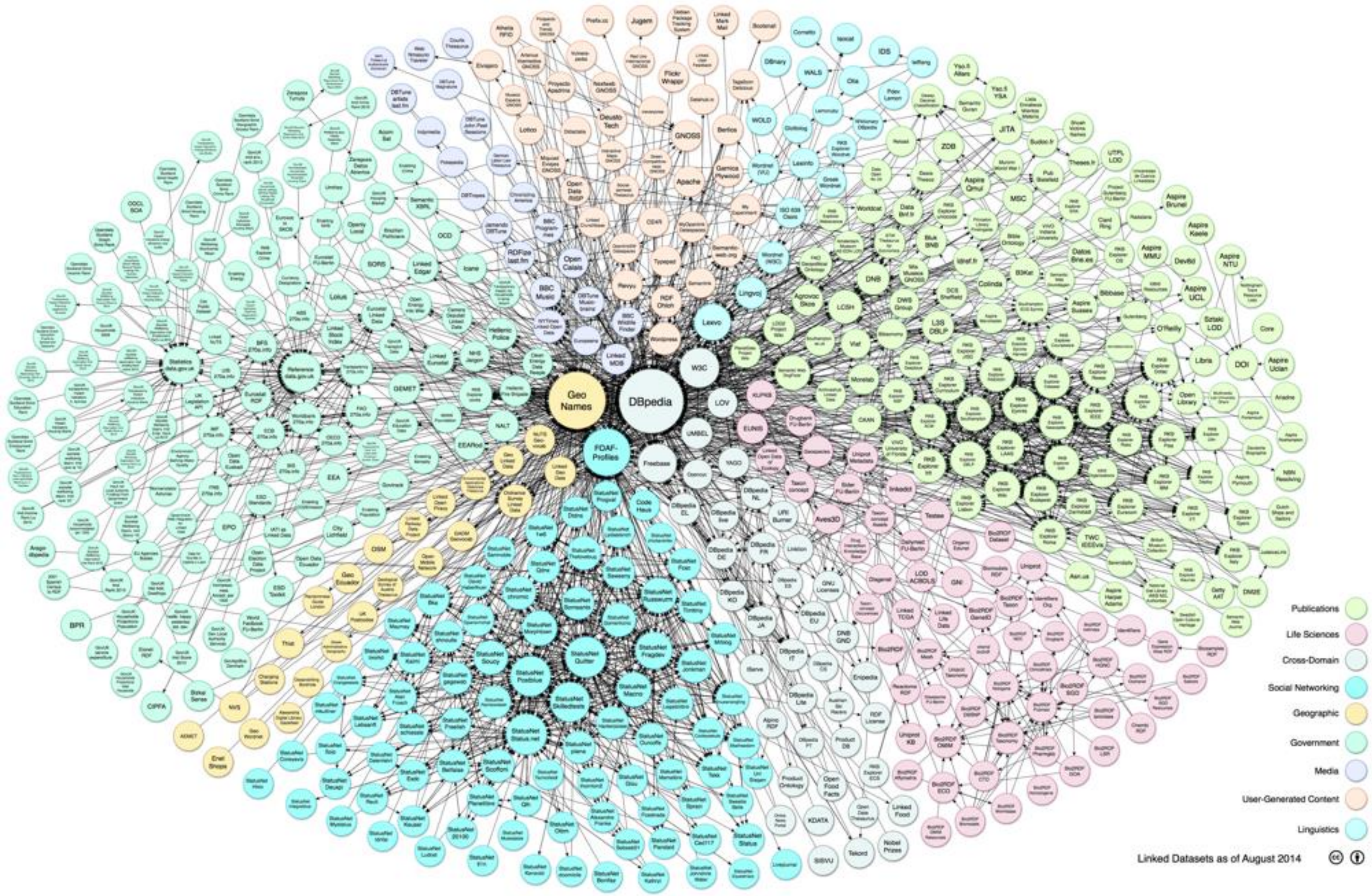
Zhishi.lemon: On Publishing Zhishi.me as Linguistic Linked Open Data

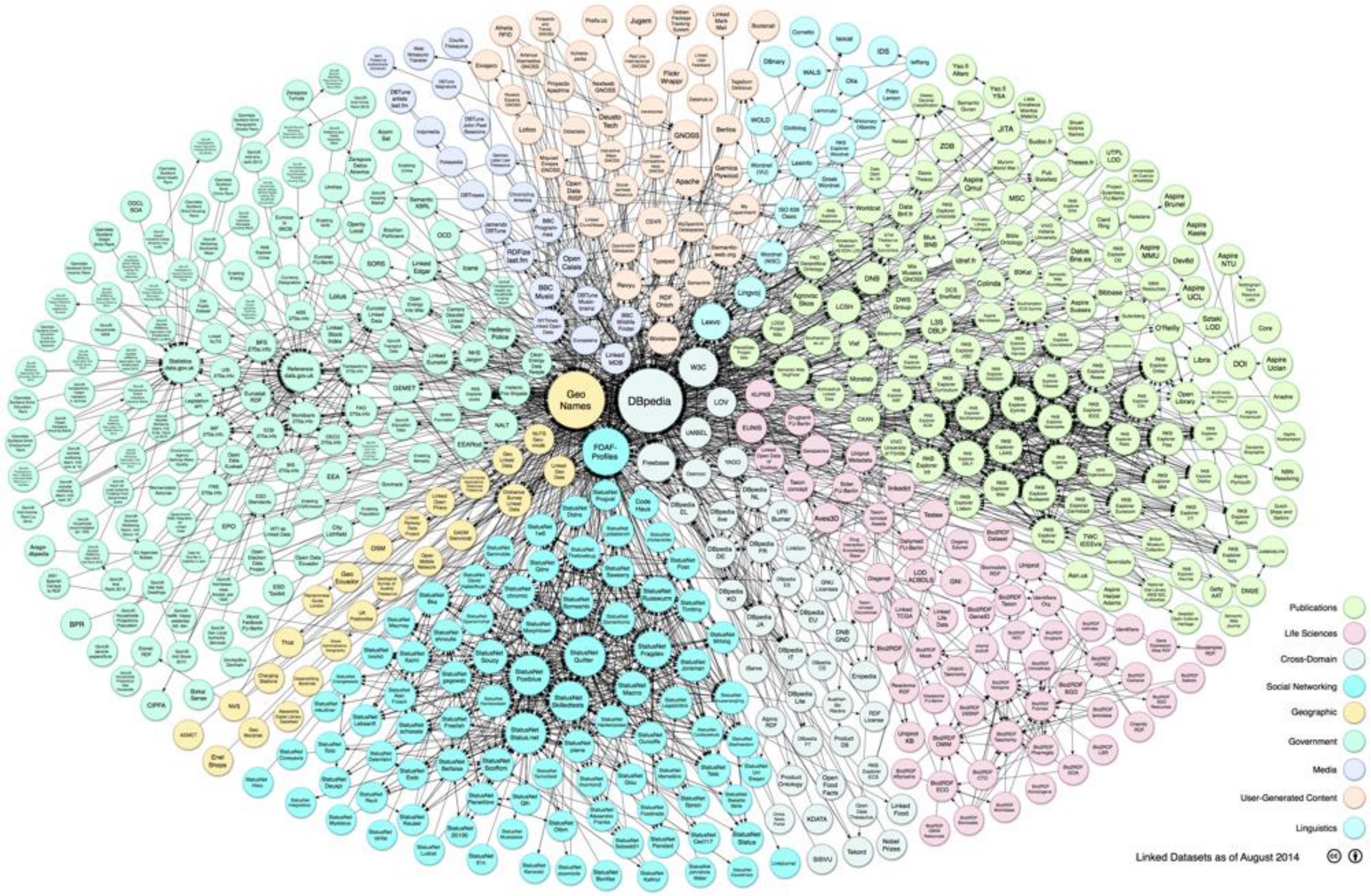
Zhijia Fang[1], Haofen Wang[1],
Jorge Gracia[2], Julia Bosque-Gil[2] and Tong Ruan[1]

[1] East China University of Science and Technology
[2] Ontology Engineering Group, Universidad Politécnica de Madrid



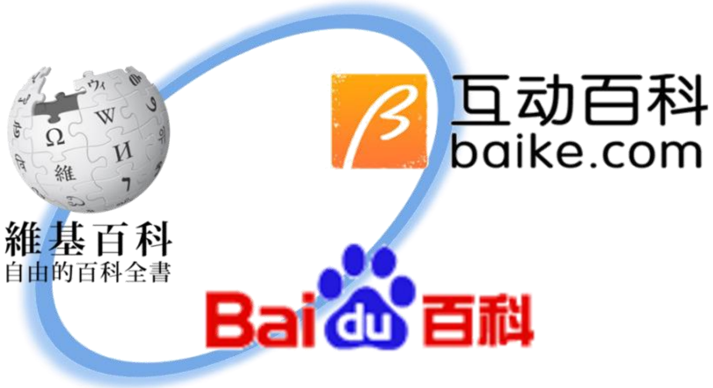
Linked Datasets as of August 2014

Linked Open Data

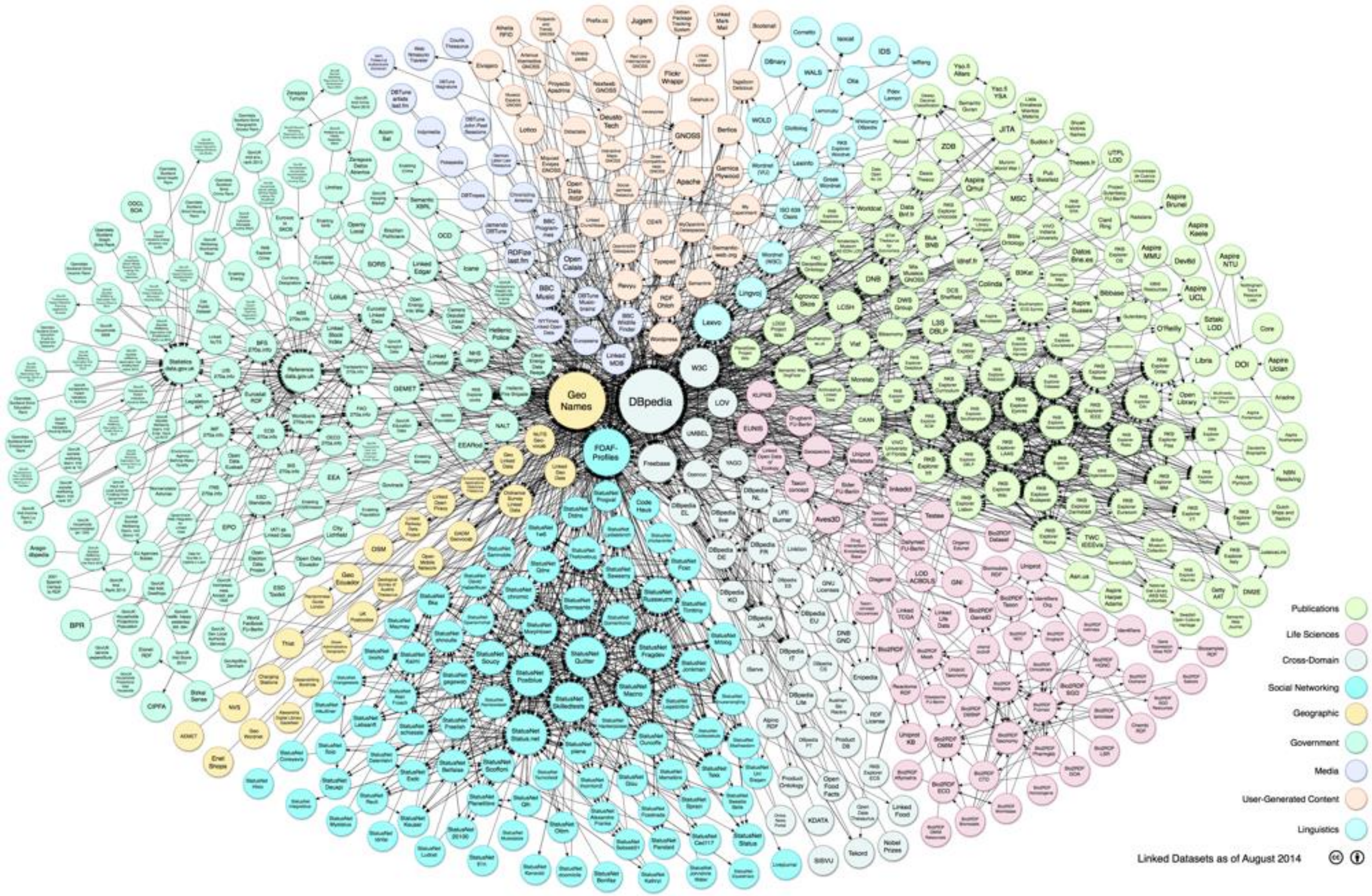


Zhishi.me

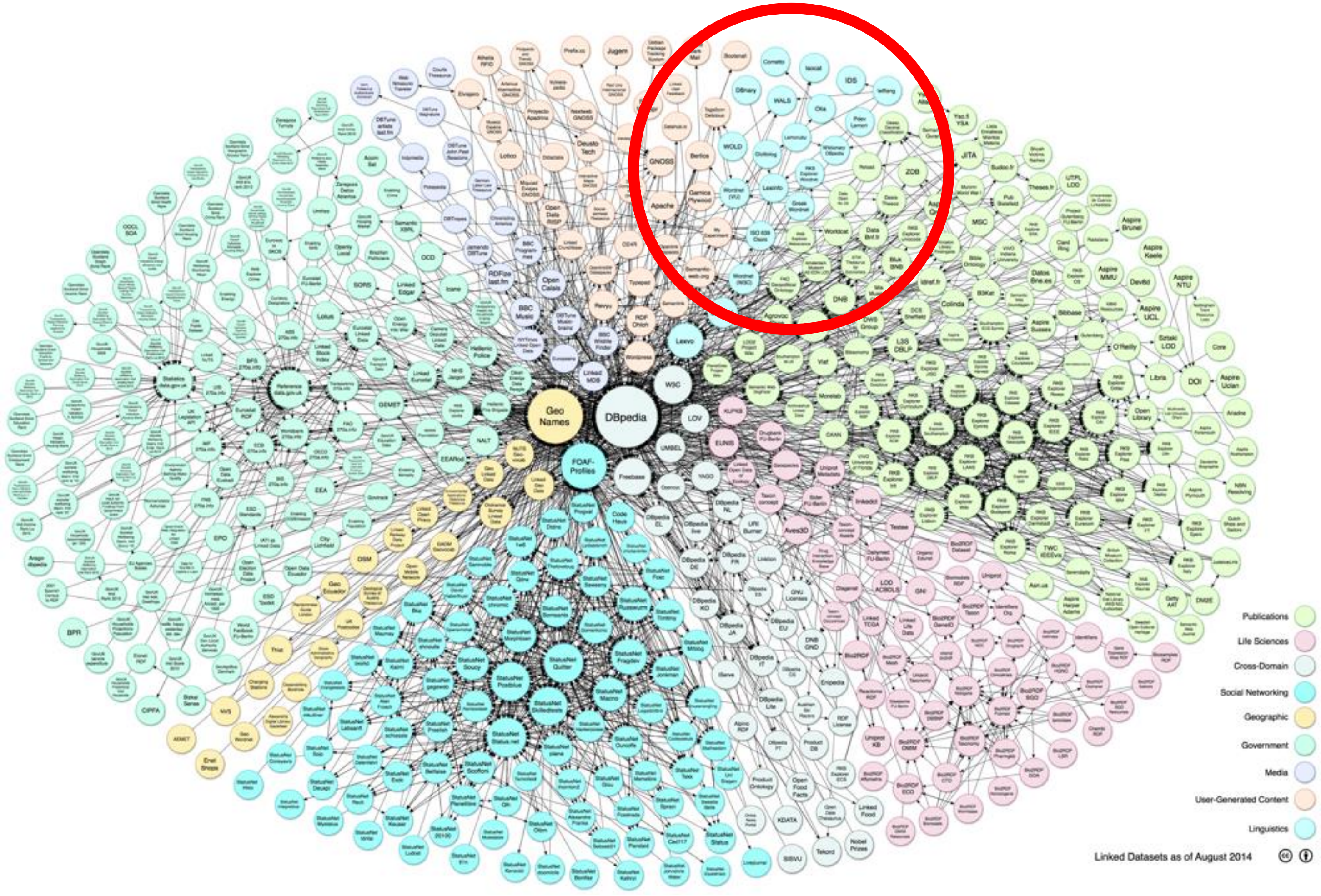
- Zhishi.me (http://zhishi.me) is the first effort to publish large scale Chinese semantic data and link them together as a Chinese Linked Open Data (CLOD).



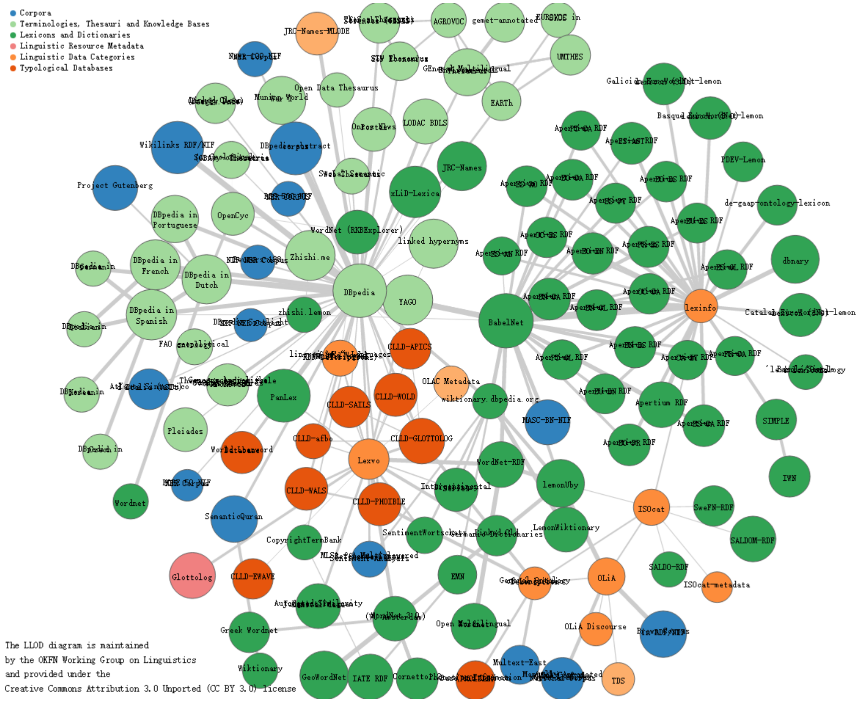
It has over 8 million distinct instances and 200 million RDF triples.


