Wiki resources

T. Mickus

Nov 25^{th}, 2020

Outline

Outline

Wikipedia & how to handle it

In broad terms

- Wikipedia is a collaborative encyclopedia

Wikipedia (/ˌwɪkɪˈpiːdiə/ (listen) *wik-ih-PEE-dee-ə* or /ˌwɪkiˈpiːdiə/ (listen) *wik-ee-PEE-dee-ə*) is a multilingual online encyclopedia created and maintained as an open collaboration project[3] by a community of volunteer editors 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 October 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 & how to handle it

In broad terms

- Wikipedia is a collaborative encyclopedia
- The English version of Wikipedia contains ≈ 6M content pages.

Wikipedia (/ˌwɪkɪˈpiːdiə/ (listen) *wik-ih-PEE-dee-ə* or /ˌwɪkiˈpiːdiə/ (listen) *wik-ee-PEE-dee-ə*) is a multilingual online encyclopedia created and maintained as an open collaboration project[3] by a community of volunteer editors 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 October 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 & how to handle it

In broad terms

- Wikipedia is a collaborative encyclopedia
- The English version of Wikipedia contains $\approx$ 6M content pages.
- Wikipedia exists in 304 languages (+ 10 inactive projects), cf. `https://en.wikipedia.org/wiki/List_of_Wikipedias`

> **Wikipedia** (/ˌwɪkɪˈpiːdiə/ (listen) *wik-ih-PEE-dee-ə* or /ˌwɪkiˈpiːdiə/ (listen) *wik-ee-PEE-dee-ə*) is a multilingual online encyclopedia created and maintained as an open collaboration project[3] by a community of volunteer editors 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 October 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 is an online encyclopedia, ie. “a reference work or compendium providing summaries of knowledge either from all branches or from a particular field or discipline.”

  NB: def. from Wikipedia

Wikipedia & how to handle it

Content pages

- Wikipedia is an online encyclopedia, ie. “a reference work or compendium providing summaries of knowledge either from all branches or from a particular field or discipline.”

  NB: def. from Wikipedia

- In the case of Wikipedia, these summaries are known as Wikipedia ‘pages’ (more precisely ‘content pages’) or ‘articles’.

Wikipedia & how to handle it

Content pages

- Wikipedia is an online encyclopedia, ie. “a reference work or compendium providing summaries of knowledge either from all branches or from a particular field or discipline.”
  NB: def. from Wikipedia

- In the case of Wikipedia, these summaries are known as Wikipedia ‘pages’ (more precisely ‘content pages’) or ‘articles’.

- As is often the case with collaborative projects, the quality and reliability of pages **vary** (both across pages, across domains and across languages).
  NB: `https://en.wikipedia.org/wiki/Wikipedia:Wikipedia_is_not_a_reliable_source`

Wikipedia & how to handle it

Infoboxes and Lead sections

There are multiple ways of extracting information from a wikipedia page.

There are multiple ways of extracting information from a wikipedia page. The NLP community generally focuses on two specific elements

1. **Lead Section:** The few first sentences of an article, that broach over a subject and attempts to give both overview and introduction to the full contents of the page (cf. `https://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style/Lead_section`).

Wikipedia & how to handle it

Infoboxes and Lead sections

There are multiple ways of extracting information from a wikipedia page. The NLP community generally focuses on two specific elements

1. **Lead Section:** The few first sentences of an article, that broach over a subject and attempts to give both overview and introduction to the full contents of the page (cf. `https://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style/Lead_section`).
2. **Infobox:** Wikipedia contains ‘infoboxes’, viz. tabulated data in attribute-value format presenting the key facts of the related page (cf. `https://en.wikipedia.org/wiki/Infobox`)

Wikipedia & how to handle it

Wikipedia for the developer

- All the content of all Wikipedias is available for download:
  `https://dumps.wikimedia.org/`

Wikipedia & how to handle it

Wikipedia for the developer

- All the content of all Wikipedias is available for download:
  `https://dumps.wikimedia.org/`

- There are many available guidelines regarding how Wikipedia articles are to be structured, in particular for wiki markup cf.
  `https://en.wikipedia.org/wiki/Help:Wikitext`

- All the content of all Wikipedias is available for download: `https://dumps.wikimedia.org/`

- There are many available guidelines regarding how Wikipedia articles are to be structured, in particular for wiki markup cf. `https://en.wikipedia.org/wiki/Help:Wikitext`

- There are many existing tools devoted to handling wiki markup: eg. the pip package `wikipedia`, the pip package `wptools`, the mediawiki parser from hell pip package or the gensim `corpora.wikicorpus` subpackage.

Outline

Elements of graph theory

Wikipedia is a graph

We can consider that Wikipedia is a graph

Elements of graph theory

Wikipedia is a graph

We can consider that Wikipedia is a graph

- where vertices V are Wikipedia pages

Elements of graph theory

Wikipedia is a graph

We can consider that Wikipedia is a graph

- where vertices V are Wikipedia pages
- and edges $E \subseteq V \times V$ are internal links

Elements of graph theory

Wikipedia is a graph

We can consider that Wikipedia is a graph

- where vertices V are Wikipedia pages
- and edges $E \subseteq V \times V$ are internal links

If a page p_i refers to another page p_j using a internal link, then $\langle v_i, v_j \rangle \in E$: one page ($p_j$) can be accessed from another (p_i), hence one vertex v_i is linked to another (v_j).

Elements of graph theory

Wikipedia is a graph

We can consider that Wikipedia is a graph

- where vertices V are Wikipedia pages
- and edges $E \subseteq V \times V$ are internal links

If a page p_i refers to another page p_j using a internal link, then $\langle v_i, v_j \rangle \in E$: one page ($p_j$) can be accessed from another (p_i), hence one vertex v_i is linked to another (v_j).

