Will there still be dictionaries in 2020?

Piek Vossen,
VU University Amsterdam

YES.

Will there be dictionary publishers in 2020?

NO.

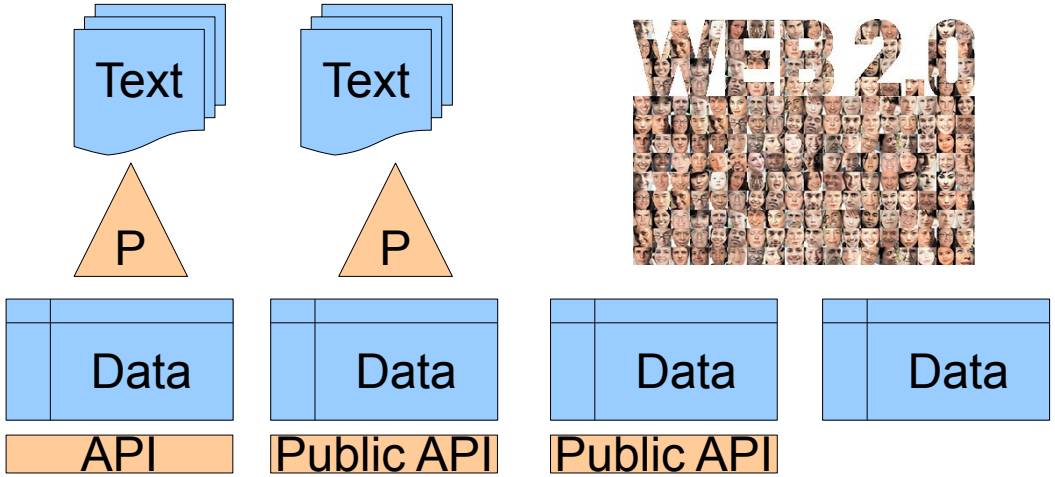
Perhaps they (publishers) are no longer what we used to think.

Who are the users of lexical data?