Geo Names
DBpedia
FOAF-Profiles
Freebase
W3C
LOV
UMBEL
YAGO
Lexvo
Statistics data.gov.uk
Reference data.gov.uk
StatusNet Status.net
Publications
Life Sciences
Cross-Domain
Social Networking
Geographic
Government
Media
User-Generated Content
Linguistics
Linked Datasets as of August 2014

Linguistic Linked Open Data Cloud



Linked Datasets as of August 2014



The LLOD diagram is maintained
by the OKFN Working Group on Linguistics
and provided under the
Creative Commons Attribution 3.0 Unported (CC BY 3.0) license

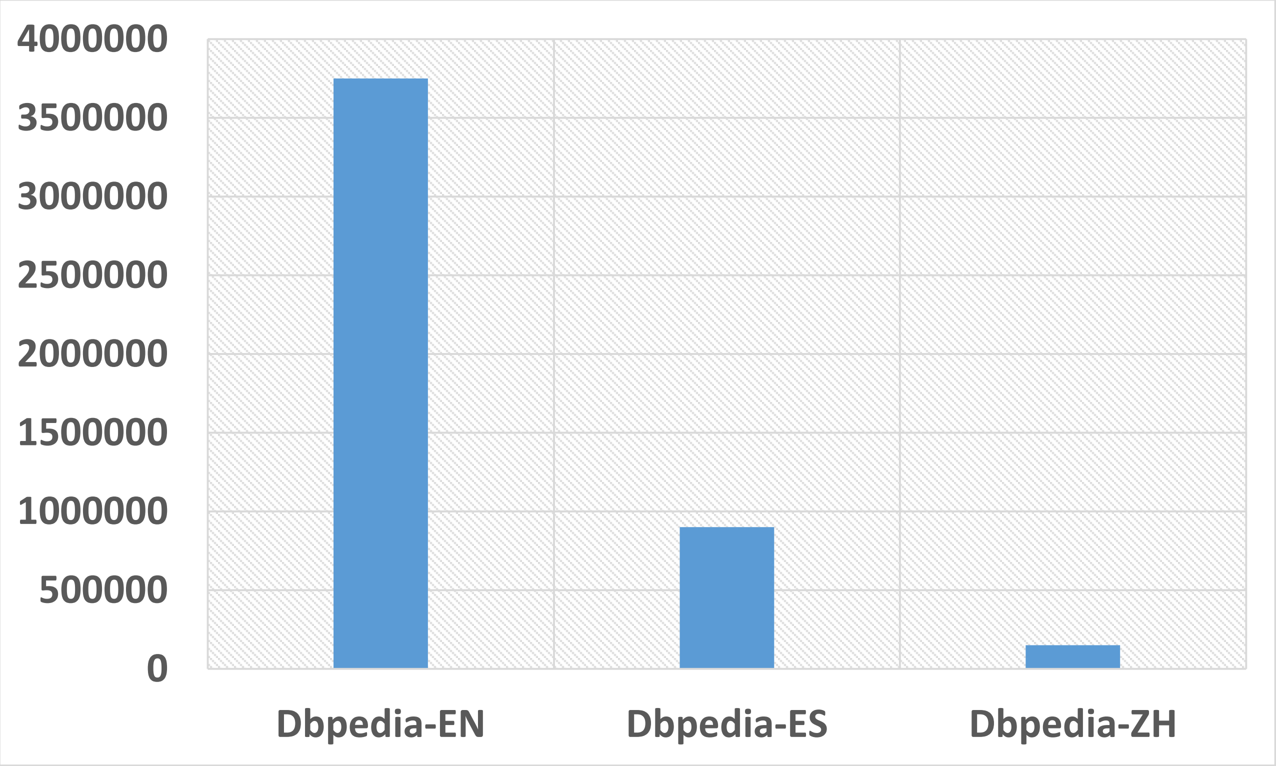
Zhishi.lemon

- A newly developed dataset based on the lemon model that constitutes the **lexical realization of Zhishi.me**.
- Zhishi.lemon combines the lemon core with the lemon translation module in order to build a linked data lexicon in Chinese with translations into **Spanish and English**.
- Links to **BabelNet and Dbpedia** have been provided as well.

Why we build Zhishi.lemon?

- Capacity

Compared to English and other prevalent languages in the LLOD cloud, resources in Chinese are scarce.



Why we build Zhishi.lemon?

- Linkability

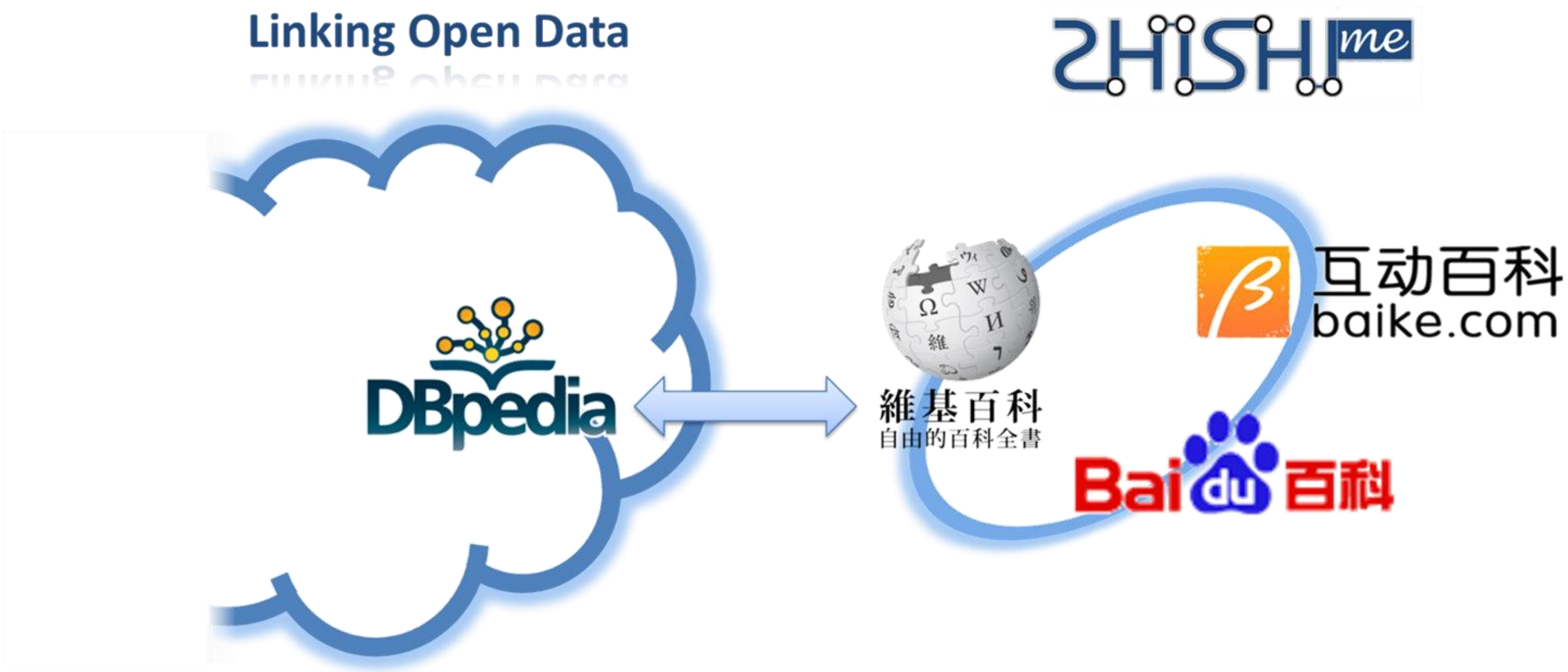
Linking Zhishi.me to the LLOD cloud can further enrich the whole cloud with additional Chinese entries.

- Usability

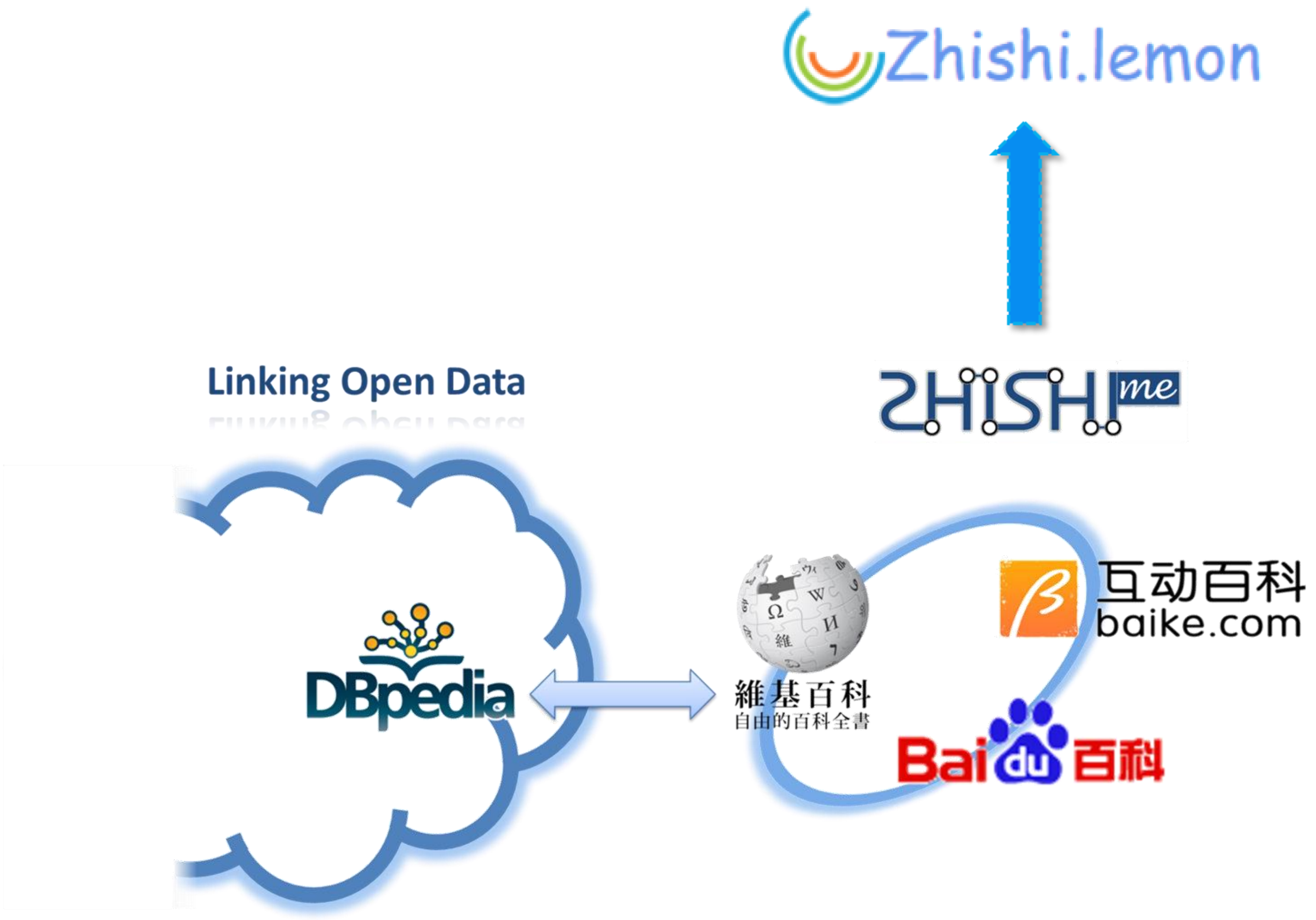
Benefit both academia and industry to better understand Chinese and to further build related applications.

Linking Zhishi.me to DBpedia and BabelNet

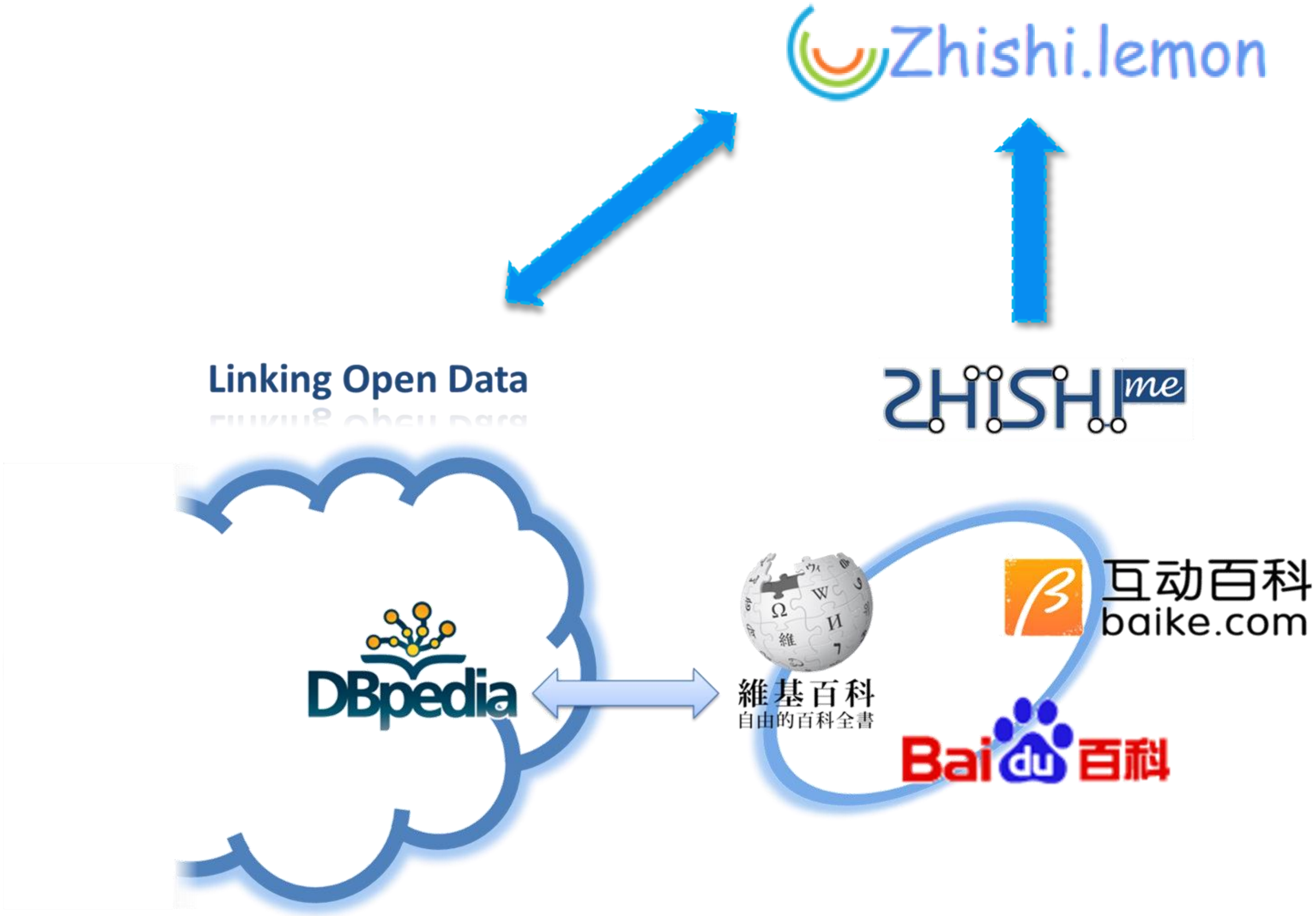
Linking Zhishi.me to DBpedia and BabelNet



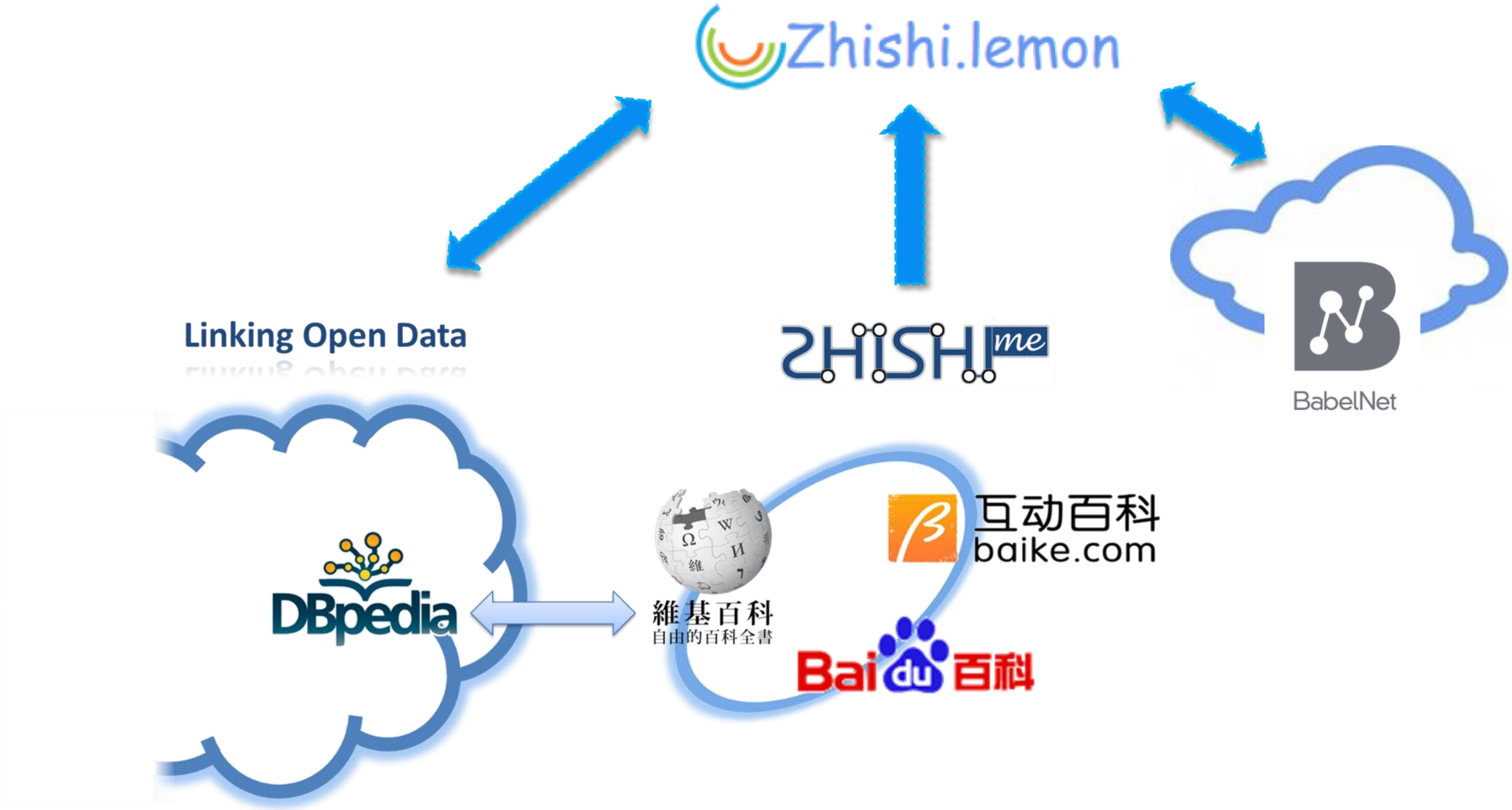
Linking Zhishi.me to DBpedia and BabelNet



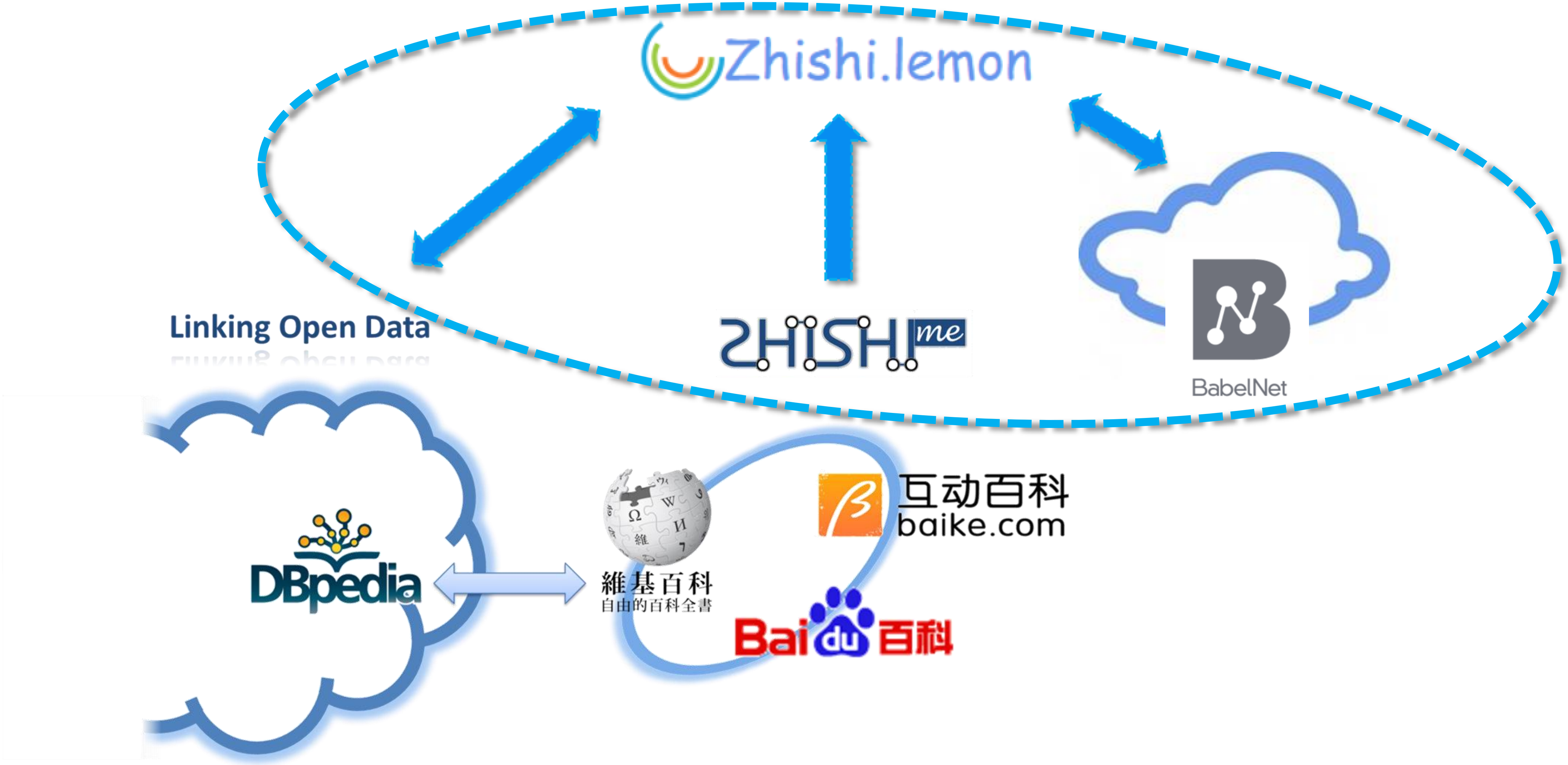
Linking Zhishi.me to DBpedia and BabelNet