More formally:

Elements of graph theory

Wikipedia is a graph

We can consider that Wikipedia is a graph

- where vertices V are Wikipedia pages
- and edges $E \subseteq V \times V$ are internal links

If a page p_i refers to another page p_j using a internal link, then $\langle v_i, v_j \rangle \in E$: one page ($p_j$) can be accessed from another (p_i), hence one vertex v_i is linked to another (v_j).

More formally:

- A graph is a pair $\langle V, E \rangle$ where V is a set of vertices, or nodes, and $E \subseteq V \times V$ is a set of oriented edges linking vertices two by two.

Elements of graph theory

Wikipedia is a graph

We can consider that Wikipedia is a graph

- where vertices V are Wikipedia pages
- and edges $E \subseteq V \times V$ are internal links

If a page p_i refers to another page p_j using a internal link, then $\langle v_i, v_j \rangle \in E$: one page ($p_j$) can be accessed from another (p_i), hence one vertex v_i is linked to another (v_j).

More formally:

- A graph is a pair $\langle V, E \rangle$ where V is a set of vertices, or nodes, and $E \subseteq V \times V$ is a set of oriented edges linking vertices two by two.
- the number of edges $\langle v_x, v_i \rangle$ arriving at a given vertex v_i is called the indegree of v_i

Elements of graph theory

Wikipedia is a graph

We can consider that Wikipedia is a graph

- where vertices V are Wikipedia pages
- and edges $E \subseteq V \times V$ are internal links

If a page p_i refers to another page p_j using a internal link, then $\langle v_i, v_j \rangle \in E$: one page ($p_j$) can be accessed from another (p_i), hence one vertex v_i is linked to another (v_j).

More formally:

- A graph is a pair $\langle V, E \rangle$ where V is a set of vertices, or nodes, and $E \subseteq V \times V$ is a set of oriented edges linking vertices two by two.
- the number of edges $\langle v_x, v_i \rangle$ arriving at a given vertex v_i is called the indegree of v_i
- the number of edges $\langle v_i, v_x \rangle$ starting from a given vertex v_i is called the outdegree of v_i

Elements of graph theory

Wikipedia is a graph

We can consider that Wikipedia is a graph

- where vertices V are Wikipedia pages
- and edges $E \subseteq V \times V$ are internal links

If a page p_i refers to another page p_j using a internal link, then $\langle v_i, v_j \rangle \in E$: one page ($p_j$) can be accessed from another (p_i), hence one vertex v_i is linked to another (v_j).

More formally:

- A graph is a pair $\langle V, E \rangle$ where V is a set of vertices, or nodes, and $E \subseteq V \times V$ is a set of oriented edges linking vertices two by two.
- the number of edges $\langle v_x, v_i \rangle$ arriving at a given vertex v_i is called the indegree of v_i
- the number of edges $\langle v_i, v_x \rangle$ starting from a given vertex v_i is called the outdegree of v_i
- By definition, $\sum_{v_i \in V} \text{indegree}\,(v_i) = \sum_{v_i \in V} \text{outdegree}\,(v_i) = \#E$

Elements of graph theory

It’s a small world after all

Many empirical graphs are ‘small world’ graphs or ‘small world networks’. Intuitively, such graphs have few ‘hubs’ (vertices with very high degrees) and many ‘neighbors’ (vertices connected to a ‘hub’).

Elements of graph theory

It's a small world after all

Many empirical graphs are 'small world' graphs or 'small world networks'. Intuitively, such graphs have few 'hubs' (vertices with very high degrees) and many 'neighbors' (vertices connected to a 'hub').

There's no absolute definition, but we generally employ the following:

Elements of graph theory

It's a small world after all

Many empirical graphs are 'small world' graphs or 'small world networks'. Intuitively, such graphs have few 'hubs' (vertices with very high degrees) and many 'neighbors' (vertices connected to a 'hub').

There's no absolute definition, but we generally employ the following:

- Let $N(v_i) = \{v_n \ : \ \langle v_i, v_n \rangle \ \in \ E\}$, the 'outgoing neighborhood' of node v_i.

Elements of graph theory

It's a small world after all

Many empirical graphs are 'small world' graphs or 'small world networks'. Intuitively, such graphs have few 'hubs' (vertices with very high degrees) and many 'neighbors' (vertices connected to a 'hub').

There's no absolute definition, but we generally employ the following:

- Let $N(v_i) = \{v_n \ : \ \langle v_i, v_n \rangle \ \in \ E\}$, the 'outgoing neighborhood' of node v_i.
- Let $C(v_i) = \frac{\#\{\langle v_n, v_m \rangle \ : \ v_n, v_m \ \in \ N(v_i) \ \wedge \ \langle v_n, v_m \rangle \ \in \ E\}}{\#N(v_i) \ (\#N(v_i) - 1)}$ the 'local clustering coefficient' of node v_i.

Elements of graph theory

It's a small world after all

Many empirical graphs are 'small world' graphs or 'small world networks'. Intuitively, such graphs have few 'hubs' (vertices with very high degrees) and many 'neighbors' (vertices connected to a 'hub').

There's no absolute definition, but we generally employ the following:

- Let $N(v_i) = \{v_n \ : \ \langle v_i, v_n \rangle \ \in \ E\}$, the 'outgoing neighborhood' of node v_i.
- Let $C(v_i) = \frac{\#\{\langle v_n, v_m \rangle \ : \ v_n, v_m \ \in \ N(v_i) \ \wedge \ \langle v_n, v_m \rangle \ \in \ E\}}{\#N(v_i) \ (\#N(v_i) - 1)}$ the 'local clustering coefficient' of node v_i.
- Let the length of the shortest path between v_i and v_j

$$L(v_i, v_j) = \begin{cases} \infty & \textit{if } N(v_i) = \emptyset \\ 1 & \textit{if } \langle v_i, v_j \rangle \in E \\ 1 + \min\limits_{v_n \in N(v_i)} L(v_n, v_j) & \textit{otherwise} \end{cases}$$

Elements of graph theory

It's a small world after all

Many empirical graphs are 'small world' graphs or 'small world networks'. Intuitively, such graphs have few 'hubs' (vertices with very high degrees) and many 'neighbors' (vertices connected to a 'hub').