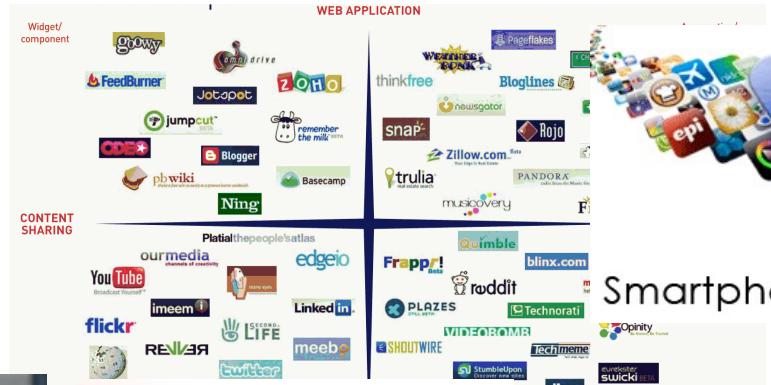
Language

Extraction And/or editing

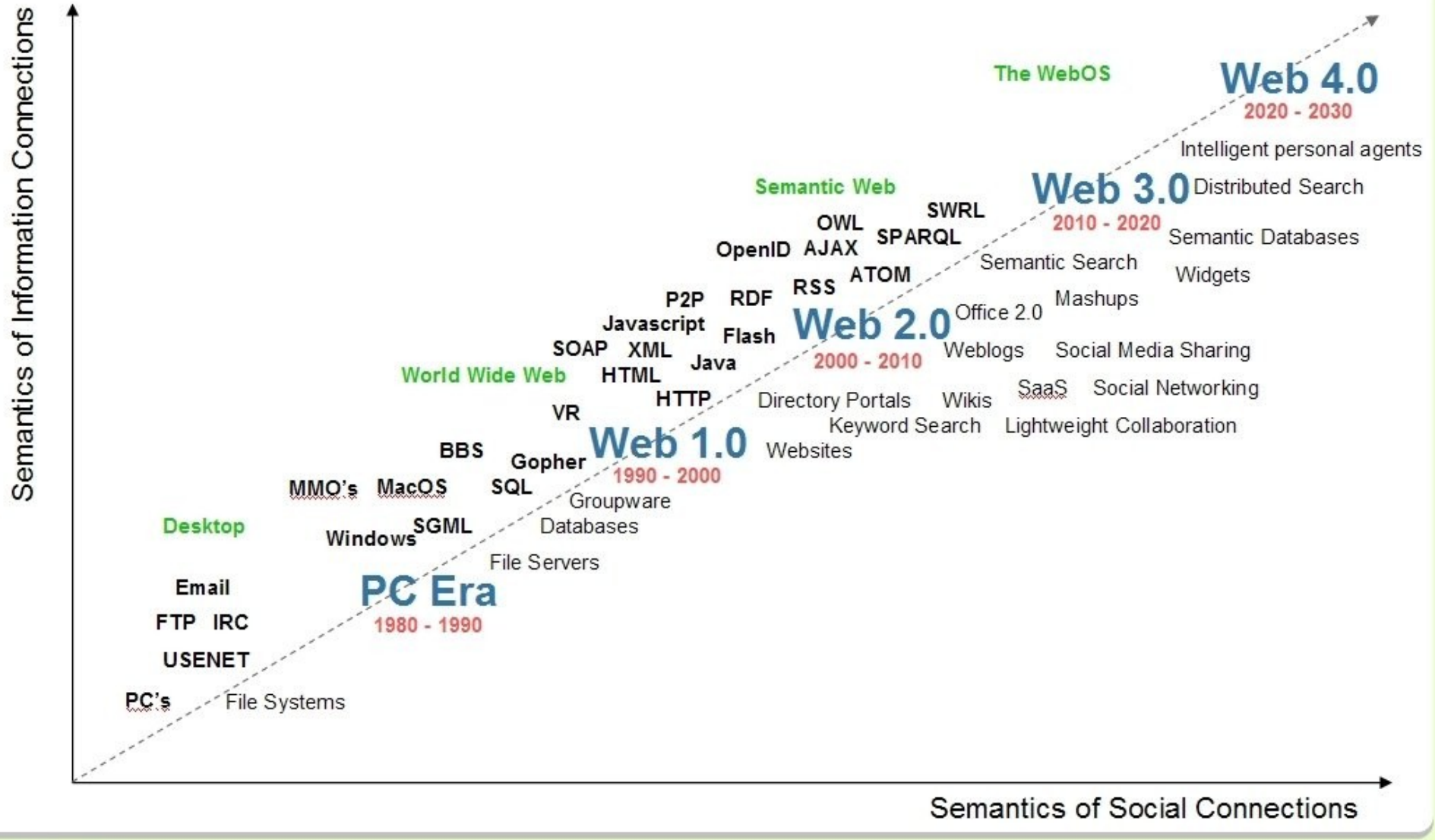
Database



Software



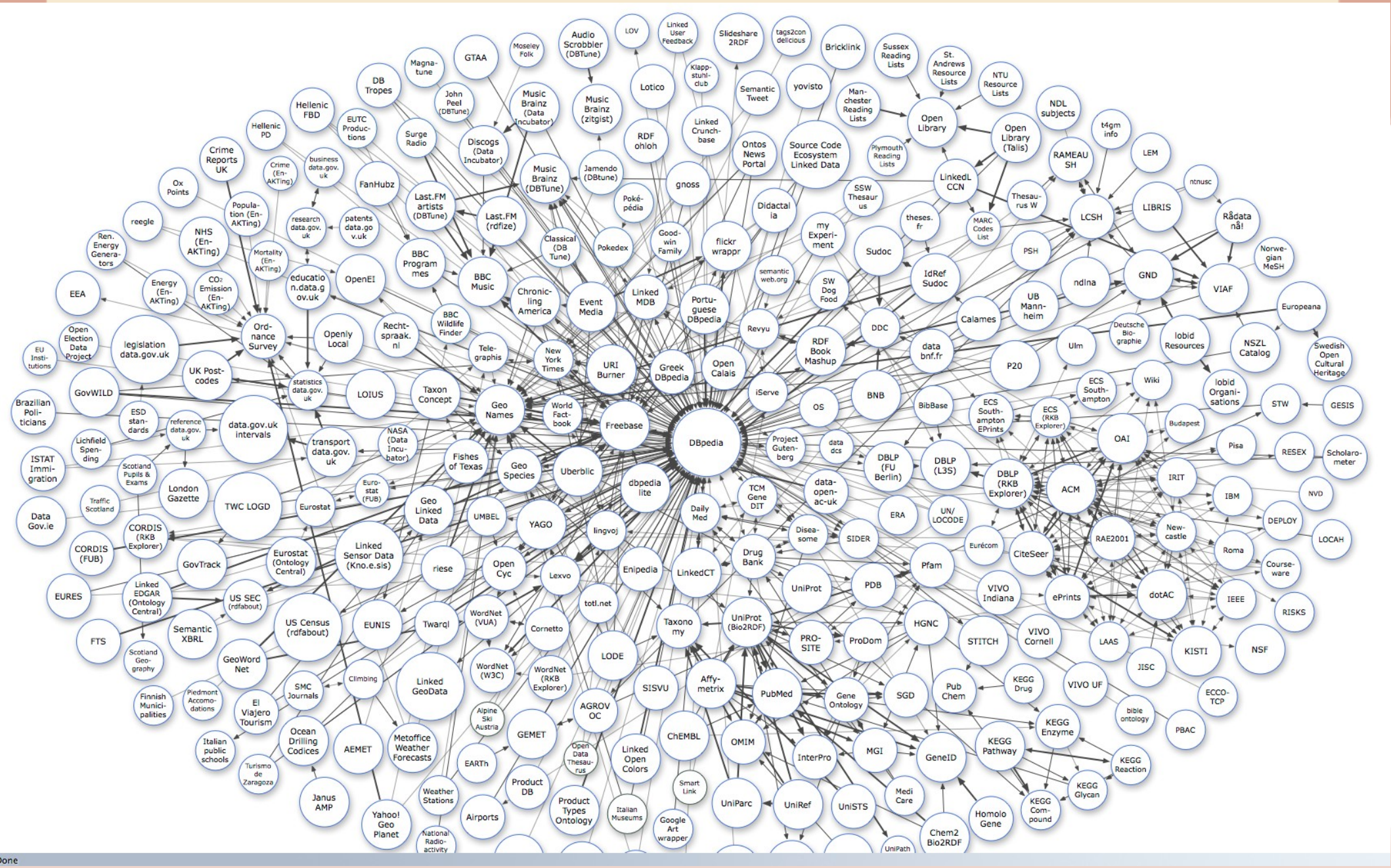
People



Source: Radar Networks & Nova Spivack, 2007 – www.radarnetworks.com

Open Linked Data

- Network of RDF repositories:
  - W3C knowledge representation format
  - 25,200,042,407 triples
- Factual data
- Conceptual data
- Lexical data
- http://richard.cyganiak.de/2007/10/lod/imagemap.html