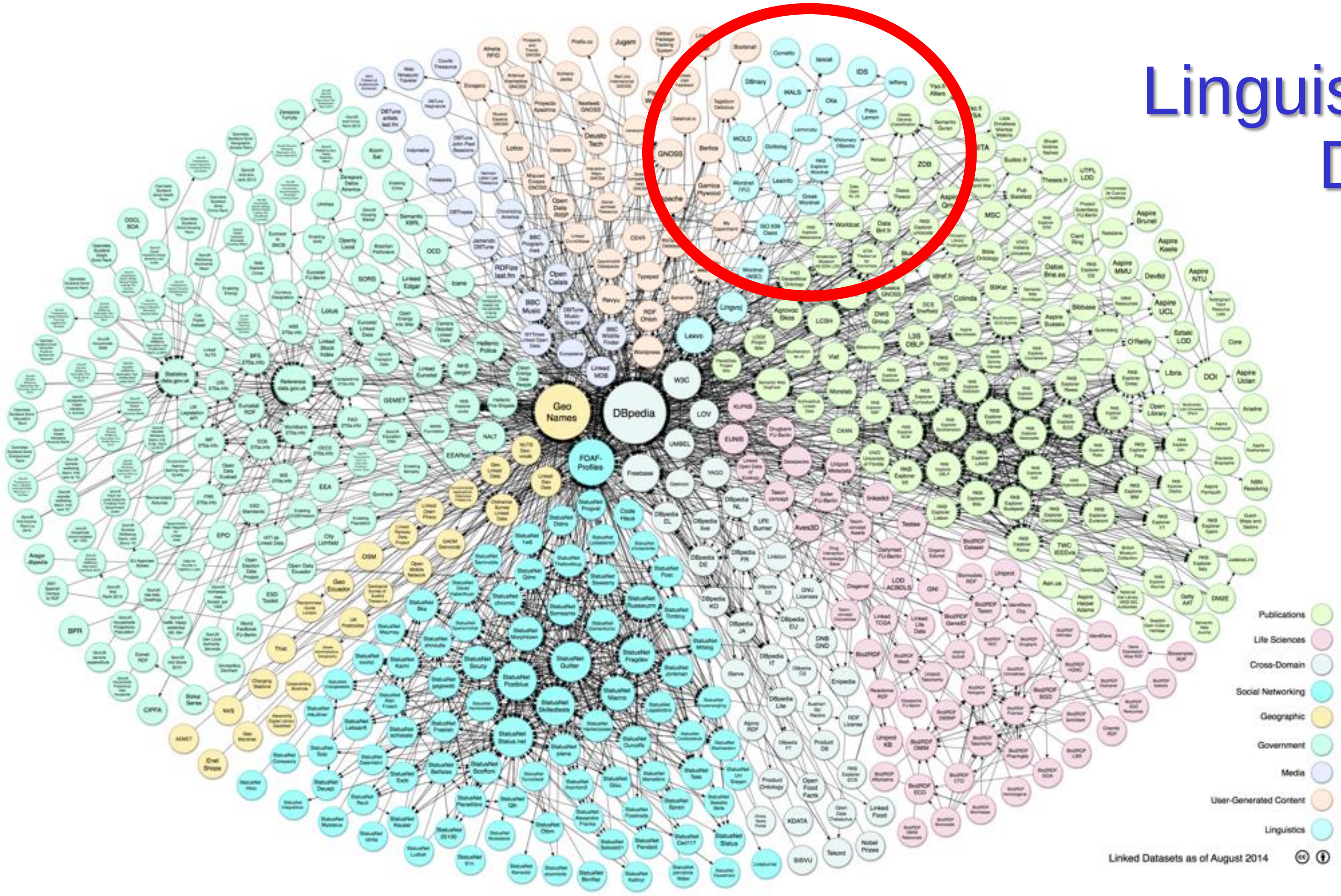
Linking Zhishi.me to DBpedia and BabelNet

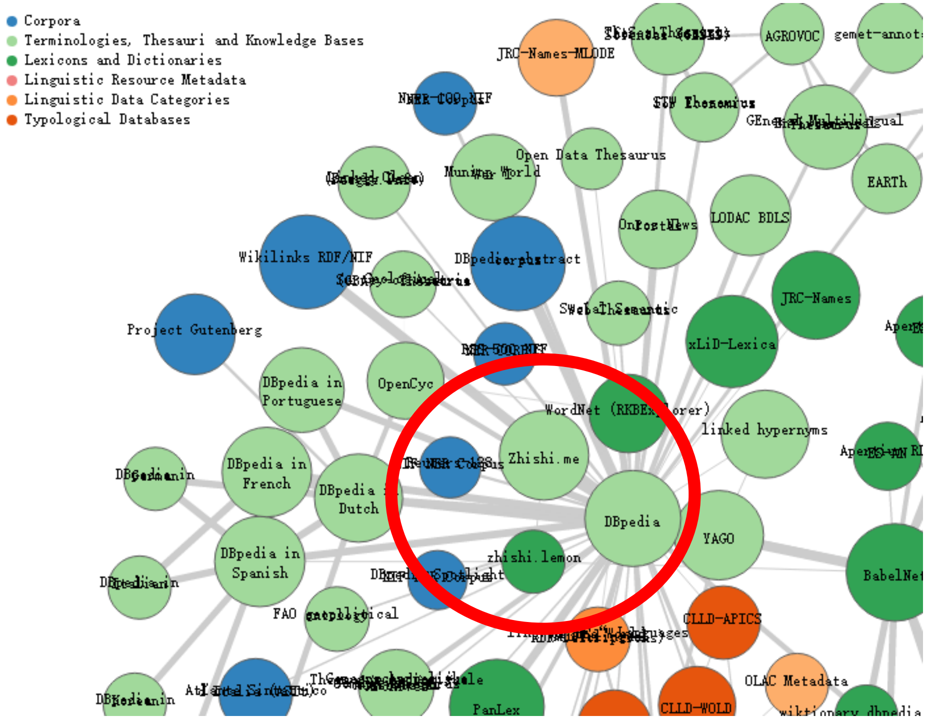


Linking Zhishi.me to DBpedia and BabelNet



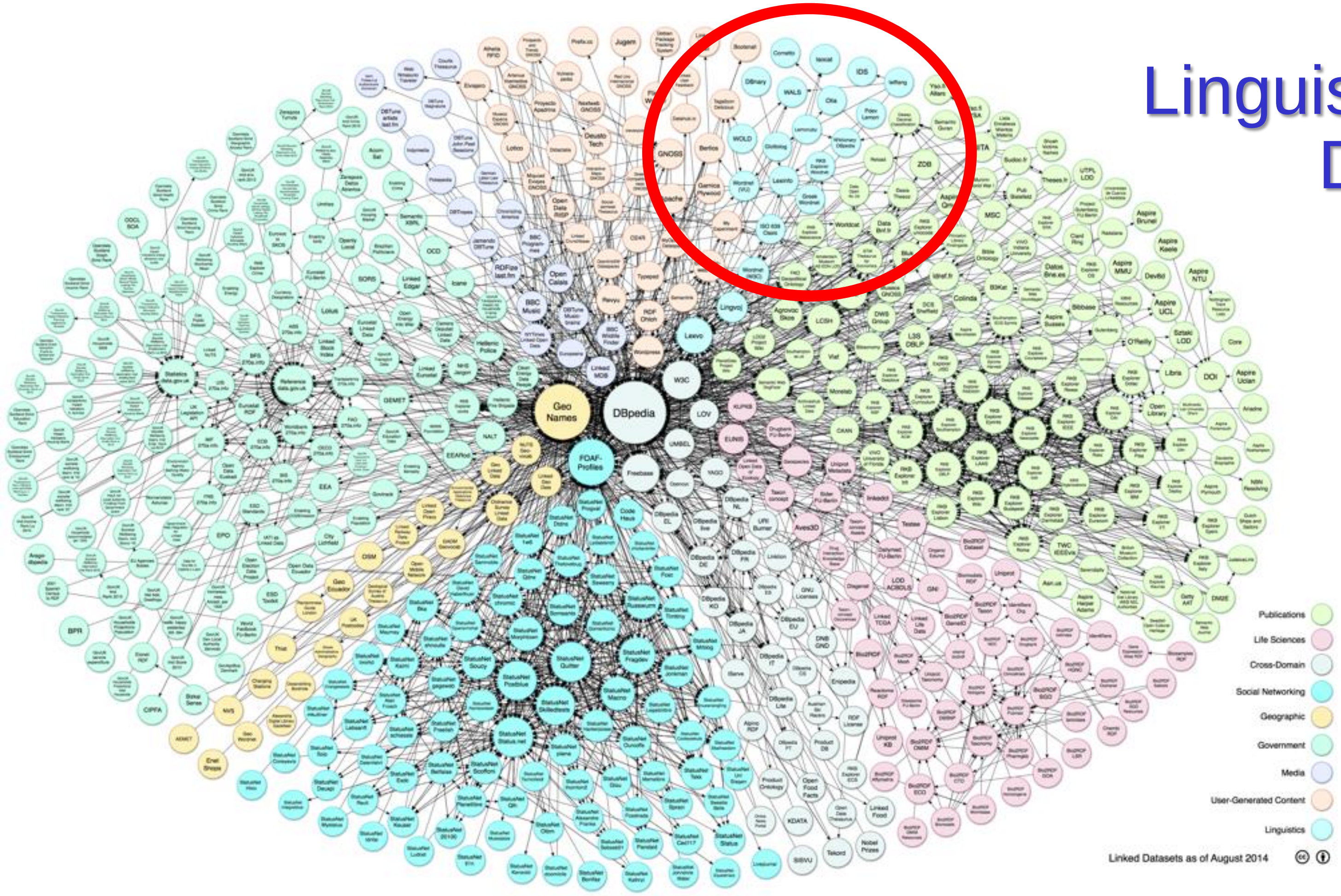
Linguistic Linked Open Data Cloud

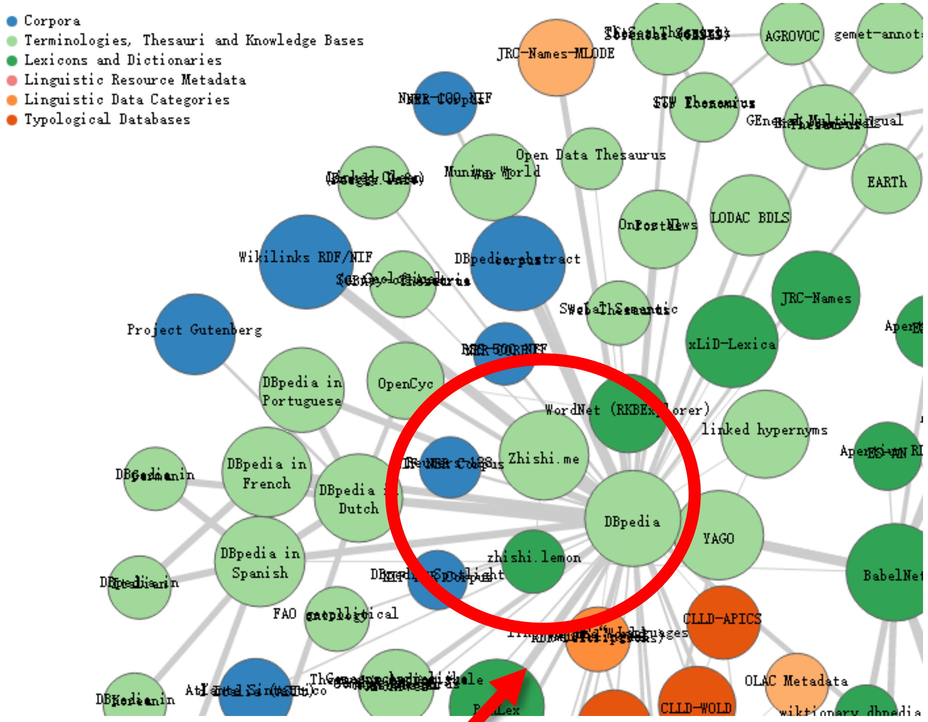




Zhishi.lemon and Zhishi.me

Linguistic Linked Open Data Cloud





How to represent the data?

Zhishi.lemon and Zhishi.me

lemon

LExicon Model for ONtologies

Why use lemon?

lemon

LExicon Model for ONtologies

Why use lemon?

- Bridge the gap between lexical and conceptual information
- De-facto standard for representing and publishing lexical resources as linked data on the Web

lemon

LExicon Model for ONtologies

Why use lemon?

- Bridge the gap between lexical and conceptual information
- De-facto standard for representing and publishing lexical resources as linked data on the Web

Who have used lemon?

lemon

LExicon Model for ONtologies

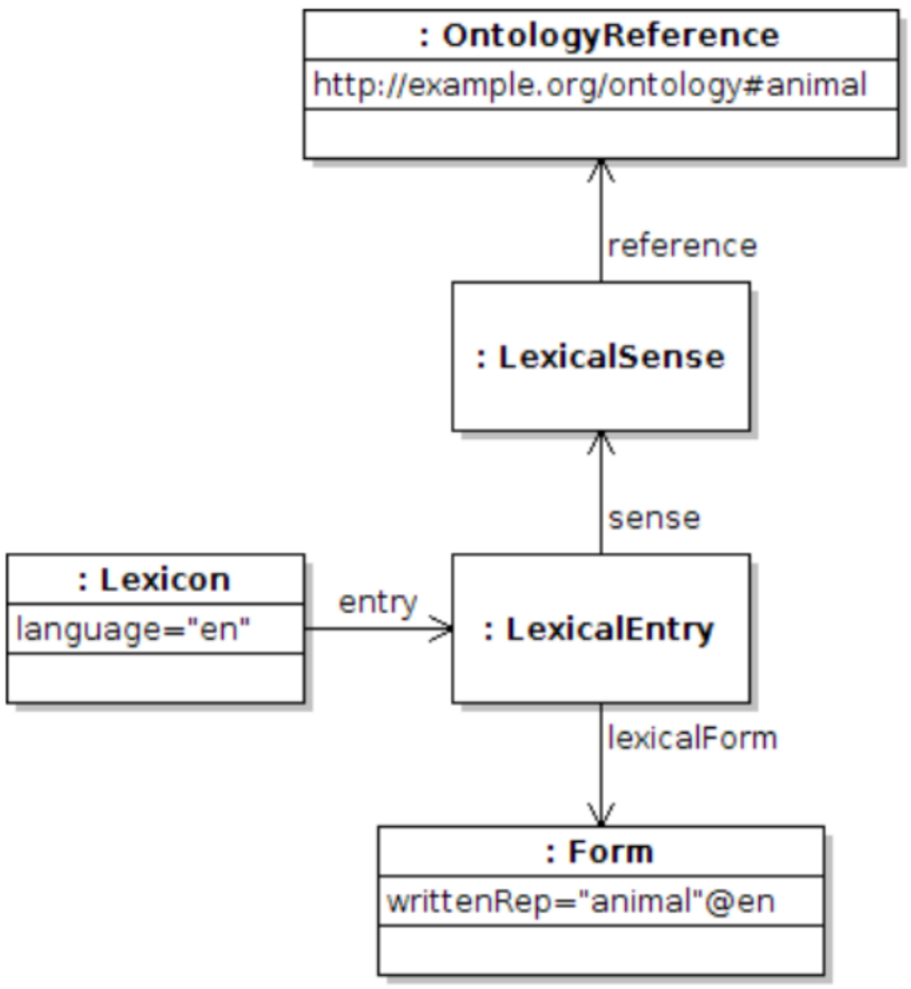
Why use lemon?

- Bridge the gap between lexical and conceptual information
- De-facto standard for representing and publishing lexical resources as linked data on the Web

- Apertium
- WordNet
- Dbpedia Wikitonary
- BabelNet
- ……

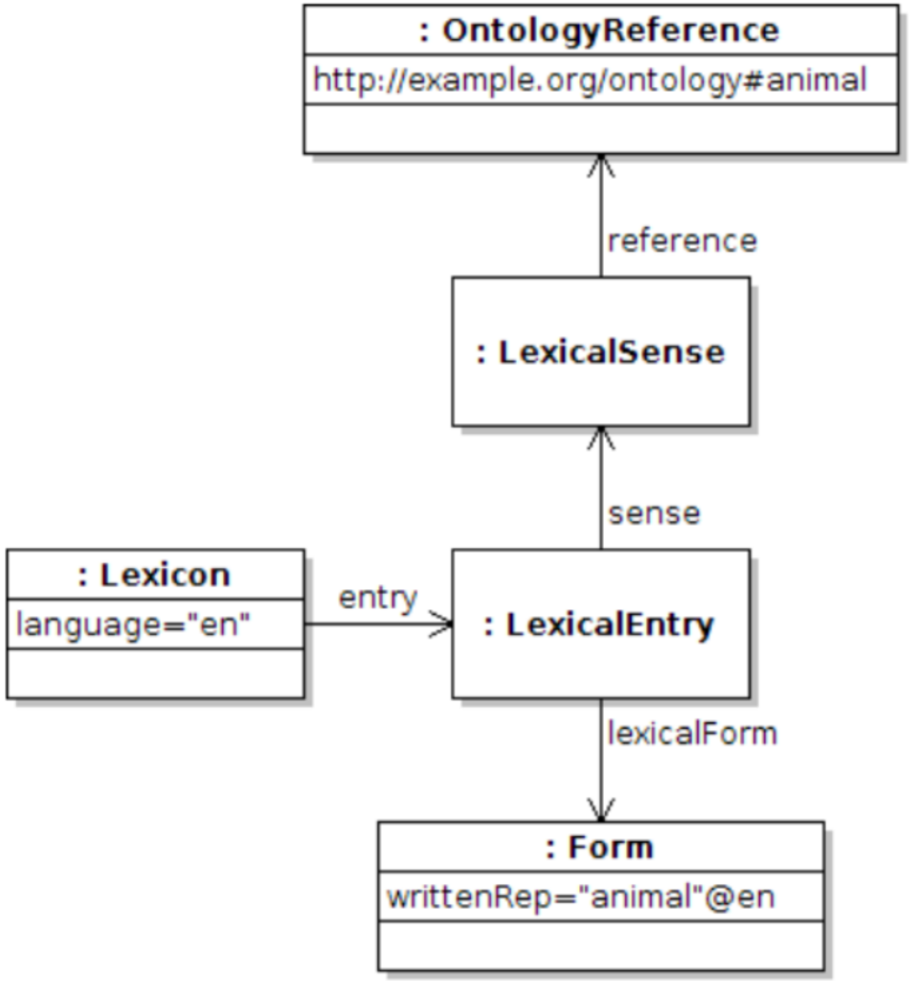
Who have used lemon?