There's no absolute definition, but we generally employ the following:

- Let $N(v_i) = \{v_n \; : \; \langle v_i, v_n \rangle \; \in \; E\}$, the 'outgoing neighborhood' of node v_i.
- Let $C(v_i) = \frac{\#\{\langle v_n, v_m \rangle \; : \; v_n, v_m \; \in \; N(v_i) \; \wedge \; \langle v_n, v_m \rangle \; \in \; E\}}{\#N(v_i) \; (\#N(v_i) - 1)}$ the 'local clustering coefficient' of node v_i.
- Let the length of the shortest path between v_i and v_j

$$L(v_i, v_j) = \begin{cases} \infty & \textit{if } N(v_i) = \emptyset \\ 1 & \textit{if } \langle v_i, v_j \rangle \in E \\ 1 + \min\limits_{v_n \in N(v_i)} L(v_n, v_j) & \textit{otherwise} \end{cases}$$

- If $\bar{L} \propto \log \#V$ and $\bar{C}$ is sufficiently large ($>$ to what a random graph with the same number of vertices would have), the graph is 'small-word'.

Elements of graph theory

It's a small world after all

Many empirical graphs are 'small world' graphs or 'small world networks'. Intuitively, such graphs have few 'hubs' (vertices with very high degrees) and many 'neighbors' (vertices connected to a 'hub').

There's no absolute definition, but we generally employ the following:

- Let $N(v_i) = \{v_n \ : \ \langle v_i, v_n \rangle \ \in \ E\}$, the 'outgoing neighborhood' of node v_i.
- Let $C(v_i) = \frac{\#\{\langle v_n, v_m \rangle \ : \ v_n, v_m \ \in \ N(v_i) \ \wedge \ \langle v_n, v_m \rangle \ \in \ E\}}{\#N(v_i) \ (\#N(v_i) - 1)}$ the 'local clustering coefficient' of node v_i.
- Let the length of the shortest path between v_i and v_j

$$L(v_i, v_j) = \begin{cases} \infty & \textit{if } N(v_i) = \emptyset \\ 1 & \textit{if } \langle v_i, v_j \rangle \in E \\ 1 + \min\limits_{v_n \in N(v_i)} L(v_n, v_j) & \textit{otherwise} \end{cases}$$

- If $\bar{L} \propto \log \#V$ and $\bar{C}$ is sufficiently large ($>$ to what a random graph with the same number of vertices would have), the graph is 'small-word'.

One can think of 'small-world' structures as the analog for graphs of 'Zipf's Law' for word frequencies.

Elements of graph theory

It's a small world after all

Many empirical graphs are 'small world' graphs or 'small world networks'. Intuitively, such graphs have few 'hubs' (vertices with very high degrees) and many 'neighbors' (vertices connected to a 'hub').

There's no absolute definition, but we generally employ the following:

- Let $N(v_i) = \{v_n \; : \; \langle v_i, v_n \rangle \; \in \; E\}$, the 'outgoing neighborhood' of node v_i.
- Let $C(v_i) = \frac{\#\{\langle v_n, v_m \rangle \; : \; v_n, v_m \; \in \; N(v_i) \; \wedge \; \langle v_n, v_m \rangle \; \in \; E\}}{\#N(v_i) \; (\#N(v_i) - 1)}$ the 'local clustering coefficient' of node v_i.
- Let the length of the shortest path between v_i and v_j

  $$L(v_i, v_j) = \begin{cases} \infty & \textit{if } N(v_i) = \emptyset \\ 1 & \textit{if } \langle v_i, v_j \rangle \in E \\ 1 + \min\limits_{v_n \in N(v_i)} L(v_n, v_j) & \textit{otherwise} \end{cases}$$

- If $\bar{L} \propto \log \#V$ and $\bar{C}$ is sufficiently large ($>$ to what a random graph with the same number of vertices would have), the graph is 'small-word'.

One can think of 'small-world' structures as the analog for graphs of 'Zipf's Law' for word frequencies.

See also: `https://en.wikipedia.org/wiki/Wikipedia:Getting_to_Philosophy`

Outline

The main usage of encyclopedias at large is to ground NLP applications into the real world, by making use of the facts that they collect.

The main usage of encyclopedias at large is to ground NLP applications into the real world, by making use of the facts that they collect.

- Typically, an important branch of research has been dedicated in converting online encyclopedias (and Wikipedia in particular) into ‘ontologies’.

Example applications of Wikipedia

Some applications of Wikipedia

The main usage of encyclopedias at large is to ground NLP applications into the real world, by making use of the facts that they collect.

- Typically, an important branch of research has been dedicated in converting online encyclopedias (and Wikipedia in particular) into ‘ontologies’.
- Other usages include deriving templates and models for converting raw data (attribute-values pairs or raw numerical data) into natural-language text.

Example applications of Wikipedia

Inducing ontologies

Ontologies are systematic definitions for classifying and structuring data. ‘Ontology learning’ is the task that focuses on inferring such systematic classifications from unstructured or partially structured data.

Example applications of Wikipedia

Inducing ontologies

Ontologies are systematic definitions for classifying and structuring data. ‘Ontology learning’ is the task that focuses on inferring such systematic classifications from unstructured or partially structured data.