DBpedia
Freebase
Geo Names
Event Media
Linked MDB
Portuguese DBpedia
BBC Music
Chronicling America
New York Times
URI Burner
Greek DBpedia
Open Calais
Revyu
RDF Book Mashup
iServe
OS
BNB
DDC
data bnf.fr
P20
Calames
UB Mannheim
Sudoc
IdRef Sudoc
PSH
LCSH
GND
VIAF
ndlna
LIBRIS
RAMEAU SH
NDL subjects
Open Library
Open Library (Talis)
LinkedL CCN
Thesaurus W
MARC Codes List
SSW Thesaurus
theses.fr
Didactalia
flickr wrappr
Good-win Family
gnoss
Ontos News Portal
Source Code Ecosystem Linked Data
Semantic Tweet
yovisto
Linked Crunchbase
RDF ohloh
Lotico
Music Brainz (zitgist)
Music Brainz (Data Incubator)
Music Brainz (DBTune)
Jamendo (DBtune)
Poké-pédia
Pokedex
Classical (DB Tune)
Last.FM (rdfize)
Last.FM artists (DBTune)
Discogs (Data Incubator)
Surge Radio
FanHubz
BBC Programmes
BBC Wildlife Finder
Tele-graphis
Recht-spraak.nl
Openly Local
OpenEI
education.data.gov.uk
research data.gov.uk
patents data.gov.uk
business data.gov.uk
Crime (En-AKTing)
Mortality (En-AKTing)
Population (En-AKTing)
CO2 Emission (En-AKTing)
Energy (En-AKTing)
NHS (En-AKTing)
Ren. Energy Generators
Crime Reports UK
Ox Points
reegle
Hellenic PD
Hellenic FBD
EUTC Productions
DB Tropes
Magna-tune
John Peel (DBTune)
GTAA
Moseley Folk
Audio Scrobbler (DBTune)
LOV
Linked User Feedback
Slideshare 2RDF
tags2con delicious
Bricklink
Sussex Reading Lists
St. Andrews Resource Lists
NTU Resource Lists
Man-chester Reading Lists
Plymouth Reading Lists
NDL subjects
t4gm info
LEM
ntnusc
Rådata nå!
Norwe-gian MeSH
Europeana
Swedish Open Cultural Heritage
NSZL Catalog
lobid Resources
lobid Organi-sations
Deutsche Bio-graphie
Ulm
Wiki
ECS South-ampton
ECS South-ampton EPrints
ECS (RKB Explorer)
OAI
Budapest
Pisa
RESEX
Scholaro-meter
STW
GESIS
IRIT
IBM
NVD
DEPLOY
LOCAH
Course-ware
RISKS
IEEE
dotAC
KISTI
NSF
JISC
LAAS
VIVO Cornell
VIVO Indiana
VIVO UF
ePrints
CiteSeer
RAE2001
New-castle
Roma
ACM
DBLP (RKB Explorer)
DBLP (L3S)
DBLP (FU Berlin)
BibBase
Project Guten-berg
data dcs
data-open-ac-uk
TCM Gene DIT
dbpedia lite
Uberblic
Geo Species
Fishes of Texas
NASA (Data Incu-bator)
transport data.gov.uk
statistics data.gov.uk
LOIUS
Taxon Concept
World Fact-book
Ordnance Survey
legislation data.gov.uk
UK Post-codes
data.gov.uk intervals
reference data.gov.uk
ESD stan-dards
Scotland Pupils & Exams
London Gazette
TWC LOGD
Eurostat
Euro-stat (FUB)
Geo Linked Data
UMBEL
YAGO
lingvoj
Daily Med
Disea-some
SIDER
ERA
UN/ LOCODE
Eurécom
Pfam
PDB
UniProt
Drug Bank
LinkedCT
Enipedia
Open Cyc
Lexvo
totl.net
Taxono-my
UniProt (Bio2RDF)
PRO-SITE
ProDom
HGNC
STITCH
PubMed
Gene Ontology
SGD
Pub Chem
KEGG Drug
KEGG Enzyme
KEGG Pathway
KEGG Reaction
KEGG Glycan
KEGG Com-pound
GeneID
MGI
InterPro
OMIM
ChEMBL
Affy-metrix
SISVU
LODE
AGROVOC
GEMET
EARTh
Alpine Ski Austria
WordNet (RKB Explorer)
WordNet (W3C)
WordNet (VUA)
Cornetto
Twarql
EUNIS
US Census (rdfabout)
US SEC (rdfabout)
Linked Sensor Data (Kno.e.sis)
Eurostat (Ontology Central)
GovTrack
Linked EDGAR (Ontology Central)
Semantic XBRL
CORDIS (RKB Explorer)
CORDIS (FUB)
Traffic Scotland
Lichfield Spen-ding
ISTAT Immi-gration
Data Gov.ie
EURES
FTS
Scotland Geo-graphy
Finnish Munici-palities
Piedmont Accomo-dations
Italian public schools
Turismo de Zaragoza
GeoWord Net
El Viajero Tourism
SMC Journals
Climbing
Ocean Drilling Codices
Janus AMP
AEMET
Metoffice Weather Forecasts
Linked GeoData
Weather Stations
National Radio-activity
Yahoo! Geo Planet
Airports
Product DB
Product Types Ontology
Open Data Thesau-rus
Linked Open Colors
Italian Museums
Smart Link
Google Art wrapper
UniParc
UniRef
UniSTS
UniPath
Chem2 Bio2RDF
Homolo Gene
Medi Care
bible ontology
PBAC
ECCO-TCP
Brazilian Poli-ticians
EU Insti-tutions
Open Election Data Project
GovWILD
EEA
Done

