0	arb.zip (+xml)	cite.arb: (.bib)
BulTreeBank Wordnet (BTB-WN)	bul	4,959	6,720	8,936	99%	CC BY 3.0	bul.zip (+xml)	cite.bul: (.bib)
Chinese Open Wordnet	cmn	42,312	61,533	79,809	100%	wordnet	cmn.zip (+xml)	cite.cmn: (.bib)

Figure: Open Multilingual WordNet Home¹

¹ <http://compling.hss.ntu.edu.sg/omw/>

Linguistic Resource Repositories



The lexicons we support



Vision and Mission

The Universal Knowledge Core (UKC) is a psycholinguistic principles based multilingual, high quality, large scale, and diversity aware machine readable lexical resource.

The key design principle underlying the UKC is to maintain a clear distinction between the language(s) used to describe the world as it is perceived and what is being described, i.e., the world itself. The Concept Core (CC) is the UKC representation of the world and it consists of a semantic network where nodes are

Figure: UKC Home¹

¹<http://ukc.disi.unitn.it>

Contents

1 Linguistic Resources

2 Knowledge Resources

3 Data Resources

Knowledge Resources

A Knowledge resource is a dataset which consists of a KB encoding information about schemas (etypes and properties).

KBs of high quality are usually called ontologies. We call them teleologies (meaning by this, ontologies with metadata which empower their practical use in knowledge and data integration).

Example

- LOV/LOV4IoT
- Schema.org
- DBpedia (schema only)
- Datascientia/liveschema (forthcoming)

Knowledge Resource Repositories

The screenshot shows the LOV website interface. At the top, there is a logo consisting of several colored squares (orange, blue, green) and a navigation menu with the following items: VOCABS, TERMS, AGENTS, and SPARQL/DUMP. Below the navigation menu is a large teal banner with the text "Linked Open Vocabularies (LOV)". Underneath the banner is a row of five dark grey buttons: "+ Suggest", "Documentation" (with a document icon), "Follow" (with a person icon), a search input field with a magnifying glass icon, and a button with a person icon. Below this row is a large teal area containing a bubble chart titled "721 Vocabularies in LOV". The chart shows various vocabularies represented by circles of different sizes and colors. The largest circles are labeled "vann", "foaf", "skos", "dcterms", and "dce". To the right of the bubble chart is a section titled "Latest insertion" with a list of recent additions:

- fiesta-priv** - FIESTA-Priv Ontology
2020-08-07
- sdm** - SPARQL endpoint metadata
2020-07-24
- oum** - Ontology of units of Measure (OM)
2020-07-24
- dg** - DINGO Ontology
2020-07-24
- sur** - The Survey Ontology
2020-07-24

Figure: Linked Open Vocabulary²

²<https://lov.linkeddata.es/dataset/lov/>

Knowledge Resource Repositories

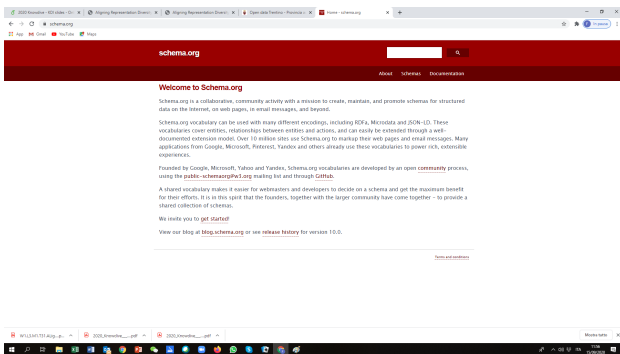


Figure: Schema.org²

²<http://www.schema.org/>

Knowledge Resource Repositories



Figure: DBpedia Home²

²<https://wiki.dbpedia.org/>

Knowledge Resource Repositories

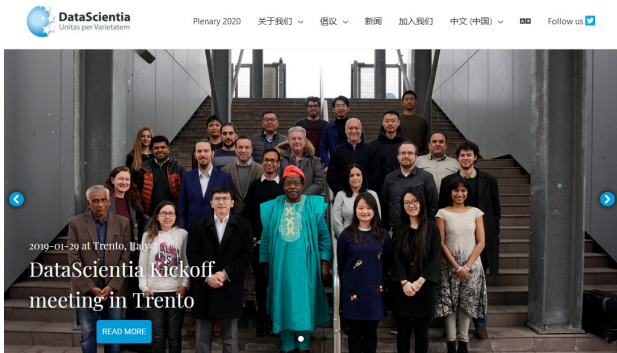


Figure: DataScientia Home²

²<http://datascientia.disi.unitn.it/zh/home-zh/>

Contents

1 Linguistic Resources

2 Knowledge Resources

3 Data Resources

Data Resources

A data resource is a dataset which consists of data in some format (tabular, unstructured, entities and property values).

Open Data: data freely available. Check the licence (lots of options).