The core path



- **Lexicon**: Represents the lexicon. Marked with a single language tag (ISO-639)
- **Lexical Entry**: An entry in a lexicon. Syntax-invariant
- **Lexical Sense**: The relationship between the entry and its ontology reference.
- **Reference**: The ontology entity
- **Form**: A form of an entry. Orthography-invariant
- **Representation**: The string. IETF lang-tagged

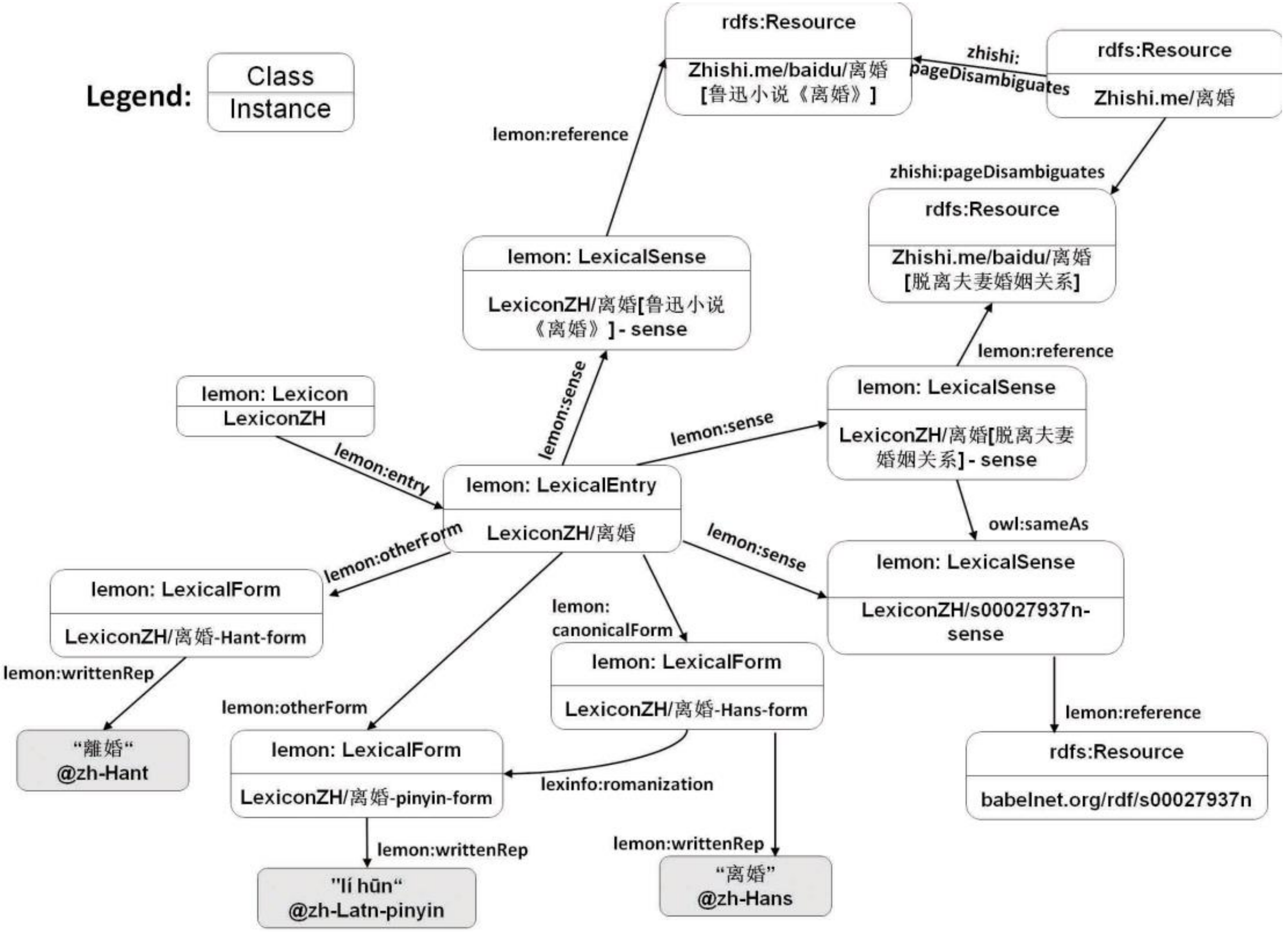
The core path



- **Lexicon**: Represents the lexicon. Marked with a single language tag (ISO-639)
- **Lexical Entry**: An entry in a lexicon. Syntax-invariant
- **Lexical Sense**: The relationship between the entry and its ontology reference.
- **Reference**: The ontology entity
- **Form**: A form of an entry. Orthography-invariant
- **Representation**: The string. IETF lang-tagged

Ontology Overview

Chinese Lexicalization Module



Legend:

Class

Instance

rdfs:Resource

Zhishi.me/baidu/离婚
[鲁迅小说《离婚》]

zhishi:
pageDisambiguates

rdfs:Resource

Zhishi.me/离婚

lemon:reference

zhishi:pageDisambiguates

rdfs:Resource

Zhishi.me/baidu/离婚
[脱离夫妻婚姻关系]

lemon: LexicalSense

LexiconZH/离婚[鲁迅小说
《离婚》]- sense

lemon:reference

lemon: LexicalSense

LexiconZH/离婚[脱离夫妻
婚姻关系]- sense

lemon: Lexicon
LexiconZH

lemon:sense

lemon:sense

lemon:entry

lemon: LexicalEntry

LexiconZH/离婚

owl:sameAs

lemon:otherForm

lemon:sense

lemon: LexicalSense

LexiconZH/s00027937n-
sense

lemon: LexicalForm

LexiconZH/离婚-Hant-form

lemon:
canonicalForm

lemon: LexicalForm

LexiconZH/离婚-Hans-form

lemon:writtenRep

lemon:otherForm

lemon:reference

"離婚"
@zh-Hant

lemon: LexicalForm

LexiconZH/离婚-pinyin-form

lexinfo:romanization

rdfs:Resource

babelnet.org/rdf/s00027937n

lemon:writtenRep

lemon:writtenRep

"lí hūn"
@zh-Latn-pinyin

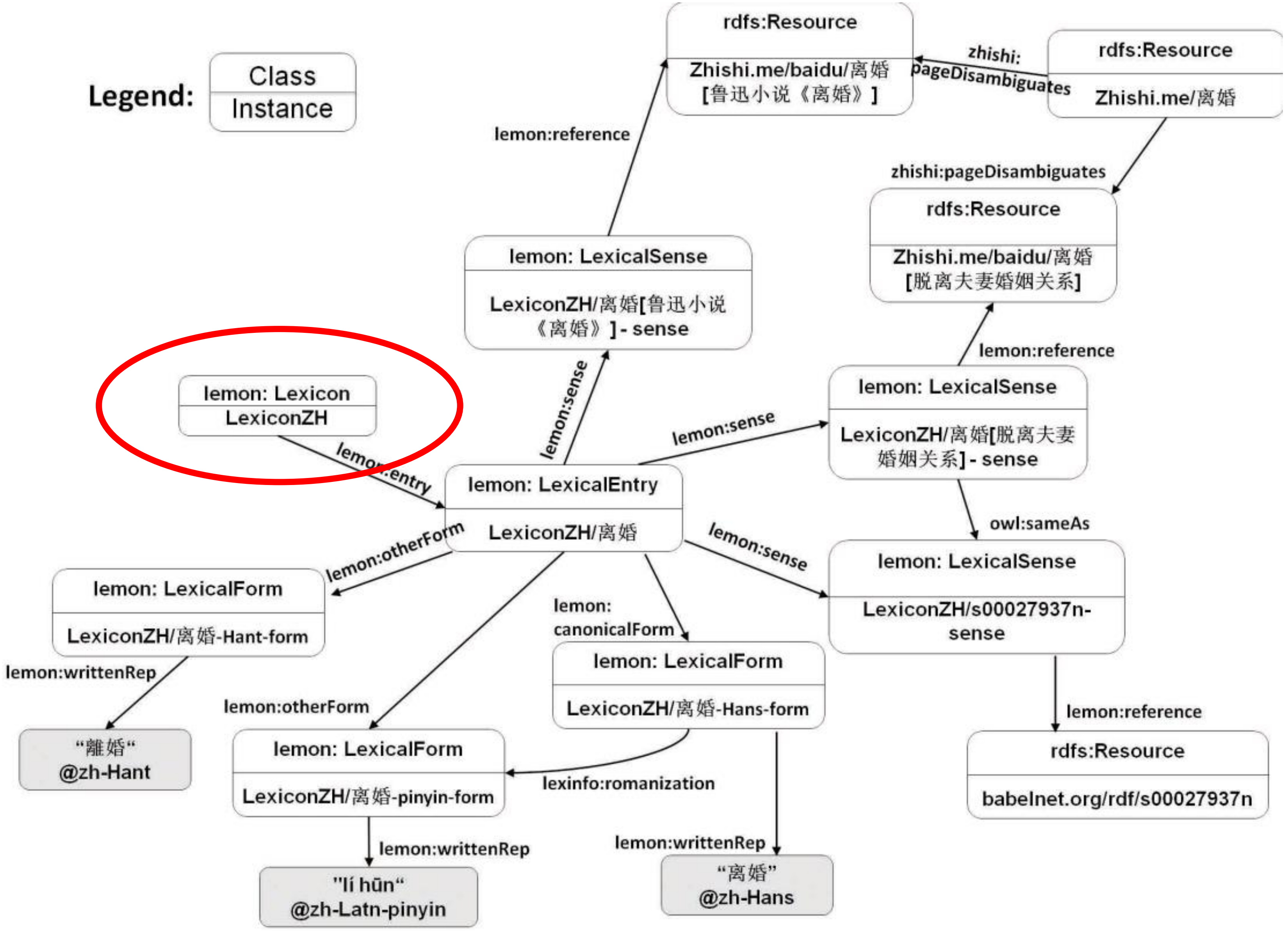
"离婚"
@zh-Hans

Chinese Lexicalization Module

Lexicon

Three languages:

- LexiconZH
- LexiconEN
- LexiconES



Legend: Class
Instance

rdfs:Resource
Zhishi.me/baidu/离婚
[鲁迅小说《离婚》]

zhishi:
pageDisambiguates

rdfs:Resource
Zhishi.me/离婚

zhishi:pageDisambiguates

rdfs:Resource
Zhishi.me/baidu/离婚
[脱离夫妻婚姻关系]