A generally well established format for Ontologies is the Resource Description Framework proposed by the W3C.

Example applications of Wikipedia

Inducing ontologies

Ontologies are systematic definitions for classifying and structuring data. 'Ontology learning' is the task that focuses on inferring such systematic classifications from unstructured or partially structured data.

A generally well established format for Ontologies is the Resource Description Framework proposed by the W3C.

- More specifically restricted to Wikipedia, the DBPedia initiative aims at extracting factual relations from Wikimedia pages and converting those to the RDF format.

Example applications of Wikipedia

Inducing ontologies

Ontologies are systematic definitions for classifying and structuring data. 'Ontology learning' is the task that focuses on inferring such systematic classifications from unstructured or partially structured data.

A generally well established format for Ontologies is the Resource Description Framework proposed by the W3C.

- More specifically restricted to Wikipedia, the DBPedia initiative aims at extracting factual relations from Wikimedia pages and converting those to the RDF format.
- **Vast enterprise**: "1,445,000 persons, 735,000 places (including 478,000 populated places), 411,000 creative works (including 123,000 music albums, 87,000 films and 19,000 video games), 241,000 organizations (including 58,000 companies and 49,000 educational institutions), 251,000 species and 6,000 diseases."

Example applications of Wikipedia

Data-to-text NLG

Natural language generation (‘NLG’) anchored into factual data is one of the ‘holy grails’ of NLP.

Example applications of Wikipedia

Data-to-text NLG

Natural language generation (‘NLG’) anchored into factual data is one of the ‘holy grails’ of NLP.

Using Wikipedia:

- Retrieve the attribute-value pairs listed in the infobox

Example applications of Wikipedia

Data-to-text NLG

Natural language generation ('NLG') anchored into factual data is one of the 'holy grails' of NLP.

Using Wikipedia:

- Retrieve the attribute-value pairs listed in the infobox
- Transform these pairs into vector representations:
  - generally speaking, you would need an embedding-based biaffine function $\mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^{d'}$: $\vec{a} = e(a)$; $\vec{v} = e(v)$; $\vec{r} = B(\vec{a}, \vec{v})$ with a the attribute, v the value and B the learned composition function.

Example applications of Wikipedia

Data-to-text NLG

Natural language generation ('NLG') anchored into factual data is one of the 'holy grails' of NLP.

Using Wikipedia:

- Retrieve the attribute-value pairs listed in the infobox
- Transform these pairs into vector representations:
  - generally speaking, you would need an embedding-based biaffine function $\mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^{d'}$: $\vec{a} = e(a)$; $\vec{v} = e(v)$; $\vec{r} = B(\vec{a}, \vec{v})$ with a the attribute, v the value and B the learned composition function.
  - in practice, a simple concatenation followed by a linear layer could suffice.

Example applications of Wikipedia

Data-to-text NLG

Natural language generation (‘NLG’) anchored into factual data is one of the ‘holy grails’ of NLP.

Using Wikipedia:

- Retrieve the attribute-value pairs listed in the infobox
- Transform these pairs into vector representations:
  - generally speaking, you would need an embedding-based biaffine function $\mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^{d'} : \vec{a} = e(a);\ \vec{v} = e(v);\ \vec{r} = B(\vec{a}, \vec{v})$ with a the attribute, v the value and B the learned composition function.
  - in practice, a simple concatenation followed by a linear layer could suffice.
  - another way of tackling this would be using copy pointer mechanisms.

Example applications of Wikipedia

Data-to-text NLG

Natural language generation ('NLG') anchored into factual data is one of the 'holy grails' of NLP.

Using Wikipedia:

- Retrieve the attribute-value pairs listed in the infobox
- Transform these pairs into vector representations:
  - generally speaking, you would need an embedding-based biaffine function $\mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^{d'} : \vec{a} = e(a);\ \vec{v} = e(v);\ \vec{r} = B(\vec{a}, \vec{v})$ with a the attribute, v the value and B the learned composition function.
  - in practice, a simple concatenation followed by a linear layer could suffice.
  - another way of tackling this would be using copy pointer mechanisms.
- These representations can be thought of as a **source**, and the lead section as a **target**.
  - Following that, any sequence-to-sequence algorithm can function.

Example applications of Wikipedia

Data-to-text NLG

Natural language generation ('NLG') anchored into factual data is one of the 'holy grails' of NLP.

Using Wikipedia:

- Retrieve the attribute-value pairs listed in the infobox
- Transform these pairs into vector representations:
  - generally speaking, you would need an embedding-based biaffine function $\mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^{d'}$: $\vec{a} = e(a)$; $\vec{v} = e(v)$; $\vec{r} = B(\vec{a}, \vec{v})$ with a the attribute, v the value and B the learned composition function.
  - in practice, a simple concatenation followed by a linear layer could suffice.
  - another way of tackling this would be using copy pointer mechanisms.
- These representations can be thought of as a **source**, and the lead section as a **target**.
  - Following that, any sequence-to-sequence algorithm can function.
  - Attention-based models in particular will prove useful, as they should allow you to use the multiple source representations efficiently.