Example

- UK Open Data
- National Bureau of Statistics, China
- data.org
- Opendata Trentino (see, among others, Unitn Open Data)
- Geonames
- Open Street Map
- DBPedia
- Data Hub

Data Resource Repositories

data.gov.uk | Find open data

Publish your data Documentation Support

BETA This is a new service – your [feedback](#) will help us to improve it

Find open data

Find data published by central government, local authorities and public bodies to help you build products and services



[Business and economy](#)

Small businesses, industry, imports, exports and trade

[Crime and justice](#)

Courts, police, prison, offenders, borders and immigration

[Defence](#)

Armed forces, health and safety, search and rescue

[Education](#)

[Environment](#)

Weather, flooding, rivers, air quality, geology and agriculture

[Government](#)

Staff numbers and pay, local councillors and department business plans

[Government spending](#)

Includes all payments by government departments over £25,000

[Mapping](#)

Addresses, boundaries, land ownership, aerial photographs, seabed and land terrain

[Society](#)

Employment, benefits, household finances, poverty and population

[Towns and cities](#)

Includes housing, urban planning, leisure, waste and energy, consumption

Figure: Open Data UK³

³<https://data.gov.uk/>

Data Resource Repositories

登录 | 注册 | English

National data 国家数据
国家统计局
National Bureau of Statistics

首页 | 月度数据 | 季度数据 | 年度数据 | 普查数据 | 地区数据 | 部门数据 | 国际数据 | 可视化产品 | 出版物 | 我的收藏 | 帮助

查数 CHASHU 如: 2012年 北京 GDP 搜索 统计 热词 GDP CPI 总人口 社会消费品零售总额 粮食产量 PMI PPI

“数据中国”再升级
为了更好地满足社会公众随时随地对数据查询的需求，我们升级了数据中国 App 客户端，旧版“数据中国”将停止维护使用，欢迎大家体验手机端数据库，并提宝贵意见。
掌上数据库客户端依然使用
扫描下方二维码
数据中国 pro (中文 英文)
掌上数据库
(安卓版用微信以外的扫描工具)

数据中国 pro
App 再升级

第三次全国农业普查
第三次全国农业普查
资料下载

统计数据产品 工作年度报表
图表中国
数据中国(中英)
如何获取统计数据
数据地图

官方微博(人民 新华 新浪 腾讯) 官方微信

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 更多>>

Figure: National Bureau of Statistics, China³

³<http://data.stats.gov.cn/>

Data Resource Repositories

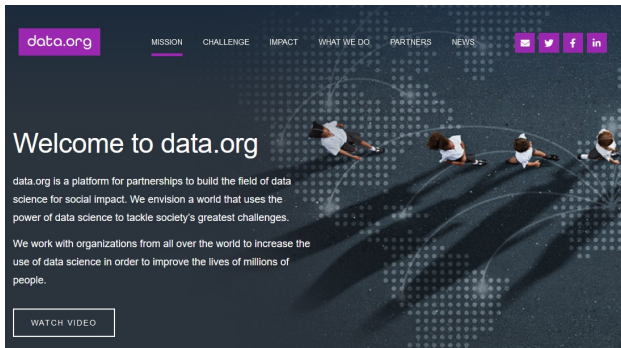


Figure: data.org³

³<https://www.data.org/>

Data Resource Repositories



Figure: Open Data Trentino³

³<http://dati.trentino.it/>

Data Resource Repositories



The GeoNames geographical database covers all countries and contains over eleven million placenames that are available for download free of charge.

[\[advanced search\]](#)

enter a location name, ex: "Paris", "Mount Everest", "New York"

Browse the names <ul style="list-style-type: none">CountriesPostal codesCountry statisticsRecent modifications	Information <ul style="list-style-type: none">About GeoNamesData SourcesUser manualAmbassadors and TeamForumBlogMailing listCommercial Support and Consulting	Download <ul style="list-style-type: none">InfoFree Gazetteer DataFree Postal Code DataPremium Data Web Services <ul style="list-style-type: none">OverviewDocumentationClient LibrariesPremium Web Services
--	---	--

Figure: Geonames Home³

³<https://www.geonames.org/>

Data Resource Repositories

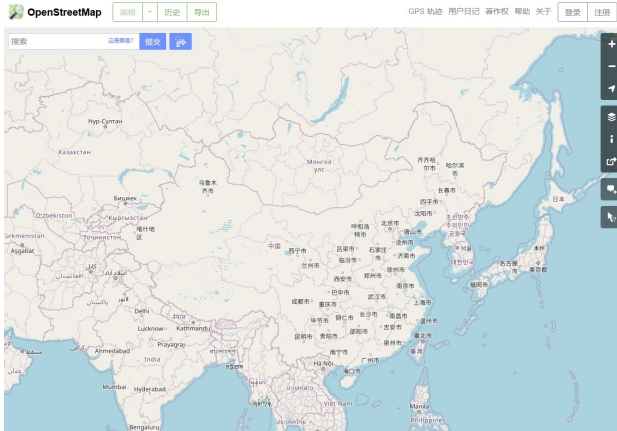


Figure: Open Street Map Home³

³<https://www.openstreetmap.org/>

Data Resource Repositories

DBpedia

Data Download

[Browse the DBpedia Datasets](#) [Go to Latest Release](#)

 Apply	 Develop	 Research	 Join	 Contribute
Explore current DBpedia projects and applications. Learn about interlinking and	Develop amazing things with our DBpedia datasets and our API. Need	Find current research papers, journal articles and conference	Get in touch with us. Become a member of the DBpedia Association, join our	Check out DBpedia's career opportunities, our challenges, website contributions,

Figure: DBpedia Home³

³<https://wiki.dbpedia.org/>

Data Resource Repositories



[ABOUT](#) [BLOG](#) [FIND DATA](#) [COLLECTIONS](#) [DOCS](#) [PRICING](#) [TOOLS](#) [CHAT](#) [LOGIN](#) [JOIN FREE](#)



We help organizations of all sizes to design, develop and scale solutions to manage their data and unleash its potential.

Let us help you!

Get in touch now »



Figure: Data Hub Home³

³<https://datahub.io/>



W1.L3.M1.T32



Resources

Phase: 1. Introduction &
Representation Diversity