lemon:reference

lemon: LexicalSense
LexiconZH/离婚[鲁迅小说《离婚》]- sense

lemon:reference

lemon: LexicalSense
LexiconZH/离婚[脱离夫妻婚姻关系]- sense

lemon: Lexicon
LexiconZH

lemon:entry

lemon:sense

lemon:sense

lemon: LexicalEntry
LexiconZH/离婚

owl:sameAs

lemon:otherForm

lemon:sense

lemon: LexicalSense
LexiconZH/s00027937n-sense

lemon: LexicalForm
LexiconZH/离婚-Hant-form

lemon:
canonicalForm

lemon:writtenRep

lemon: LexicalForm
LexiconZH/离婚-Hans-form

lemon:reference

lemon:otherForm

rdfs:Resource
babelnet.org/rdf/s00027937n

"離婚"
@zh-Hant

lemon: LexicalForm
LexiconZH/离婚-pinyin-form

lexinfo:romanization

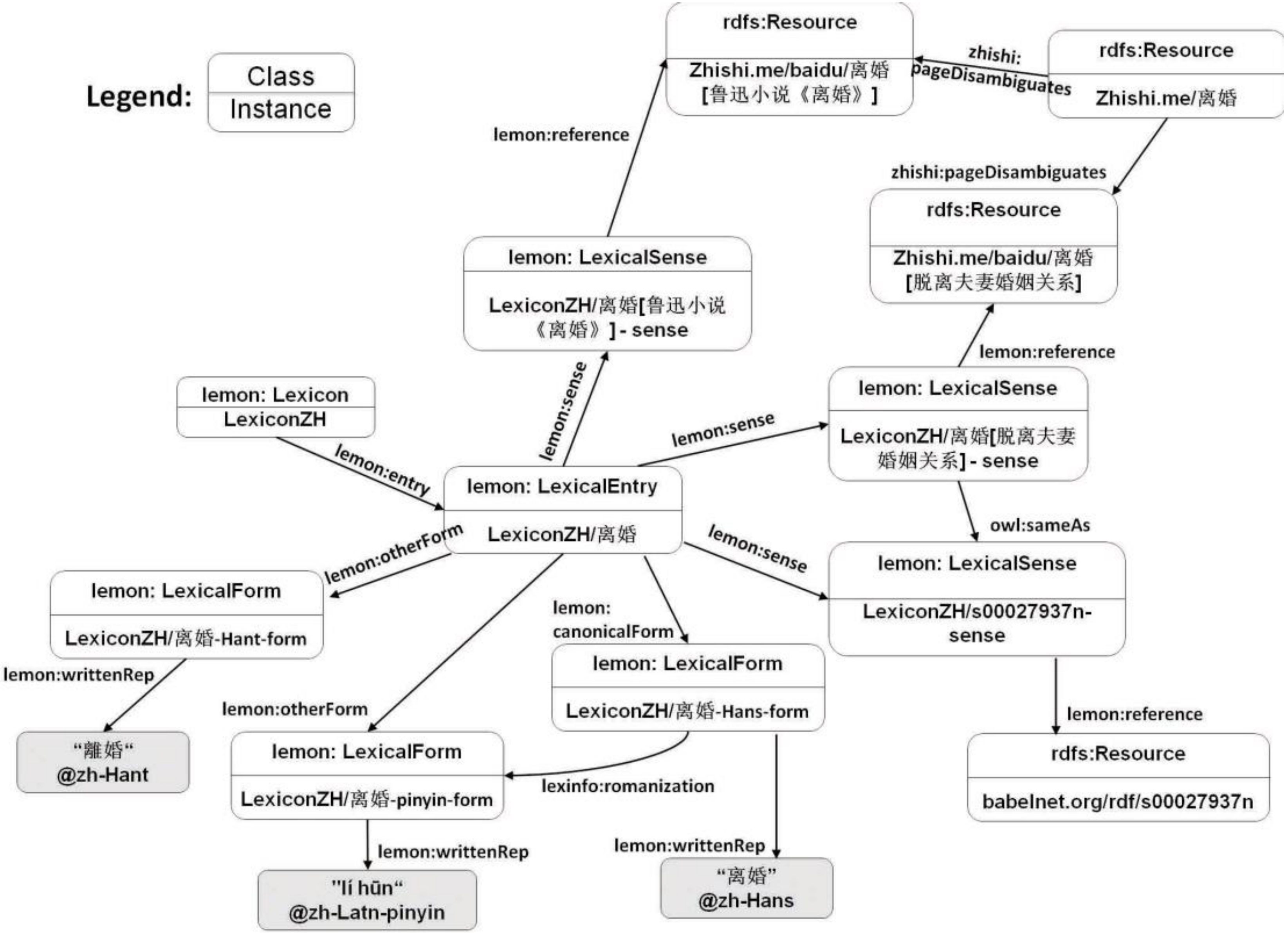
lemon:writtenRep

lemon:writtenRep

"lí hūn"
@zh-Latn-pinyin

"离婚"
@zh-Hans

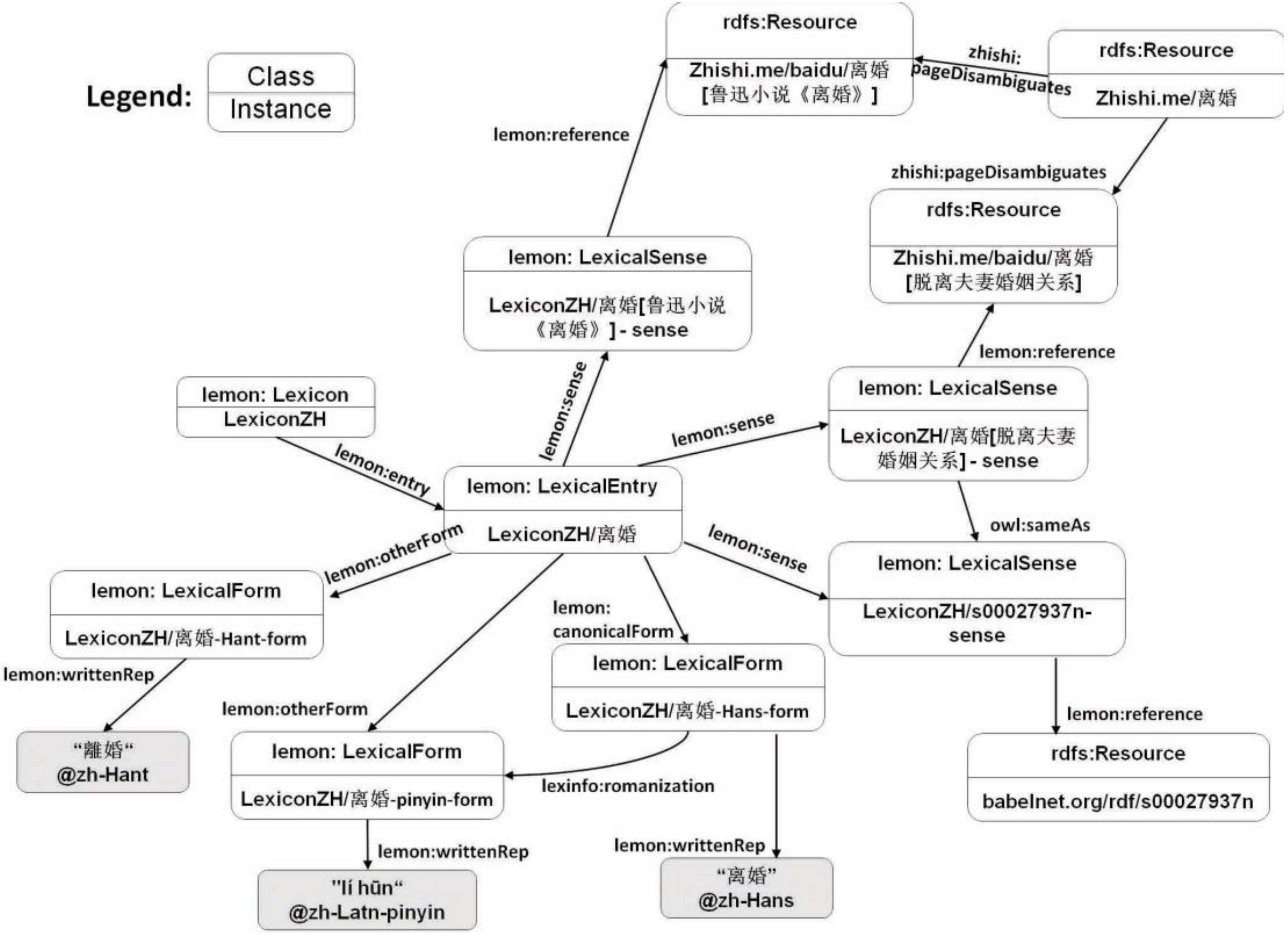
Chinese Lexicalization Module



Chinese Lexicalization Module

LexicalEntry & LexicalSense

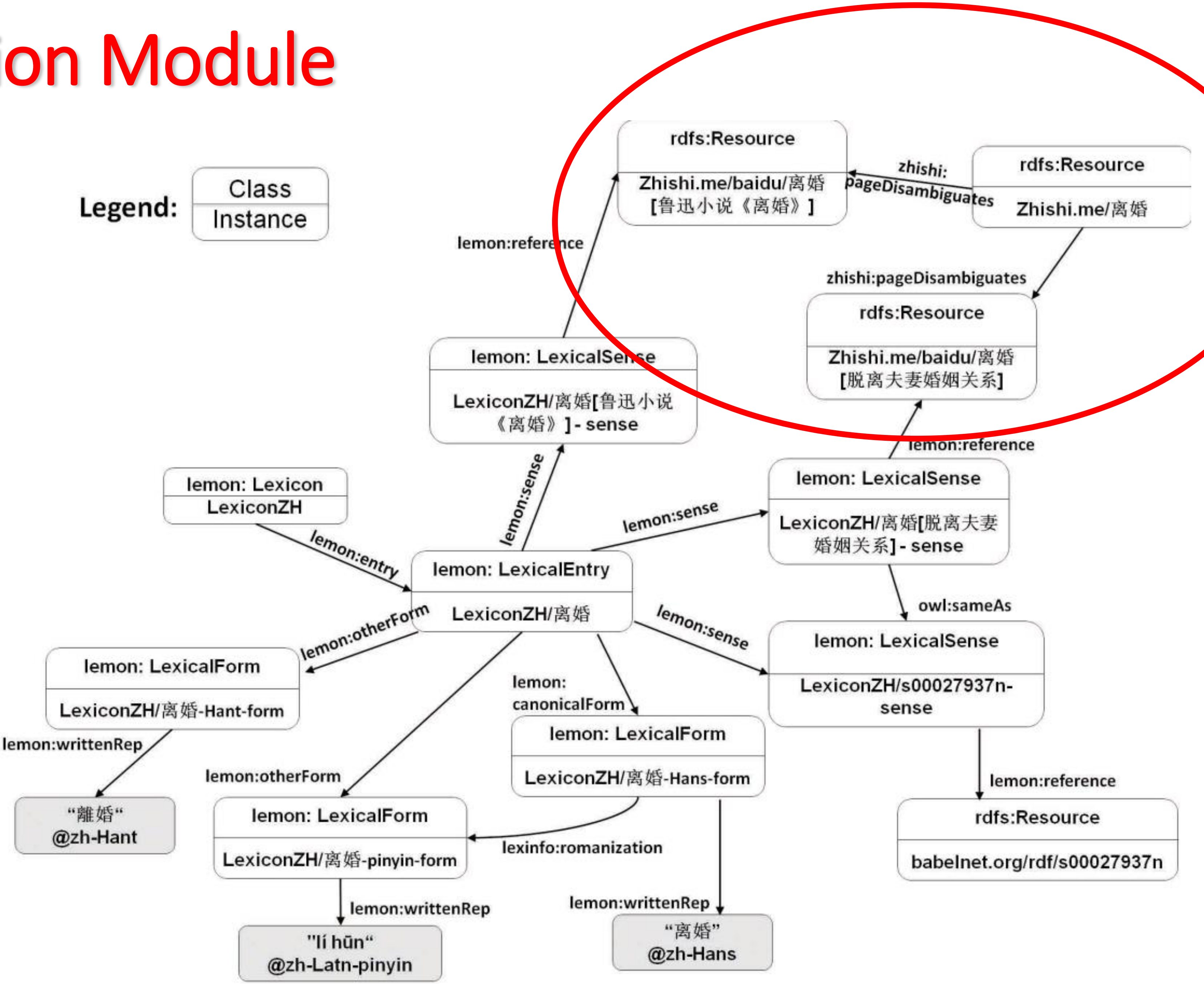
- The title of an article in an encyclopedia site is usually ambiguous.
- URIs in Zhishi.me provide semantics for each lexical entry.



Chinese Lexicalization Module

LexicalEntry & LexicalSense

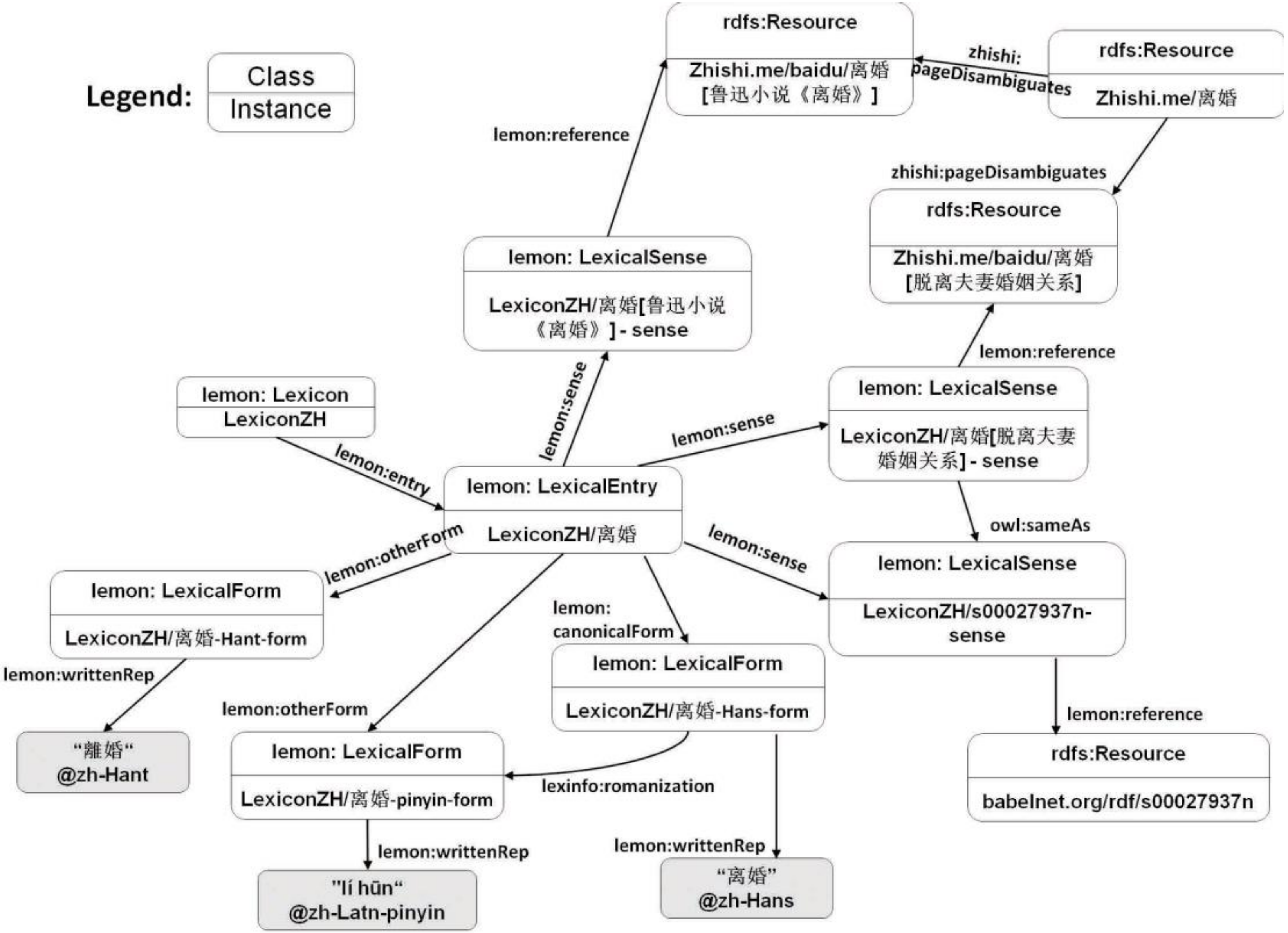
- The title of an article in an encyclopedia site is usually ambiguous.
- URIs in Zhishi.me provide semantics for each lexical entry.



Chinese Lexicalization Module

LexicalEntry & LexicalSense

- The title of an article in an encyclopedia site is usually ambiguous.
- URIs in Zhishi.me provide semantics for each lexical entry.



Legend: Class / Instance

rdfs:Resource — Zhishi.me/baidu/离婚[鲁迅小说《离婚》]

rdfs:Resource — Zhishi.me/离婚

zhishi: pageDisambiguates

zhishi:pageDisambiguates

rdfs:Resource — Zhishi.me/baidu/离婚[脱离夫妻婚姻关系]

lemon:reference

lemon: LexicalSense — LexiconZH/离婚[鲁迅小说《离婚》]- sense

lemon:reference

lemon: LexicalSense — LexiconZH/离婚[脱离夫妻婚姻关系]- sense

lemon: Lexicon — LexiconZH

lemon:entry

lemon:sense

lemon:sense

lemon: LexicalEntry — LexiconZH/离婚

owl:sameAs

lemon:otherForm

lemon:sense

lemon: LexicalSense — LexiconZH/s00027937n-sense

lemon: LexicalForm — LexiconZH/離婚-Hant-form

lemon: canonicalForm

lemon: LexicalForm — LexiconZH/离婚-Hans-form

lemon:writtenRep

lemon:otherForm

lemon:reference

“離婚“ @zh-Hant

lemon: LexicalForm — LexiconZH/离婚-pinyin-form

lexinfo:romanization

rdfs:Resource — babelnet.org/rdf/s00027937n

lemon:writtenRep

lemon:writtenRep

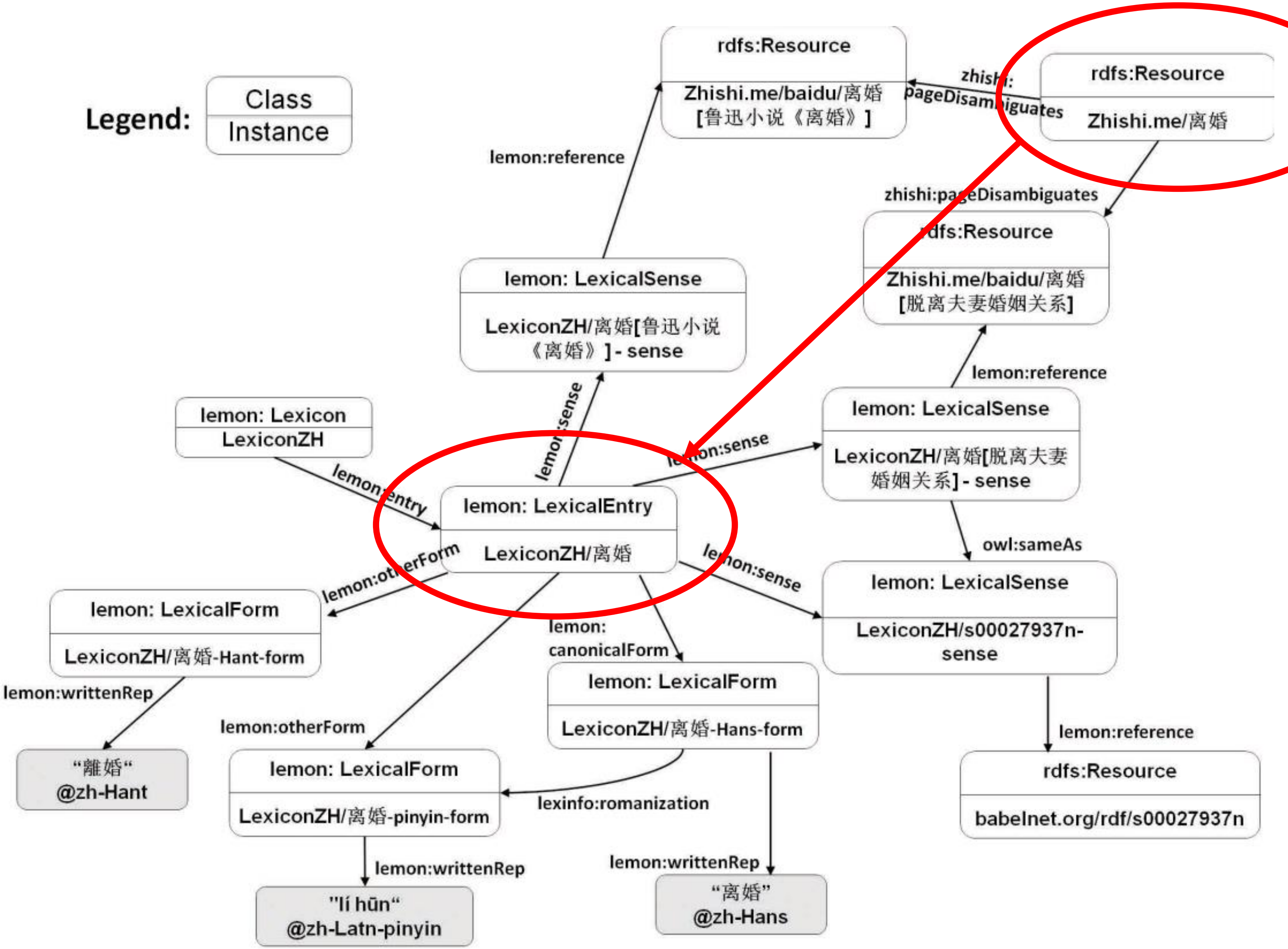
"lí hūn“ @zh-Latn-pinyin

“离婚” @zh-Hans

Chinese Lexicalization Module

LexicalEntry & LexicalSense

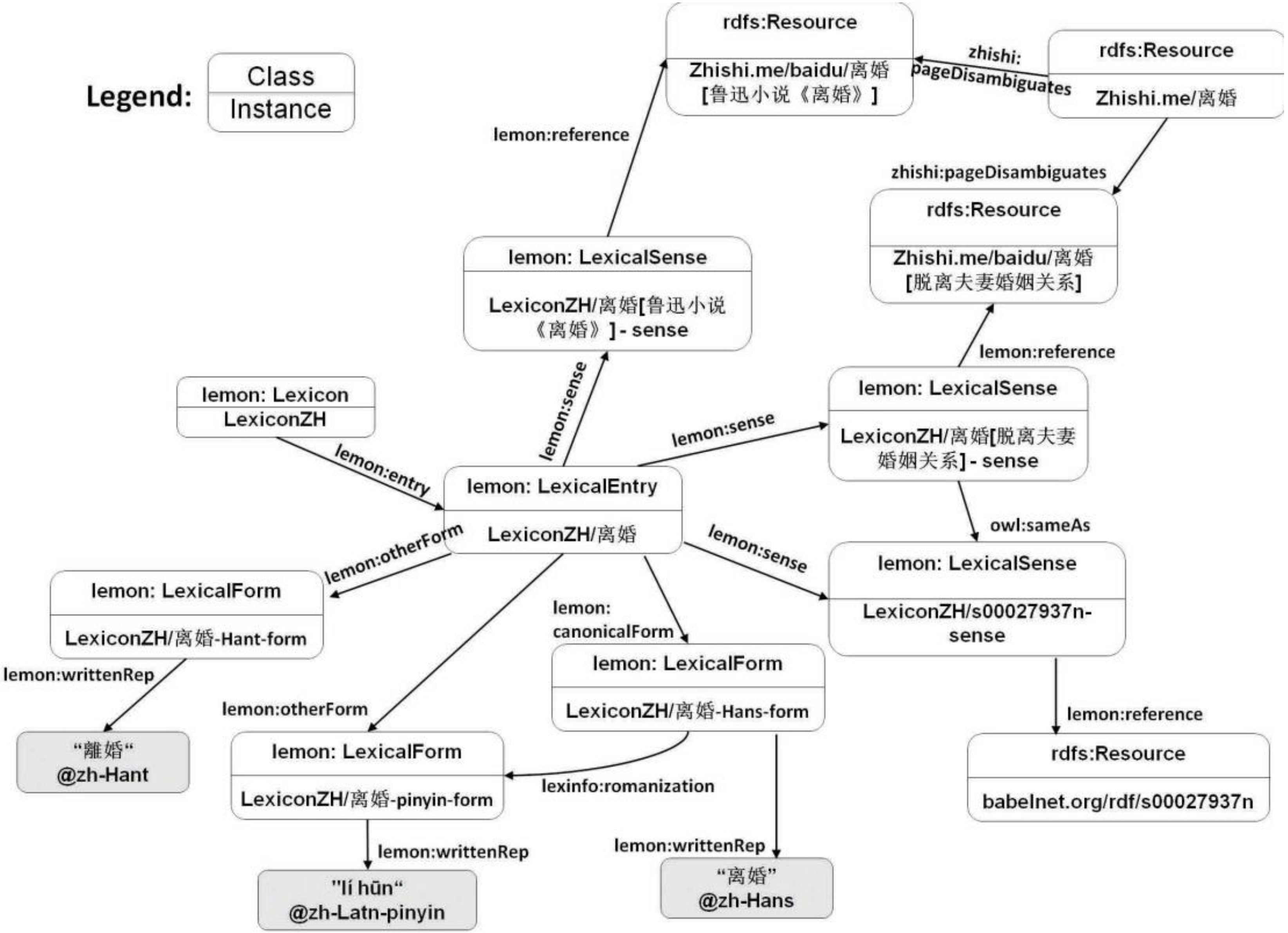
- The title of an article in an encyclopedia site is usually ambiguous.
- URIs in Zhishi.me provide semantics for each lexical entry.



Chinese Lexicalization Module

LexicalEntry & LexicalSense

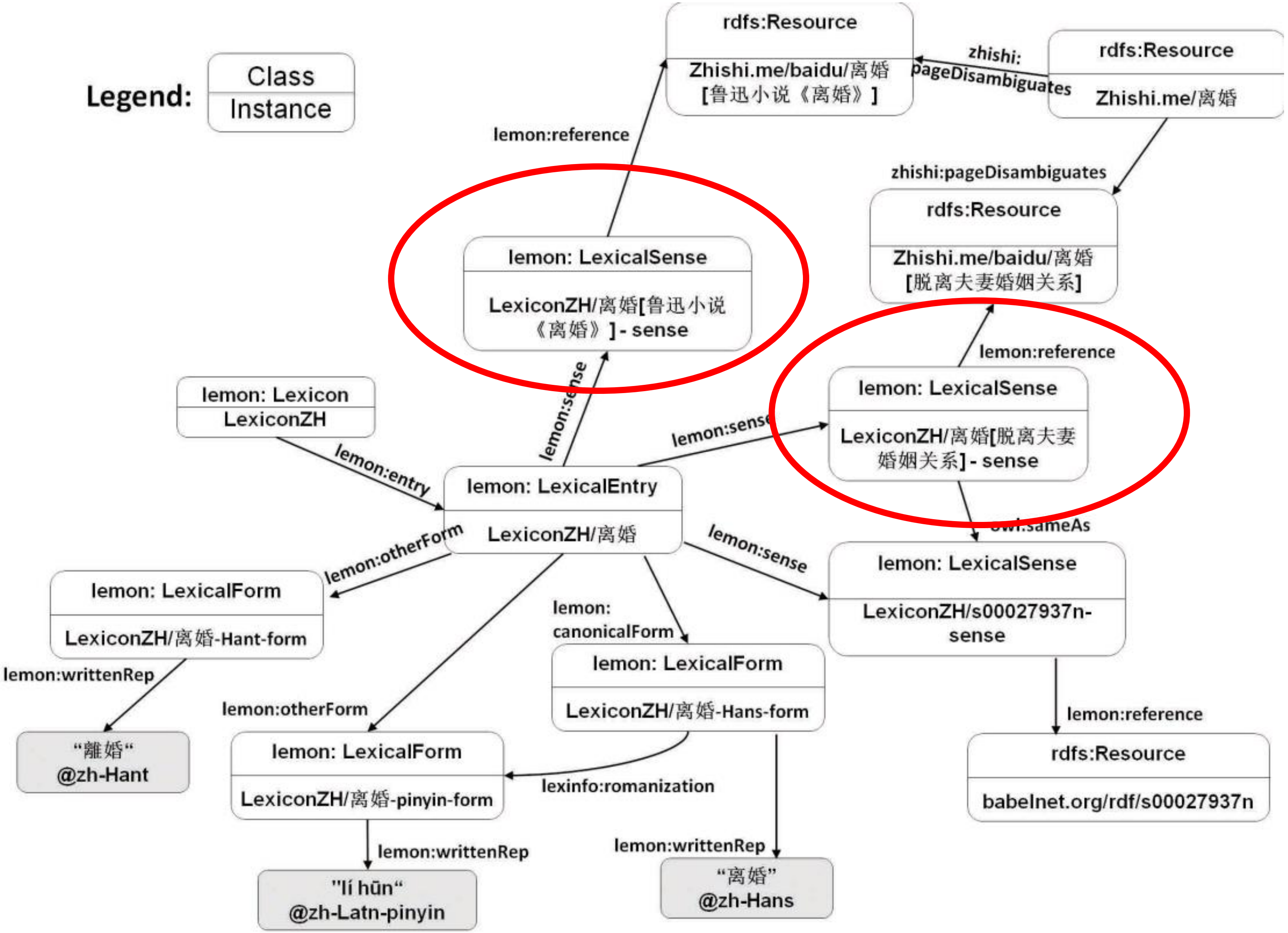
- The title of an article in an encyclopedia site is usually ambiguous.
- URIs in Zhishi.me provide semantics for each lexical entry.