Outline

Basic concepts

Working with dictionary: Definitions

- a dictionary is a collection of definition

Here's a concrete example:

- if the definition is `random: lacking a definite plan, purpose, or pattern`

Basic concepts

Working with dictionary: Definitions

- a dictionary is a collection of definition
- a definition couples a definiendum with definientia

Here's a concrete example:

- if the definition is `random: lacking a definite plan, purpose, or pattern`

Basic concepts

Working with dictionary: Definitions

- a dictionary is a collection of definition
- a definition couples a definiendum with definientia
- the definiendum (pl. definienda) is the word for which a definition is given

Here's a concrete example:

- if the definition is `random: lacking a definite plan, purpose, or pattern`
- the definiendum is `random`

Basic concepts

Working with dictionary: Definitions

- a dictionary is a collection of definition
- a definition couples a definiendum with definientia
- the definiendum (pl. definienda) is the word for which a definition is given
- a definiens (pl. definientia) is a word that occurs in order to define a definiendum

Here's a concrete example:

- if the definition is `random: lacking a definite plan, purpose, or pattern`
- the definiendum is `random`
- the definientia are `lacking`, `a`, `definite`, `plan`, `,`, `purpose`, `,`, `or`, `pattern`

Basic concepts

Working with dictionary: Definitions

- a dictionary is a collection of definition
- a definition couples a definiendum with definientia
- the definiendum (pl. definienda) is the word for which a definition is given
- a definiens (pl. definientia) is a word that occurs in order to define a definiendum
- technically, one definition corresponds to one sense.

Here's a concrete example:

- if the definition is `random: lacking a definite plan, purpose, or pattern`
- the definiendum is `random`
- the definientia are `lacking`, `a`, `definite`, `plan`, `,`, `purpose`, `,`, `or`, `pattern`
- there might be other definitions for the definiendum `random`, eg. `relating to, having, or being elements or events with definite probability of occurrence`

Basic concepts

Working with dictionary: Definitions

- definitions (more precisely, definientia) are generally organized according to a 'Genus + Differentium' pattern

- If the definition is `phosphorescent: emitting light without appreciable heat as by slow oxidation of phosphorous`...

Basic concepts

Working with dictionary: Definitions

- definitions (more precisely, definientia) are generally organized according to a 'Genus + Differentium' pattern
- the Genus (pl. genera) corresponds to a definiens (or a series of definientia) that gives the broad semantic category of the definiendum

- If the definition is `phosphorescent: emitting light without appreciable heat as by slow oxidation of phosphorous`...
  - the genus is `emitting light`

Basic concepts

Working with dictionary: Definitions

- definitions (more precisely, definientia) are generally organized according to a 'Genus + Differentium' pattern
- the Genus (pl. genera) corresponds to a definiens (or a series of definientia) that gives the broad semantic category of the definiendum
- the Differentium (pl. differentia) corresponds to a definiens (or a series of definientia) that give the finer semantics within the broad category that is the Genus

- If the definition is `phosphorescent: emitting light without appreciable heat as by slow oxidation of phosphorous`...
  - the genus is `emitting light`
  - the differentia are `without appreciable heat as by slow oxidation of phosphorous`

Basic concepts

Working with dictionary: Definitions

- definitions (more precisely, definientia) are generally organized according to a 'Genus + Differentium' pattern
- the Genus (pl. genera) corresponds to a definiens (or a series of definientia) that gives the broad semantic category of the definiendum
- the Differentium (pl. differentia) corresponds to a definiens (or a series of definientia) that give the finer semantics within the broad category that is the Genus
- the genus is *generally* a hypernym of the definiendum.

- If the definition is `phosphorescent: emitting light without appreciable heat as by slow oxidation of phosphorous`...
  - the genus is `emitting light`
  - the differentia are `without appreciable heat as by slow oxidation of phosphorous`

Basic concepts

Working with dictionary: Definitions

- definitions (more precisely, definientia) are generally organized according to a 'Genus + Differentium' pattern
- the Genus (pl. genera) corresponds to a definiens (or a series of definientia) that gives the broad semantic category of the definiendum
- the Differentium (pl. differentia) corresponds to a definiens (or a series of definientia) that give the finer semantics within the broad category that is the Genus
- the genus is *generally* a hypernym of the definiendum.
- Not all definitions follow this pattern.

- If the definition is `phosphorescent: emitting light without appreciable heat as by slow oxidation of phosphorous`...
  - the genus is `emitting light`
  - the differentia are `without appreciable heat as by slow oxidation of phosphorous`

Basic concepts

Working with dictionary: Definitions

- definitions (more precisely, definientia) are generally organized according to a 'Genus + Differentium' pattern
- the Genus (pl. genera) corresponds to a definiens (or a series of definientia) that gives the broad semantic category of the definiendum
- the Differentium (pl. differentia) corresponds to a definiens (or a series of definientia) that give the finer semantics within the broad category that is the Genus
- the genus is *generally* a hypernym of the definiendum.
- Not all definitions follow this pattern. Other patterns include: synonymy,

- If the definition is `phosphorescent: emitting light without appreciable heat as by slow oxidation of phosphorous`...
  - the genus is `emitting light`
  - the differentia are `without appreciable heat as by slow oxidation of phosphorous`
- Synonymy-based definition: `eradication: extirpation`.

Basic concepts

Working with dictionary: Definitions

- definitions (more precisely, definientia) are generally organized according to a 'Genus + Differentium' pattern
- the Genus (pl. genera) corresponds to a definiens (or a series of definientia) that gives the broad semantic category of the definiendum
- the Differentium (pl. differentia) corresponds to a definiens (or a series of definientia) that give the finer semantics within the broad category that is the Genus
- the genus is *generally* a hypernym of the definiendum.
- Not all definitions follow this pattern. Other patterns include: synonymy, meta-linguistic reference...

- If the definition is `phosphorescent: emitting light without appreciable heat as by slow oxidation of phosphorous`...
  - the genus is `emitting light`
  - the differentia are `without appreciable heat as by slow oxidation of phosphorous`
- Synonymy-based definition: `eradication: extirpation`.
- Meta-linguistic reference: `guardians: plural of "guardian"`.

Outline

Example applications of dictionaries

WSD & WSI

- WSD stands for ‘word sense disambiguation’. Given an ambiguous word in context and an inventory of meanings, find which meaning applies to the word.