DBPedia

- 3.64 million things, out of which 1.83 million are classified in a consistent Ontology, including 416,000 persons, 526,000 places, 106,000 music albums, 60,000 films, 17,500 video games, 169,000 organisations, 183,000 species and 5,400 diseases.
- The DBpedia data set features labels and abstracts for these 3.64 million things in up to 97 different languages; 2,724,000 links to images and 6,300,000 links to external web pages; 6,200,000 external links into other RDF datasets, 740,000 Wikipedia categories, and 2,900,000 YAGO categories. The dataset consists of 1 billion pieces of information (RDF triples) out of which 385 million were extracted from the English edition of Wikipedia and roughly 665 million were extracted from other language editions and links to external datasets.
- The DBpedia knowledge base allows you to ask quite surprising queries against Wikipedia, for instance “Give me all cities in New Jersey with more than 10,000 inhabitants” or “Give me all Italian musicians from the 18th century”.

About: Bled

An Entity of Type : Municipalities of Slovenia, from Named Graph : http://dbpedia.org, within Data Space : dbpedia.org

dbpedia-owl:areaTotal	188500000.000000 (xsd:double)
dbpedia-owl:country	dbpedia:Slovenia
dbpedia-owl:leaderName	dbpedia:Slovenian_National_Party
dbpedia-owl:leaderTitle	Mayor
dbpedia-owl:populationTotal	10899 (xsd:integer)
dbpedia-owl:thumbnail	http://upload.wikimedia.org/wikipedia/commons/thu
dbpedia-owl:timeZone	dbpedia:Central_European_Time dbpedia:Central_European_Summer_Time
dbpedia-owl:type	dbpedia:Municipalities_of_Slovenia
dbpedia-owl:utcOffset	+01 +02
dbpedia-owl:wikiPageExternalLink	http://www.geopedia.si/Geopedia_en.html#L5567_ http://www.blejskiotok.si http://www.bled.si http://wikitravel.org/en/Bled http://www.bled-castle.si
dbpprop:areaTotalKm	189 (xsd:integer)
dbpprop:coordinatesDisplay	title
dbpprop:coordinatesRegion	SI
dbpprop:coordinatesType	1.0
dbpprop:imageCaption	Bled: lake, castle and parish church
dbpprop:imageFlag	BledMuniFlag.gif
dbpprop:imageMap	Obcine Slovenija 2007 Bled.svg
dbpprop:imageShield	Bled Coat of Arms.png
dbpprop:imageSkyline	Bledi-tó.jpg
dbpprop:latd	46 (xsd:integer)
dbpprop:latm	22 (xsd:integer)
dbpprop:latns	N
dbpprop:leaderName	Janez Fajfar
dbpprop:leaderTitle	Mayor
dbpprop:mapCaption	Location of the Municipality of Bled in Slovenia
dbpprop:name	Bled
dbpprop:nativeNameLang	sl
dbpprop:populationAsOf	2002 (xsd:integer)
dbpprop:populationDensityKm	auto
dbpprop:populationTotal	10899 (xsd:integer)
dbpprop:pushpinLabelPosition	left
dbpprop:pushpinMap	Slovenia
dbpprop:pushpinMapCaption	Location of the Town of Bled in Slovenia
dbpprop:settlementType	Town and Municipality
dbpprop:subdivisionType	dbpedia:List_of_sovereign_states
dbpprop:timezone	dbpedia:Central_European_Time
dbpprop:timezone1Dst	dbpedia:Central_European_Summer_Time
dbpprop:unitPref	Metric
dbpprop:utcOffset	+01
dbpprop:utcOffset1Dst	+02
dbpprop:wikiPageUsesTemplate	dbpedia:Template:Infobox_settlement
dcterms:subject	category:Municipalities_of_Slovenia category:Bled category:Populated_places_in_the_Municipality_of_Bled
grs:point	46.36666666666667 14.116666666666667
rdf:type	owl:Thing yago:Municipality108626283 dbpedia-owl:PopulatedPlace dbpedia-owl:Settlement dbpedia-owl:Place dbpedia-owl:Town http://schema.org/Place gml:_Feature yago:MunicipalitiesOfSlovenia yago:GeoclassPopulatedPlace