Chinese Lexicalization Module

LexicalEntry & LexicalSense

- The title of an article in an encyclopedia site is usually ambiguous.
- URIs in Zhishi.me provide semantics for each lexical entry.



Legend: Class / Instance

rdfs:Resource — Zhishi.me/baidu/离婚[鲁迅小说《离婚》]

rdfs:Resource — Zhishi.me/离婚

zhishi: pageDisambiguates

zhishi:pageDisambiguates

rdfs:Resource — Zhishi.me/baidu/离婚[脱离夫妻婚姻关系]

lemon:reference

lemon: LexicalSense — LexiconZH/离婚[鲁迅小说《离婚》]- sense

lemon: LexicalSense — LexiconZH/离婚[脱离夫妻婚姻关系]- sense

lemon: Lexicon — LexiconZH

lemon:entry

lemon:sense

lemon: LexicalEntry — LexiconZH/离婚

owl:sameAs

lemon: LexicalSense — LexiconZH/s00027937n-sense

lemon:otherForm

lemon: LexicalForm — LexiconZH/离婚-Hant-form

lemon: canonicalForm

lemon: LexicalForm — LexiconZH/离婚-Hans-form

lemon:writtenRep

"離婚" @zh-Hant

lemon: LexicalForm — LexiconZH/离婚-pinyin-form

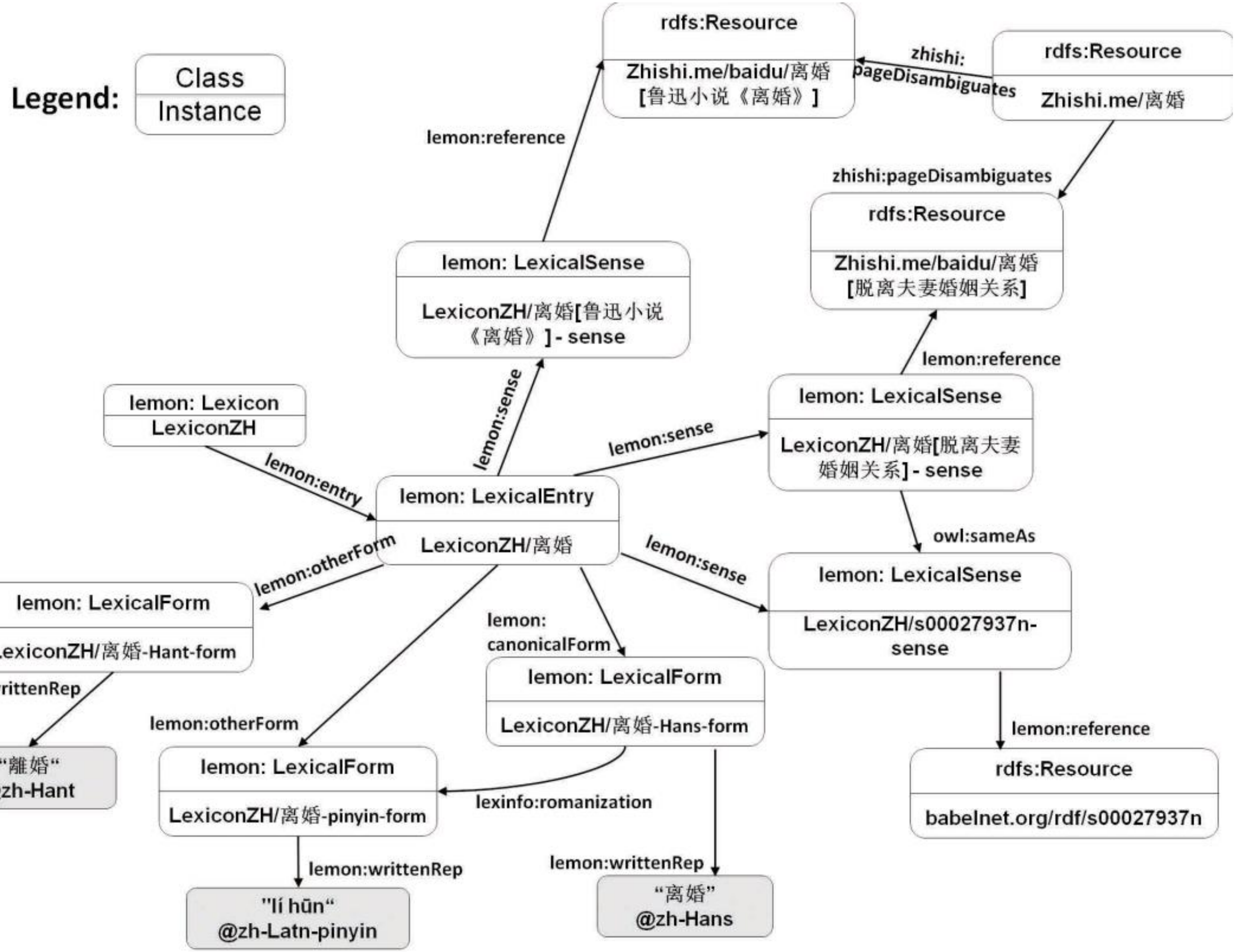
lexinfo:romanization

rdfs:Resource — babelnet.org/rdf/s00027937n

"lí hūn" @zh-Latn-pinyin

"离婚" @zh-Hans

Chinese Lexicalization Module



Legend: Class / Instance

rdfs:Resource
Zhishi.me/baidu/离婚
[鲁迅小说《离婚》]

zhishi:
pageDisambiguates

rdfs:Resource
Zhishi.me/离婚

lemon:reference

zhishi:pageDisambiguates

rdfs:Resource
Zhishi.me/baidu/离婚
[脱离夫妻婚姻关系]

lemon: LexicalSense
LexiconZH/离婚[鲁迅小说《离婚》] - sense

lemon:reference

lemon: Lexicon
LexiconZH

lemon:sense

lemon:sense

lemon: LexicalSense
LexiconZH/离婚[脱离夫妻婚姻关系] - sense

lemon:entry

lemon: LexicalEntry
LexiconZH/离婚

owl:sameAs

lemon:otherForm

lemon:sense

lemon: LexicalSense
LexiconZH/s00027937n-sense

lemon: LexicalForm
LexiconZH/离婚-Hant-form

lemon:
canonicalForm

lemon:writtenRep

lemon: LexicalForm
LexiconZH/离婚-Hans-form

lemon:otherForm

lemon:reference

"離婚"
@zh-Hant

lemon: LexicalForm
LexiconZH/离婚-pinyin-form

lexinfo:romanization

rdfs:Resource
babelnet.org/rdf/s00027937n

lemon:writtenRep

lemon:writtenRep

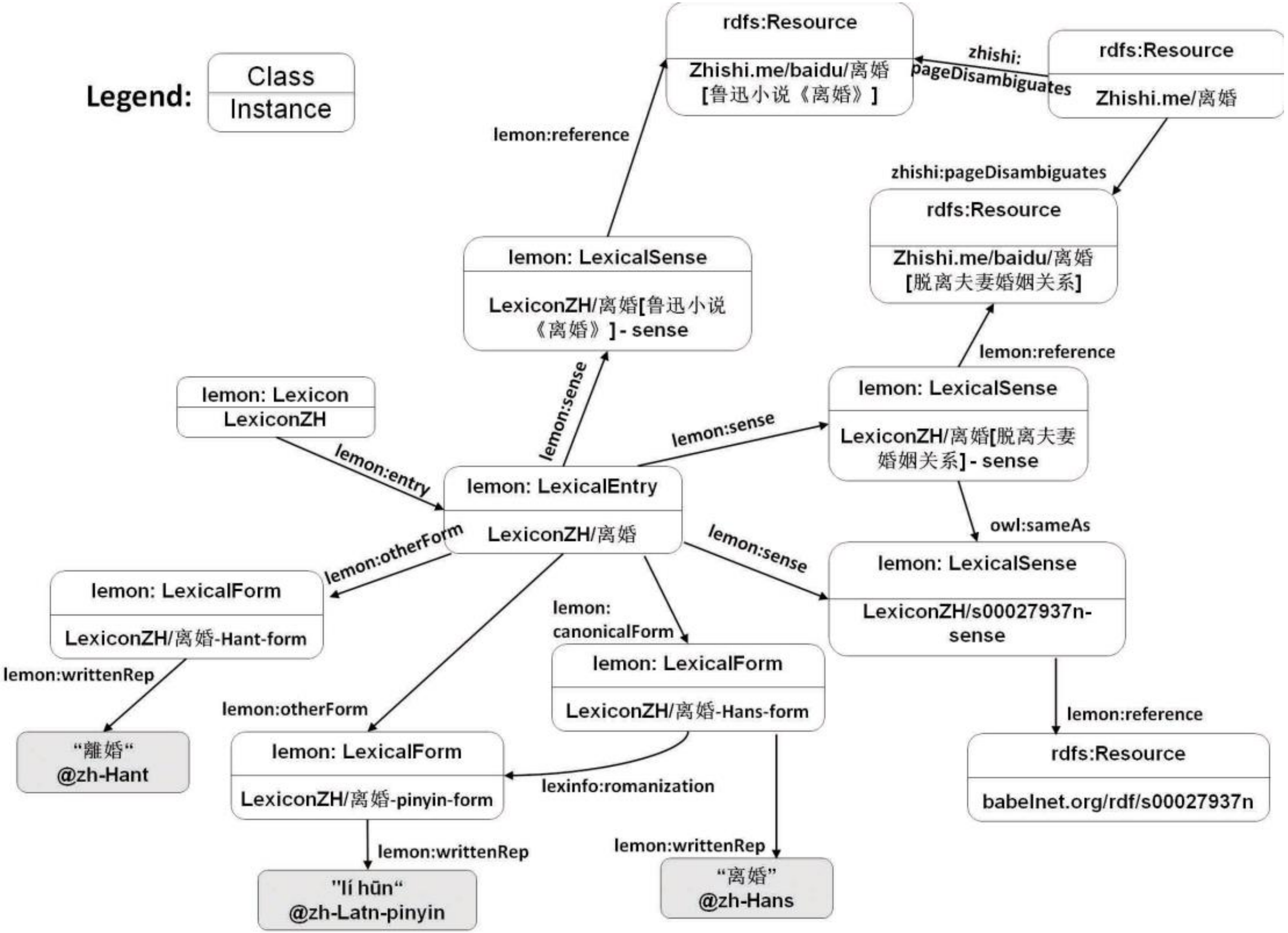
"lí hūn"
@zh-Latn-pinyin

"离婚"
@zh-Hans

Chinese Lexicalization Module

LexicalEntry & LexicalSense

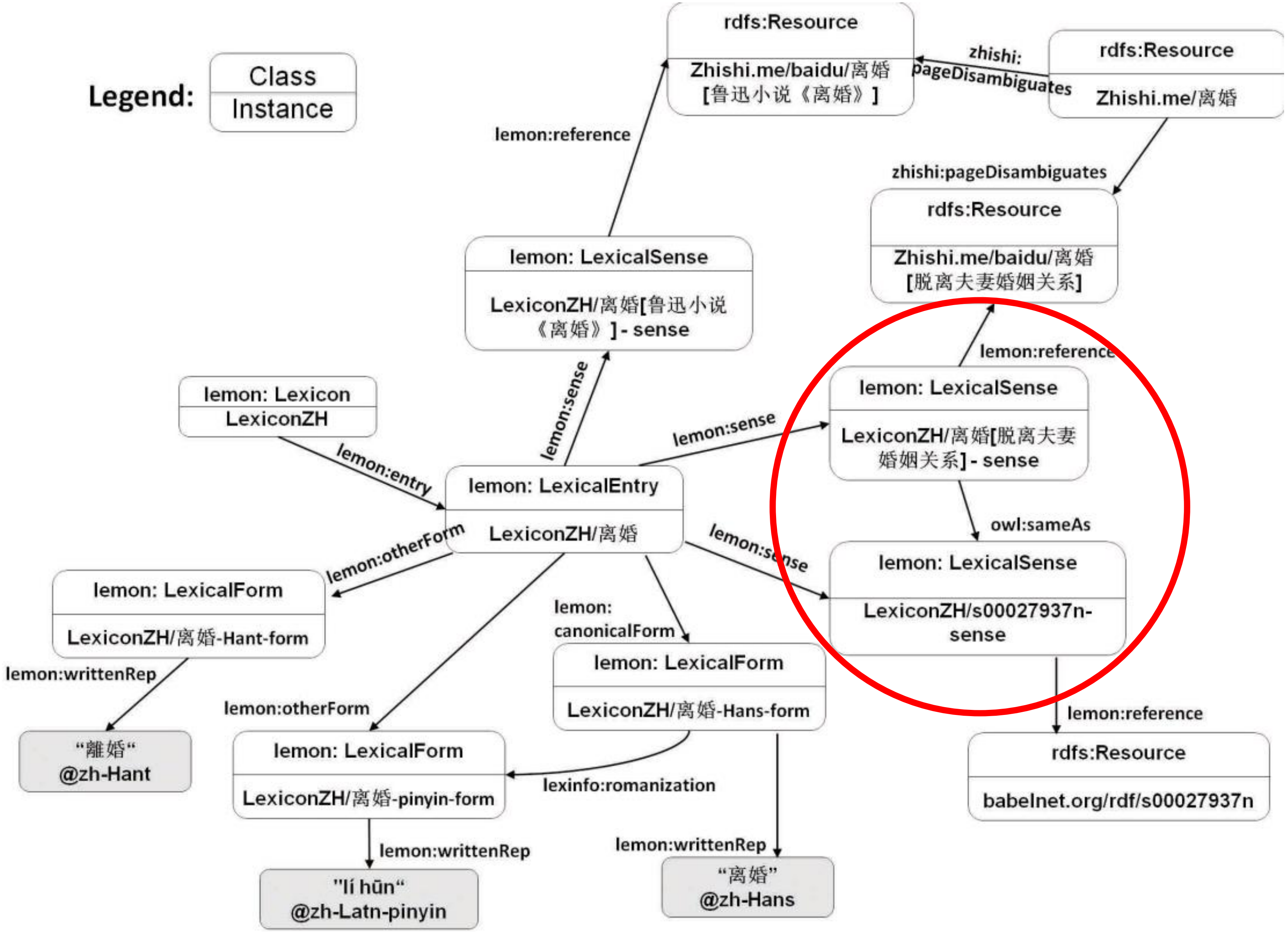
- We link the lexical senses that associate a same lexical entry with two semantically equivalent ontology descriptions by using a owl:sameAs relation.



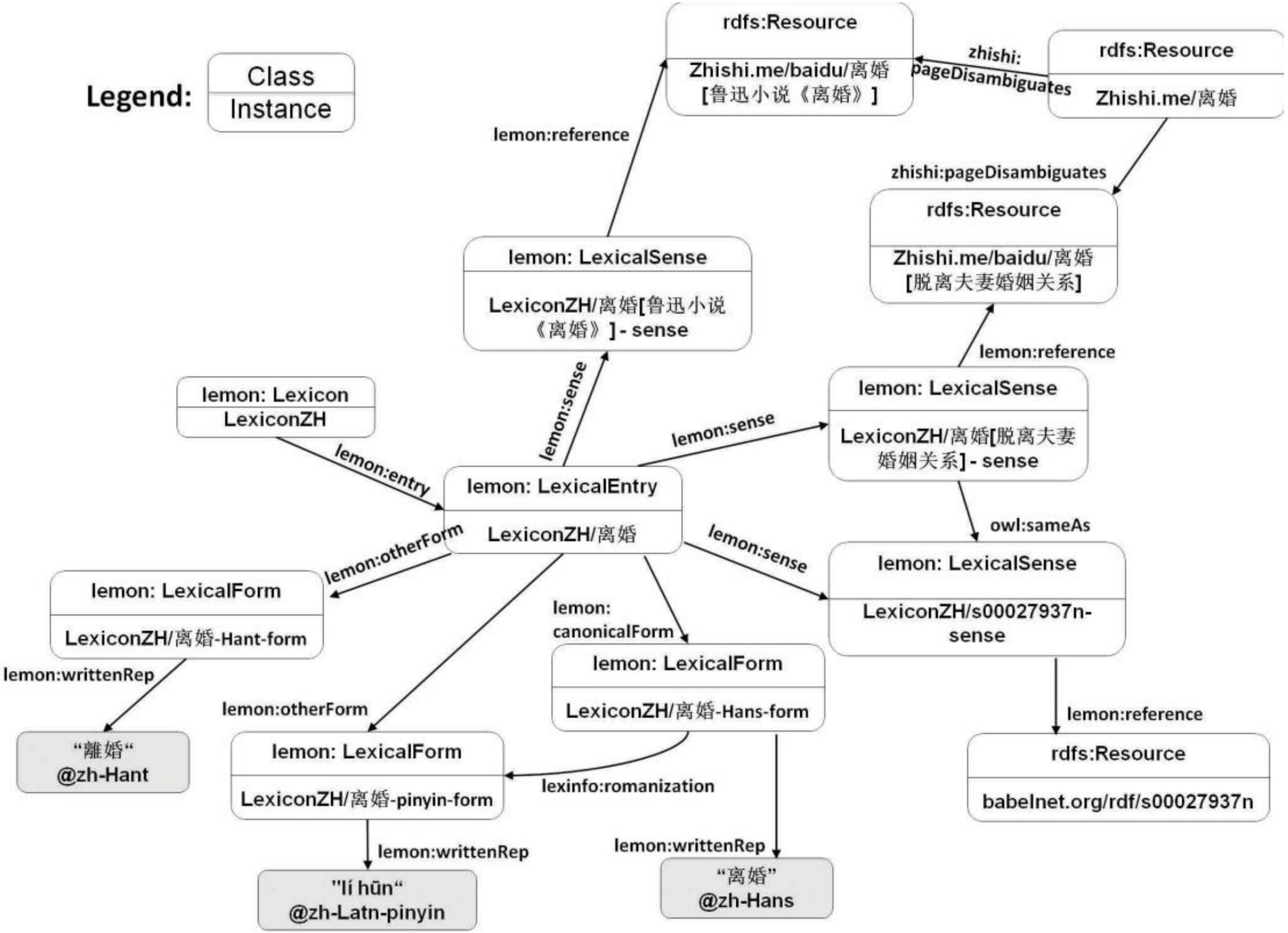
Chinese Lexicalization Module

LexicalEntry & LexicalSense

- We link the lexical senses that associate a same lexical entry with two semantically equivalent ontology descriptions by using a owl:sameAs relation.