Example applications of dictionaries

WSD & WSI

- WSD stands for ‘word sense disambiguation’. Given an ambiguous word in context and an inventory of meanings, find which meaning applies to the word.
- WSI for ‘word sense induction’. Given an ambigous word in context, find its meaning (without a precomputed inventory).

Example applications of dictionaries

WSD & WSI

- WSD stands for ‘word sense disambiguation’. Given an ambiguous word in context and an inventory of meanings, find which meaning applies to the word.
- WSI for ‘word sense induction’. Given an ambigous word in context, find its meaning (without a precomputed inventory).

Dictionaries can be used either as an inventory of meanings for WSD, or as labels to learn a WSI architecture in a supervised setup.

Example applications of dictionaries

WSD & WSI

- WSD stands for ‘word sense disambiguation’. Given an ambiguous word in context and an inventory of meanings, find which meaning applies to the word.
- WSI for ‘word sense induction’. Given an ambigous word in context, find its meaning (without a precomputed inventory).

Dictionaries can be used either as an inventory of meanings for WSD, or as labels to learn a WSI architecture in a supervised setup.

- in those WSI setups, the basic idea is to transform definitions into labels or targets.

Example applications of dictionaries

WSD & WSI

- WSD stands for ‘word sense disambiguation’. Given an ambiguous word in context and an inventory of meanings, find which meaning applies to the word.
- WSI for ‘word sense induction’. Given an ambigous word in context, find its meaning (without a precomputed inventory).

Dictionaries can be used either as an inventory of meanings for WSD, or as labels to learn a WSI architecture in a supervised setup.

- in those WSI setups, the basic idea is to transform definitions into labels or targets.

- (Chang and Chen, 2019): a ‘simple’ way is to convert the definientia into a single sentence embedding, and then learn the mapping from the (contextual) embedding onto the definientia embedding.

Example applications of dictionaries

‘Old school’ WSD

(Gaume, Hathout, and Muller, 2004)

The basic idea is to turn the dictionary itself into a graph, and explore the graph structure

Example applications of dictionaries

'Old school' WSD

(Gaume, Hathout, and Muller, 2004)

The basic idea is to turn the dictionary itself into a graph, and explore the graph structure

- We noted that genera are hypernyms of their respective definienda

Example applications of dictionaries

'Old school' WSD

(Gaume, Hathout, and Muller, 2004)

The basic idea is to turn the dictionary itself into a graph, and explore the graph structure

- We noted that genera are hypernyms of their respective definienda
- Therefore, by transitivity, the genus of the genus of word *w* should also be a hypernym of *w*

Example applications of dictionaries

'Old school' WSD

(Gaume, Hathout, and Muller, 2004)

The basic idea is to turn the dictionary itself into a graph, and explore the graph structure

- We noted that genera are hypernyms of their respective definienda
- Therefore, by transitivity, the genus of the genus of word w should also be a hypernym of w
- More generally, if a definiens contributes to the meaning of its definiendum w ,then the definiens of a definiens contributes indirectly to the meaning of w. We can think of it as "taking one more step" in the graph linking definienda to their definientia.

Example applications of dictionaries

‘Old school’ WSD

- Let a set D of definitions $\langle w^i, \langle w^i_1 \ \ldots \ w^i_n \rangle\rangle$, where w^i is a definiendum and $\langle w^i_1 \ \ldots \ w^i_n \rangle$ associated definientia. Let the vocabulary $V = \bigcup\limits_{\langle w^i, d_{w^i} \rangle \in D} \left(\{w^i\} \cup d_{w^i} \right)$

Example applications of dictionaries

‘Old school’ WSD

- Let a set D of definitions $\langle w^i, \langle w^i_1 \ \ldots \ w^i_n \rangle\rangle$, where w^i is a definiendum and $\langle w^i_1 \ \ldots \ w^i_n \rangle$ associated definientia. Let the vocabulary $V = \bigcup_{\langle w^i, d_{w^i} \rangle \in D} \left(\{w^i\} \cup d_{w^i} \right)$

- Let the graph $\mathscr{G}$ defined over $V \cup D \times V \cup D$, such as $\langle w^i, d^j \rangle$ and $\langle d^j, w^k \rangle$ are edges of $\mathscr{G}$ iff. $\exists d^j$ such as $d^j = \langle w^i, \langle w^i_1 \ \ldots w^k \ldots \ w^i_n \rangle\rangle \in D$.

Example applications of dictionaries

'Old school' WSD

- Let a set D of definitions $\langle w^i, \langle w^i_1 \ \ldots \ w^i_n \rangle\rangle$, where w^i is a definiendum and $\langle w^i_1 \ \ldots \ w^i_n \rangle$ associated definientia. Let the vocabulary $V = \bigcup_{\langle w^i, d_{w^i} \rangle \in D} \left(\{w^i\} \cup d_{w^i} \right)$
- Let the graph $\mathscr{G}$ defined over $V \cup D \times V \cup D$, such as $\langle w^i, d^j \rangle$ and $\langle d^j, w^k \rangle$ are edges of $\mathscr{G}$ iff. $\exists d^j$ such as $d^j = \langle w^i, \langle w^i_1 \ \ldots w^k \ldots \ w^i_n \rangle\rangle \in D$. Let the connectivity $|V \cup D| \times |V \cup D|$ matrix $\hat{\mathscr{G}}$ the matrix notation of $\mathscr{G}$, where $\hat{\mathscr{G}}_{ij} = 1$ iff. $\langle v^i, v^j \rangle$ is an edge of $\mathscr{G}$.