About: Bled

An Entity of Type : Municipalities of Slovenia, from Named Graph : http://dbpedia.org, within Data Space : dbpedia.org

geo:geometry	▪ POINT(14.1167 46.3667)
geo:lat	▪ 46.366665 (xsd:float)
geo:long	▪ 14.116667 (xsd:float)
foaf:depiction	▪ http://upload.wikimedia.org/wikipedia/commons/f/f9/Bledi-t%C3%B
foaf:homepage	▪ http://www.bled.si
foaf:name	▪ Bled
foaf:page	▪ http://en.wikipedia.org/wiki/Bled
is dbpedia-owl:birthPlace of	▪ dbpedia:Prince_Andrew_of_Yugoslavia ▪ dbpedia:Klemen_Pretnar ▪ dbpedia:Sara_Isaković ▪ dbpedia:Peter_Florjančič ▪ dbpedia:Aleksandar_Obradović ▪ dbpedia:Špela_Pretnar ▪ dbpedia:Klemen_Mohorič
is dbpedia-owl:city of	▪ dbpedia:IEDC-Bled_School_of_Management ▪ dbpedia:Lake_Bled
is dbpedia-owl:deathPlace of	▪ dbpedia:Julius_von_Payer ▪ dbpedia:Duchess_Marie_Antoinette_of_Mecklenburg
is dbpedia-owl:isPartOf of	▪ dbpedia:Bodešče ▪ dbpedia:Koritno,_Bled ▪ dbpedia:Selo_pri_Bledu ▪ dbpedia:Kupljenik ▪ dbpedia:Ribno ▪ dbpedia:Zasip ▪ dbpedia:Obrne ▪ dbpedia:Bohinjska_Bela ▪ dbpedia:Slamniki
is dbpedia-owl:location of	▪ dbpedia:Bled_Castle
is dbpedia-owl:nearestCity of	▪ dbpedia:Vogel_Ski_Resort ▪ dbpedia:Kobla_Ski_Resort
is dbpedia-owl:wikiPageRedirects of	▪ dbpedia:Bled,_Slovenia ▪ dbpedia:Veldes
is dbpprop:basedIn of	▪ dbpedia:HK_MK_Bled

What about lexical knowledge?

- Textual information will become referential knowledge:
  - Planning software for emails & calendars:
    - Dates, places, people
    - Events such as meetings, conferences, etc.
  - Biographies and bibliographical references
- Unstructured text is converted to structured knowledge through text-mining and text-understanding
- Who will be the owner of the meaning of words, representing the de-facto-standard?

Summary

- RDF repositories of knowledge → understandable by machines
- Created manually by experts, automatically and ***socially***
- Lexical knowledge is also knowledge
- Persistent identifiers (PID, URI) to make reference to knowledge objects
- Unstructured text converted to structured knowledge
- Every word becomes a hyperlink to formally structured knowledge by adding the lemma or concept PID to each word
- Either you establish the PID or, if the PID turns out to point to Wikipedia, you link your PID to theirs
- Applications (Apps) will be developed on top of structured knowledge, so you make sure you provide lexical knowledge with added value that is useful for computers