Chinese Lexicalization Module



Legend: Class

Instance

rdfs:Resource

Zhishi.me/baidu/离婚[鲁迅小说《离婚》]

zhishi: pageDisambiguates

rdfs:Resource

Zhishi.me/离婚

lemon:reference

zhishi:pageDisambiguates

rdfs:Resource

Zhishi.me/baidu/离婚[脱离夫妻婚姻关系]

lemon: LexicalSense

LexiconZH/离婚[鲁迅小说《离婚》] - sense

lemon:reference

lemon: Lexicon

LexiconZH

lemon:sense

lemon:sense

lemon: LexicalSense

LexiconZH/离婚[脱离夫妻婚姻关系] - sense

lemon:entry

lemon: LexicalEntry

LexiconZH/离婚

owl:sameAs

lemon:otherForm

lemon:sense

lemon: LexicalSense

LexiconZH/s00027937n-sense

lemon: LexicalForm

LexiconZH/离婚-Hant-form

lemon: canonicalForm

lemon: LexicalForm

LexiconZH/离婚-Hans-form

lemon:writtenRep

lemon:otherForm

lemon:reference

“離婚“ @zh-Hant

lemon: LexicalForm

LexiconZH/离婚-pinyin-form

lexinfo:romanization

rdfs:Resource

babelnet.org/rdf/s00027937n

lemon:writtenRep

lemon:writtenRep

"lí hūn“ @zh-Latn-pinyin

“离婚” @zh-Hans

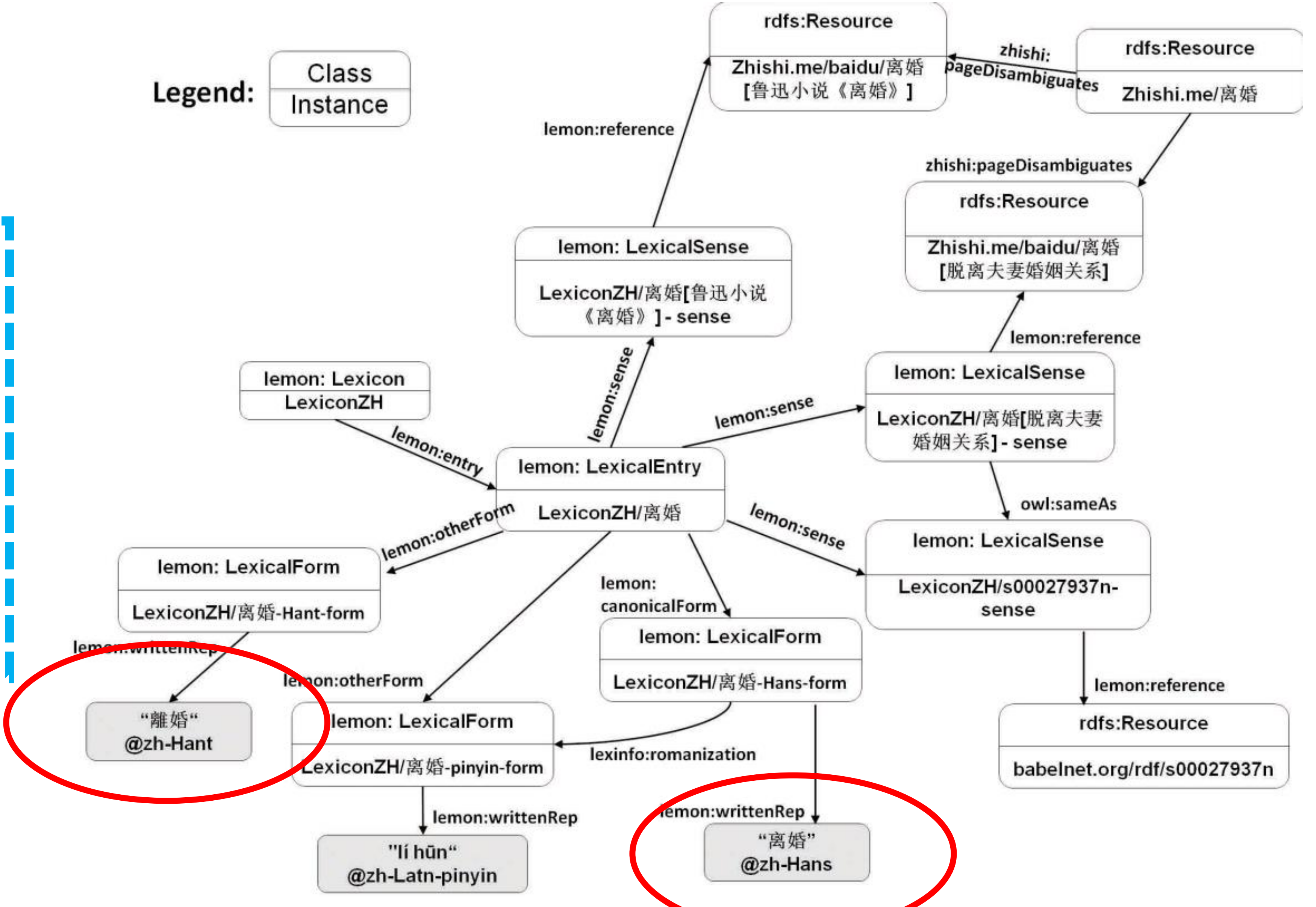
Chinese Lexicalization Module

Chinese form

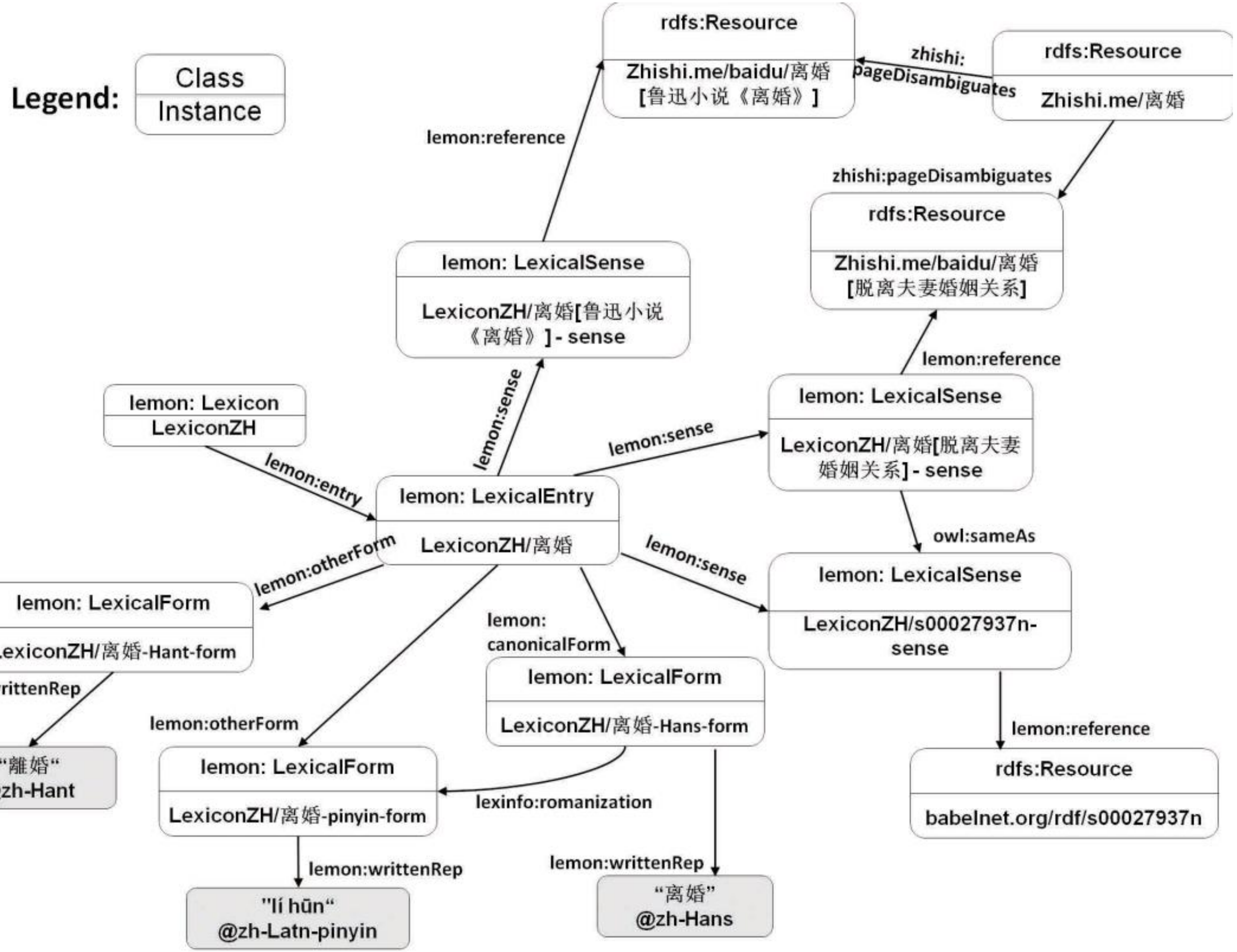
- Modeling both simplified and traditional Chinese characters
- Using different language codes

zh-Hans => simplified Chinese

zh-Hant => traditional Chinese



Chinese Lexicalization Module



Legend: Class / Instance

rdfs:Resource
Zhishi.me/baidu/离婚
[鲁迅小说《离婚》]

zhishi:
pageDisambiguates

rdfs:Resource
Zhishi.me/离婚

lemon:reference

zhishi:pageDisambiguates

rdfs:Resource
Zhishi.me/baidu/离婚
[脱离夫妻婚姻关系]

lemon: LexicalSense
LexiconZH/离婚[鲁迅小说
《离婚》] - sense

lemon:reference

lemon: Lexicon
LexiconZH

lemon:sense

lemon:sense

lemon: LexicalSense
LexiconZH/离婚[脱离夫妻
婚姻关系] - sense

lemon:entry

lemon: LexicalEntry
LexiconZH/离婚

owl:sameAs

lemon:otherForm

lemon:sense

lemon: LexicalSense
LexiconZH/s00027937n-
sense

lemon: LexicalForm
LexiconZH/离婚-Hant-form

lemon:
canonicalForm

lemon:writtenRep

lemon: LexicalForm
LexiconZH/离婚-Hans-form

lemon:otherForm

lemon:reference

“離婚“
@zh-Hant

lemon: LexicalForm
LexiconZH/离婚-pinyin-form

lexinfo:romanization

rdfs:Resource
babelnet.org/rdf/s00027937n

lemon:writtenRep

lemon:writtenRep

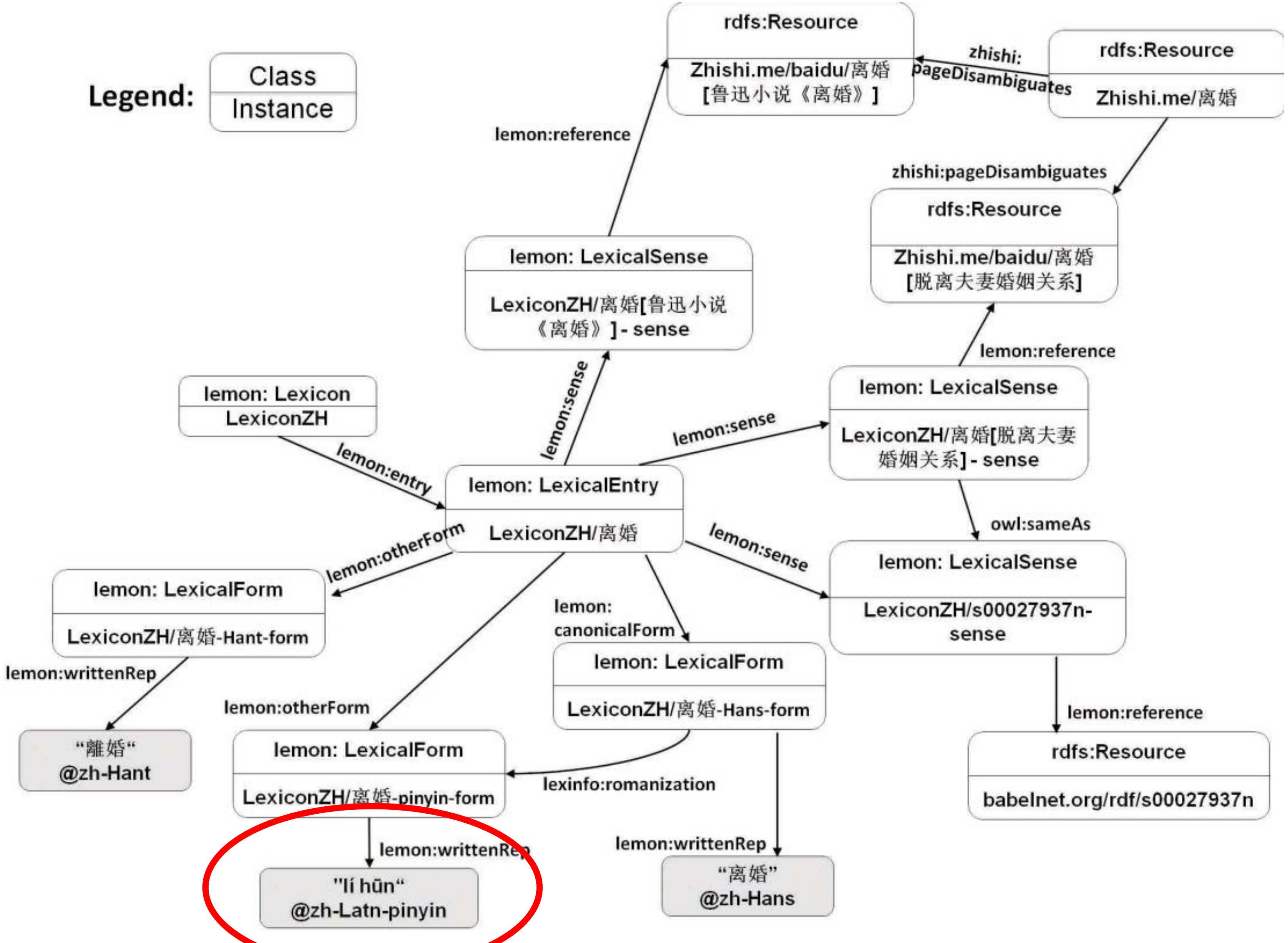
"lí hūn“
@zh-Latn-pinyin

“离婚”
@zh-Hans

Chinese Lexicalization Module

Chinese Romanization

- Chinese use “Hanyu Pinyin” as a common romanization standard.
- zh-Latn-pinyin => Language code of pinyin



Legend: Class / Instance

rdfs:Resource
Zhishi.me/baidu/离婚[鲁迅小说《离婚》]

zhishi:pageDisambiguates

rdfs:Resource
Zhishi.me/离婚

lemon:reference

zhishi:pageDisambiguates

rdfs:Resource
Zhishi.me/baidu/离婚[脱离夫妻婚姻关系]