Example applications of dictionaries

'Old school' WSD

- Let a set D of definitions $\langle w^i, \langle w^i_1 \; \dots \; w^i_n \rangle \rangle$, where w^i is a definiendum and $\langle w^i_1 \; \dots \; w^i_n \rangle$ associated definientia. Let the vocabulary $V = \bigcup_{\langle w^i, d_{w^i} \rangle \in D} \left(\{w^i\} \cup d_{w^i} \right)$
- Let the graph $\mathscr{G}$ defined over $V \cup D \times V \cup D$, such as $\langle w^i, d^j \rangle$ and $\langle d^j, w^k \rangle$ are edges of $\mathscr{G}$ iff. $\exists d^j$ such as $d^j = \langle w^i, \langle w^i_1 \; \dots w^k \dots \; w^i_n \rangle \rangle \in D$. Let the connectivity $|V \cup D| \times |V \cup D|$ matrix $\hat{\mathscr{G}}$ the matrix notation of $\mathscr{G}$, where $\hat{\mathscr{G}}_{ij} = 1$ iff. $\langle v^i, v^j \rangle$ is an edge of $\mathscr{G}$.
- Let the matrix $\tilde{\mathscr{G}}$ the normalized version of $\hat{\mathscr{G}}$. Hence all rows of $\tilde{\mathscr{G}}$ sum up to 1 and can be seen as probability distributions.
  - If the row corresponds to a definition d, it defines how likely words from V are its definientia.

Example applications of dictionaries

'Old school' WSD

- Let a set D of definitions $\langle w^i, \langle w^i_1 \; \dots \; w^i_n \rangle\rangle$, where w^i is a definiendum and $\langle w^i_1 \; \dots \; w^i_n \rangle$ associated definientia. Let the vocabulary $V = \bigcup_{\langle w^i, d_{w^i}\rangle \in D} \left(\{w^i\} \cup d_{w^i}\right)$
- Let the graph $\mathcal{G}$ defined over $V \cup D \times V \cup D$, such as $\langle w^i, d^j \rangle$ and $\langle d^j, w^k \rangle$ are edges of $\mathcal{G}$ iff. $\exists d^j$ such as $d^j = \langle w^i, \langle w^i_1 \; \dots w^k \dots \; w^i_n \rangle\rangle \in D$. Let the connectivity $|V \cup D| \times |V \cup D|$ matrix $\hat{\mathcal{G}}$ the matrix notation of $\mathcal{G}$, where $\hat{\mathcal{G}}_{ij} = 1$ iff. $\langle v^i, v^j \rangle$ is an edge of $\mathcal{G}$.
- Let the matrix $\tilde{\mathcal{G}}$ the normalized version of $\hat{\mathcal{G}}$. Hence all rows of $\tilde{\mathcal{G}}$ sum up to 1 and can be seen as probability distributions.
  - If the row corresponds to a definition d, it defines how likely words from V are its definientia.
  - If the row corresponds to a word w, it simply links to its possible definitions in D.

Example applications of dictionaries

‘Old school’ WSD

- Let a set D of definitions $\langle w^i, \langle w^i_1 \ \dots \ w^i_n \rangle\rangle$, where w^i is a definiendum and $\langle w^i_1 \ \dots \ w^i_n \rangle$ associated definientia. Let the vocabulary $V = \bigcup_{\langle w^i, d_{w^i} \rangle \in D} \left(\{w^i\} \cup d_{w^i} \right)$

- Let the graph $\mathscr{G}$ defined over $V \cup D \times V \cup D$, such as $\langle w^i, d^j \rangle$ and $\langle d^j, w^k \rangle$ are edges of $\mathscr{G}$ iff. $\exists d^j$ such as $d^j = \langle w^i, \langle w^i_1 \ \dots w^k \dots \ w^i_n \rangle\rangle \in D$. Let the connectivity $|V \cup D| \times |V \cup D|$ matrix $\hat{\mathscr{G}}$ the matrix notation of $\mathscr{G}$, where $\hat{\mathscr{G}}_{ij} = 1$ iff. $\langle v^i, v^j \rangle$ is an edge of $\mathscr{G}$.

- Let the matrix $\tilde{\mathscr{G}}$ the normalized version of $\hat{\mathscr{G}}$. Hence all rows of $\tilde{\mathscr{G}}$ sum up to 1 and can be seen as probability distributions.
  - If the row corresponds to a definition d, it defines how likely words from V are its definientia.
  - If the row corresponds to a word w, it simply links to its possible definitions in D.

- We can transcribe hypernymy transitivity using the chain rule

Example applications of dictionaries

'Old school' WSD

- Let a set D of definitions $\langle w^i, \langle w^i_1 \; \dots \; w^i_n \rangle \rangle$, where w^i is a definiendum and $\langle w^i_1 \; \dots \; w^i_n \rangle$ associated definientia. Let the vocabulary $V = \bigcup_{\langle w^i, d_{w^i} \rangle \in D} \left(\{w^i\} \cup d_{w^i} \right)$

- Let the graph $\mathscr{G}$ defined over $V \cup D \times V \cup D$, such as $\langle w^i, d^j \rangle$ and $\langle d^j, w^k \rangle$ are edges of $\mathscr{G}$ iff. $\exists d^j$ such as $d^j = \langle w^i, \langle w^i_1 \; \dots w^k \dots \; w^i_n \rangle \rangle \in D$. Let the connectivity $|V \cup D| \times |V \cup D|$ matrix $\hat{\mathscr{G}}$ the matrix notation of $\mathscr{G}$, where $\hat{\mathscr{G}}_{ij} = 1$ iff. $\langle v^i, v^j \rangle$ is an edge of $\mathscr{G}$.