lemon: LexicalSense
LexiconZH/离婚[鲁迅小说《离婚》]- sense

lemon:reference

lemon: Lexicon
LexiconZH

lemon:sense

lemon:sense

lemon: LexicalSense
LexiconZH/离婚[脱离夫妻婚姻关系]- sense

lemon:entry

lemon: LexicalEntry
LexiconZH/离婚

owl:sameAs

lemon:otherForm

lemon:sense

lemon: LexicalSense
LexiconZH/s00027937n-sense

lemon: LexicalForm
LexiconZH/离婚-Hant-form

lemon: canonicalForm

lemon: LexicalForm
LexiconZH/离婚-Hans-form

lemon:writtenRep

lemon:otherForm

lemon:reference

“離婚“
@zh-Hant

lemon: LexicalForm
LexiconZH/离婚-pinyin-form

lexinfo:romanization

rdfs:Resource
babelnet.org/rdf/s00027937n

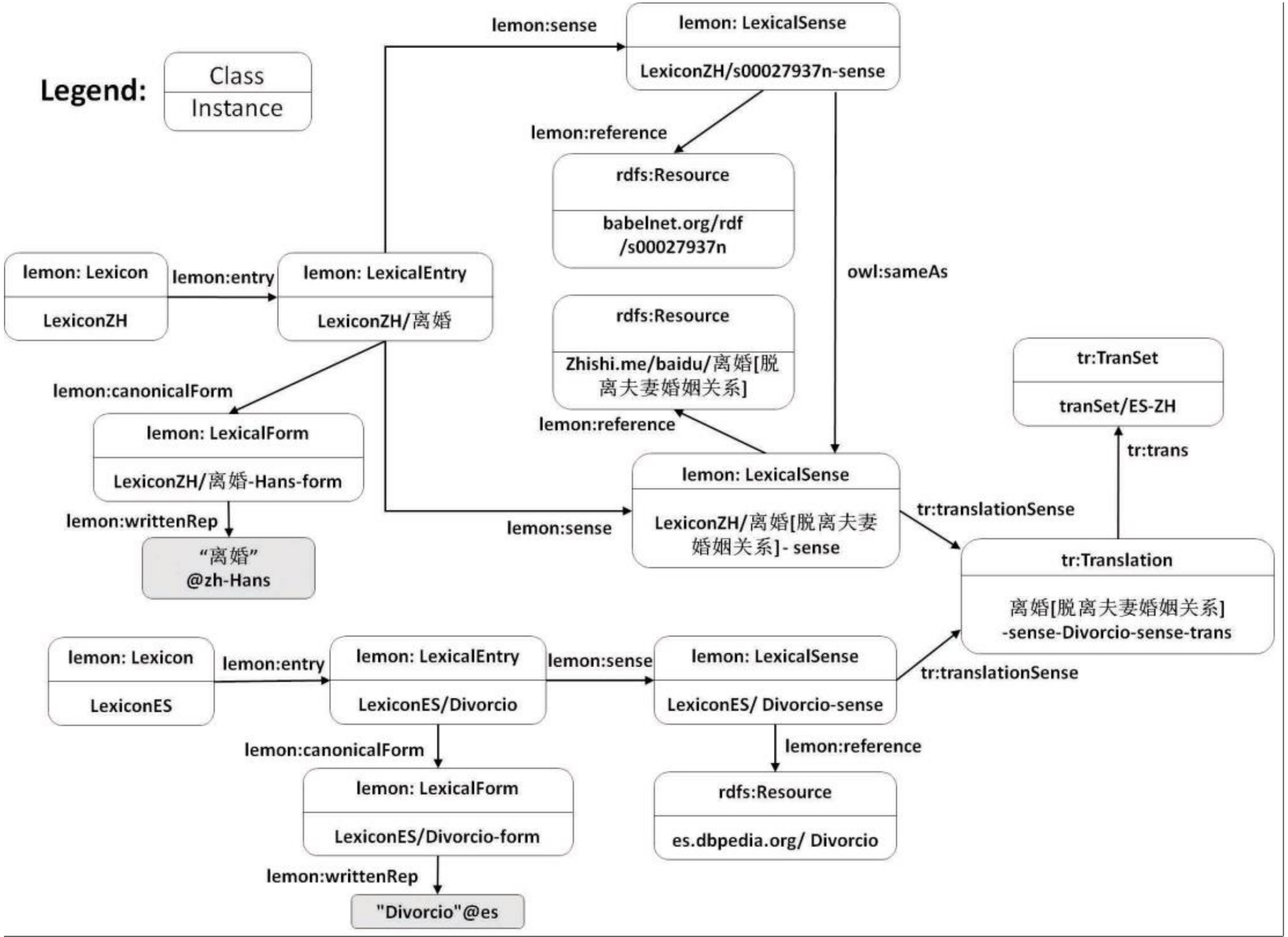
lemon:writtenRep

lemon:writtenRep

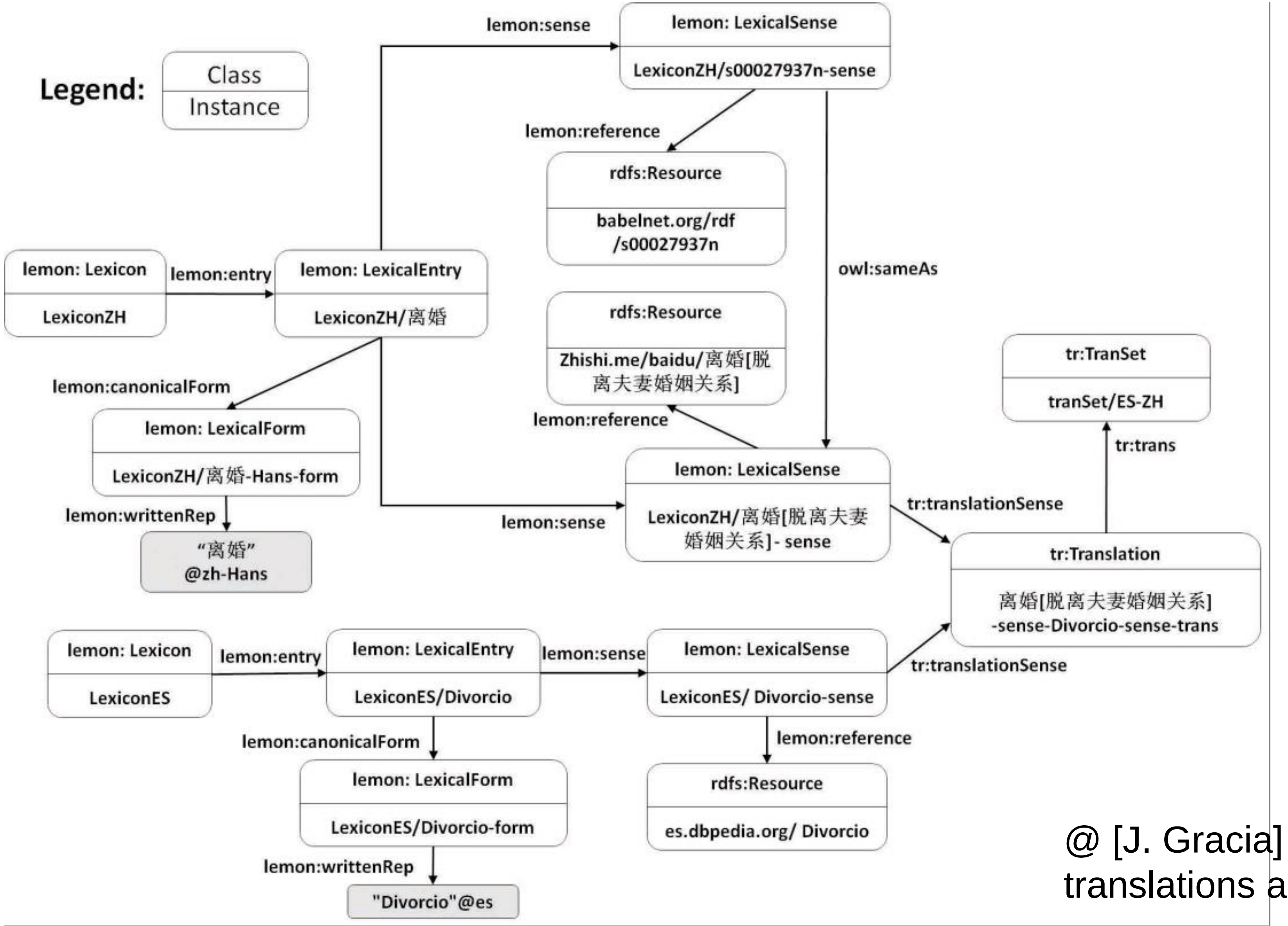
"lí hūn“
@zh-Latn-pinyin

“离婚”
@zh-Hans

Multilingual Translation Module



Multilingual Translation Module

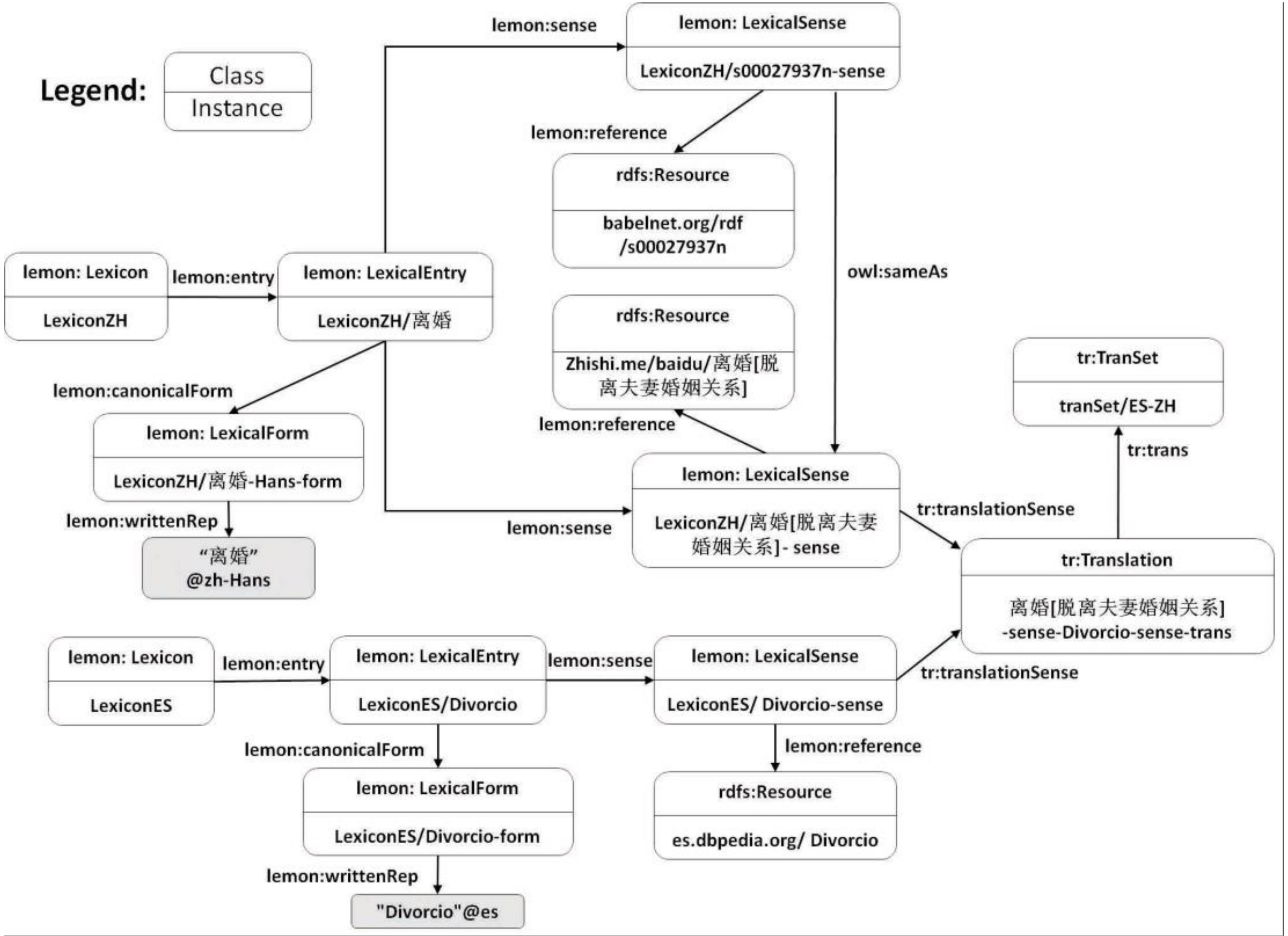


Translation Module

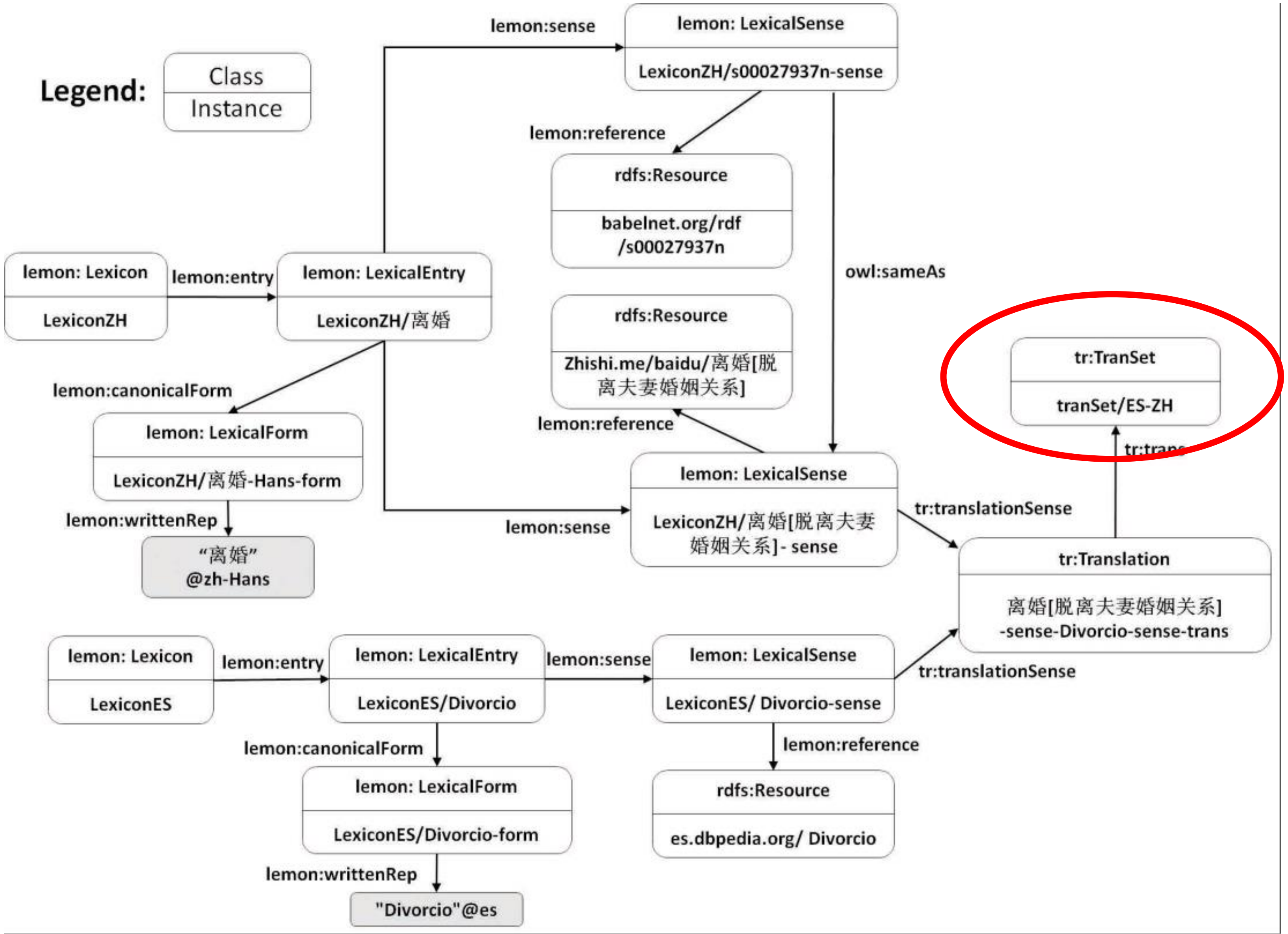
- Translation relations can be inferred between terms in different languages when they refer to the same ontology entity.

@ [J. Gracia] Enabling language resources to expose translations as linked data on the web. 2014.

Multilingual Translation Module



Multilingual Translation Module



Legend: Class / Instance

lemon: Lexicon — LexiconZH

lemon:entry

lemon: LexicalEntry — LexiconZH/离婚

lemon:sense

lemon: LexicalSense — LexiconZH/s00027937n-sense

lemon:reference

rdfs:Resource — babelnet.org/rdf /s00027937n

owl:sameAs

rdfs:Resource — Zhishi.me/baidu/离婚[脱离夫妻婚姻关系]

lemon:reference

lemon:canonicalForm

lemon: LexicalForm — LexiconZH/离婚-Hans-form

lemon:writtenRep

"离婚" @zh-Hans

lemon:sense

lemon: LexicalSense — LexiconZH/离婚[脱离夫妻婚姻关系]- sense

tr:translationSense

tr:TranSet — tranSet/ES-ZH

tr:trans

tr:Translation — 离婚[脱离夫妻婚姻关系] -sense-Divorcio-sense-trans

lemon: Lexicon — LexiconES

lemon:entry

lemon: LexicalEntry — LexiconES/Divorcio

lemon:sense

lemon: LexicalSense — LexiconES/ Divorcio-sense

tr:translationSense

lemon:canonicalForm

lemon: LexicalForm — LexiconES/Divorcio-form

lemon:writtenRep

"Divorcio"@es

lemon:reference

rdfs:Resource — es.dbpedia.org/ Divorcio

Translation Set

- TranslationSet is designed to group a set of translations, which facilitates querying.
- TranslationSet/ES-ZH
- TranslationSet/EN-ZH
- TranslationSet/ES-EN

Zhishi.lemon Statistics

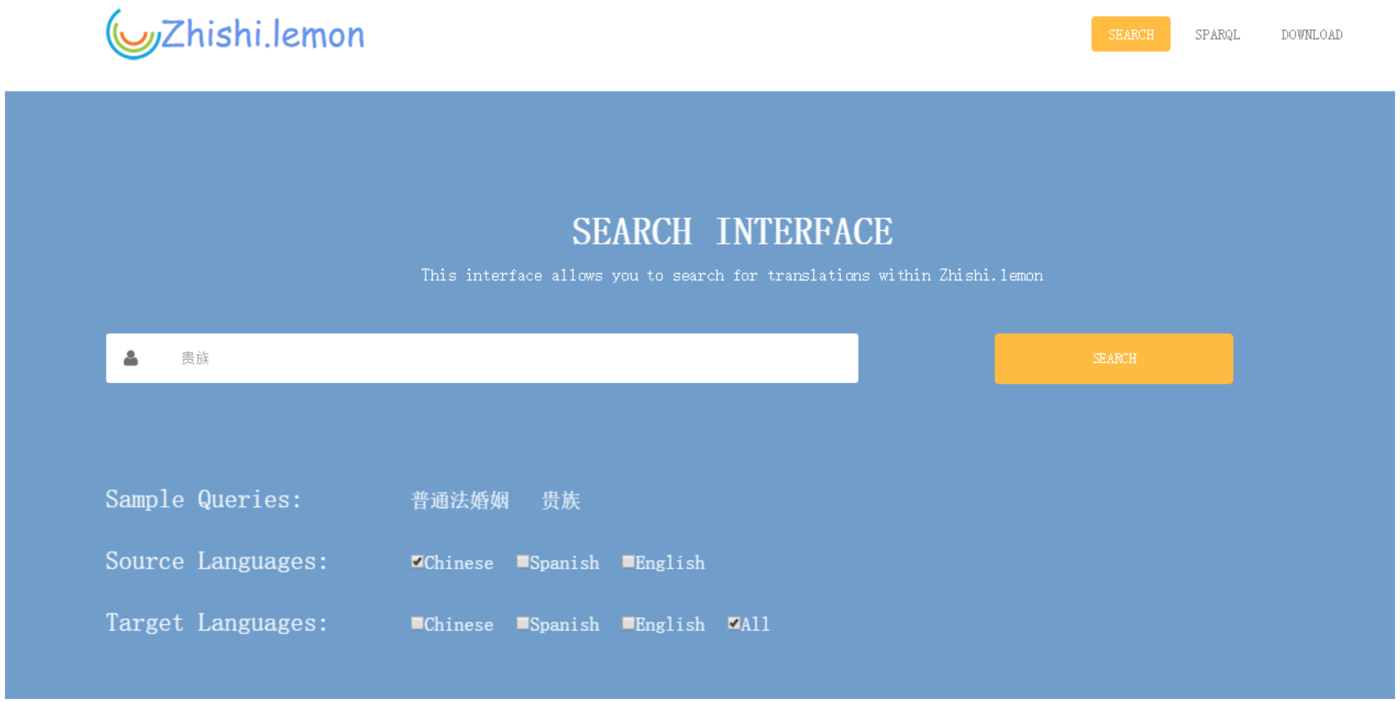
Items	Value
links: Babel Net	16,424
links: DBpedia - en	218,654
links: DBpedia - es	77,392
links: Zhishi.me	229,606
Total Translations	364,765
Total Triples	7,036,338

Zhishi.lemon vs CWN

	CWN	Zhishi.lemon
Word/Lexical Entry	12,726	215,608
Sense/Lexical Sense	34,358	523,585
Lexical Relation/Translation	47,250	364,765

The Zhishi.lemon platform

- **Lookup Service** [http://lemon.zhishi.me/search.html]



SEARCH INTERFACE

This interface allows you to search for translations within Zhishi.lemon

贵族

Sample Queries: 普通法婚姻 贵族

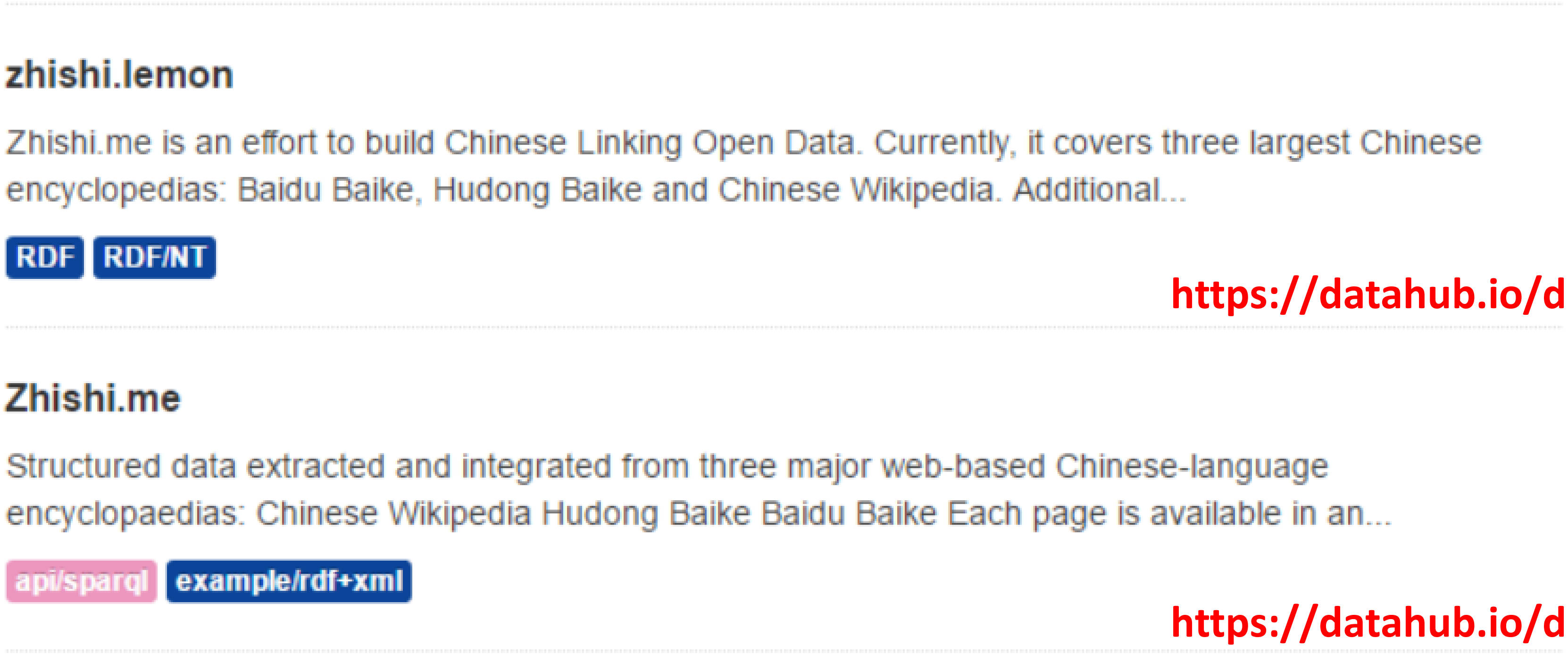
Source Languages: ☑Chinese □Spanish □English

Target Languages: □Chinese □Spanish □English ☑All

- **SPARQL Endpoint** [http://lemon.zhishi.me/sparql.html]

  - Jena TDB is used to store the extracted triples and to provide querying capabilities.

Data Availability (Data Hub)



zhishi.lemon

Zhishi.me is an effort to build Chinese Linking Open Data. Currently, it covers three largest Chinese encyclopedias: Baidu Baike, Hudong Baike and Chinese Wikipedia. Additional...

RDF RDF/NT

https://datahub.io/dataset/zhishi-lemon

Zhishi.me

Structured data extracted and integrated from three major web-based Chinese-language encyclopaedias: Chinese Wikipedia Hudong Baike Baidu Baike Each page is available in an...

api/sparql example/rdf+xml

https://datahub.io/dataset/zhishi-me

Conclusion and Future Work

- **Zhishi.lemon**: a newly developed dataset that constitutes the lexical realization of Zhishi.me.

- Supporting for three languages (Chinese, Spanish and English)

- Links to Dbpedia and BabelNet.

Conclusion and Future Work

- Transform more Chinese resources and integrate them into Zhishi.lemon.

- Leverage Zhishi.lemon to build real-world multilingual applications.

- Identify new translations from Chinese into other prevalent languages.

Thanks !