- Let the matrix $\tilde{\mathscr{G}}$ the normalized version of $\hat{\mathscr{G}}$. Hence all rows of $\tilde{\mathscr{G}}$ sum up to 1 and can be seen as probability distributions.
  - If the row corresponds to a definition d, it defines how likely words from V are its definientia.
  - If the row corresponds to a word w, it simply links to its possible definitions in D.

- We can transcribe hypernymy transitivity using the chain rule: the probability that word w^j occurs within k steps as a definiens for the definition d^i will be given by $\left(\tilde{\mathscr{G}}^k \right)_{ij}$.

Example applications of dictionaries

'Old school' WSD

- Let a set D of definitions $\langle w^i, \langle w_1^i \;\ldots\; w_n^i \rangle\rangle$, where w^i is a definiendum and $\langle w_1^i \;\ldots\; w_n^i \rangle$ associated definientia. Let the vocabulary $V = \bigcup_{\langle w^i, d_{w^i} \rangle \in D} \left(\{w^i\} \cup d_{w^i}\right)$
- Let the graph $\mathscr{G}$ defined over $V \cup D \times V \cup D$, such as $\langle w^i, d^j \rangle$ and $\langle d^j, w^k \rangle$ are edges of $\mathscr{G}$ iff. $\exists d^j$ such as $d^j = \langle w^i, \langle w_1^i \;\ldots w^k \ldots\; w_n^i \rangle\rangle \in D$. Let the connectivity $|V \cup D| \times |V \cup D|$ matrix $\hat{\mathscr{G}}$ the matrix notation of $\mathscr{G}$, where $\hat{\mathscr{G}}_{ij} = 1$ iff. $\langle v^i, v^j \rangle$ is an edge of $\mathscr{G}$.
- Let the matrix $\tilde{\mathscr{G}}$ the normalized version of $\hat{\mathscr{G}}$. Hence all rows of $\tilde{\mathscr{G}}$ sum up to 1 and can be seen as probability distributions.
  - If the row corresponds to a definition d, it defines how likely words from V are its definientia.
  - If the row corresponds to a word w, it simply links to its possible definitions in D.
- We can transcribe hypernymy transitivity using the chain rule: the probability that word w^j occurs within k steps as a definiens for the definition d^i will be given by $\left(\tilde{\mathscr{G}}^k\right)_{ij}$.

Disambiguate a word token w_α by retrieving the node for one of its definition d^j whose probability distribution, according to $\left(\tilde{\mathscr{G}}^k\right)$, matches best with the context of w_α.

Example applications of dictionaries

Crossword solver

(Hill et al., 2016): You can use definitions as meaning inventories, so you can learn to retrieve words based on what how they are defined, which may be useful for Information Retrieval

Example applications of dictionaries

Crossword solver

(Hill et al., 2016): You can use definitions as meaning inventories, so you can learn to retrieve words based on what how they are defined, which may be useful for Information Retrieval

- as a proof of concept, a crossword solver

Example applications of dictionaries

Crossword solver

(Hill et al., 2016): You can use definitions as meaning inventories, so you can learn to retrieve words based on what how they are defined, which may be useful for Information Retrieval

- as a proof of concept, a crossword solver
- train a model to retrieve or recompute the embedding of the definiendum, based on the sequence of definientia

Example applications of dictionaries

Crossword solver

(Hill et al., 2016): You can use definitions as meaning inventories, so you can learn to retrieve words based on what how they are defined, which may be useful for Information Retrieval

- as a proof of concept, a crossword solver
- train a model to retrieve or recompute the embedding of the definiendum, based on the sequence of definientia
- test on crosswords: for each cross word definition, select the most probable embedding that fits in the grid

Example applications of dictionaries

Definition Modeling

(Noraset et al., 2017): can you evaluate word embeddings content using a dictionary? You can see it as the inverse function of the previous crossword solver

Example applications of dictionaries

Definition Modeling

(Noraset et al., 2017): can you evaluate word embeddings content using a dictionary? You can see it as the inverse function of the previous crossword solver

- Given a set of embeddings $E = \vec{w}^1 \ \ldots \ \vec{w}^n$, we want to assert that the vector representations adequately capture the meaning of the words they are purported to represent

Example applications of dictionaries

Definition Modeling

(Noraset et al., 2017): can you evaluate word embeddings content using a dictionary? You can see it as the inverse function of the previous crossword solver

- Given a set of embeddings $E = \vec{w}^1 \ \ldots \ \vec{w}^n$, we want to assert that the vector representations adequately capture the meaning of the words they are purported to represent
- If $\vec{w}^i$ is a good representation of w^i, then the dictionary definition of w^i can be reconstructed from $\vec{w}^i$

Example applications of dictionaries

Definition Modeling

(Noraset et al., 2017): can you evaluate word embeddings content using a dictionary? You can see it as the inverse function of the previous crossword solver

- Given a set of embeddings $E = \vec{w^1} \ \ldots \ \vec{w^n}$, we want to assert that the vector representations adequately capture the meaning of the words they are purported to represent
- If $\vec{w^i}$ is a good representation of w^i, then the dictionary definition of w^i can be reconstructed from $\vec{w^i}$
- How can you test that? Initialize a recurrent neural network with $\vec{w^i}$, and have it produce the corresponding definientia $\langle w^i_1 \ \ldots \ w^i_n \rangle$.

Example applications of dictionaries

Definition Modeling

(Noraset et al., 2017): can you evaluate word embeddings content using a dictionary? You can see it as the inverse function of the previous crossword solver

- Given a set of embeddings $E = \vec{w^1} \ \ldots \ \vec{w^n}$, we want to assert that the vector representations adequately capture the meaning of the words they are purported to represent

- If $\vec{w^i}$ is a good representation of w^i, then the dictionary definition of w^i can be reconstructed from $\vec{w^i}$

- How can you test that? Initialize a recurrent neural network with $\vec{w^i}$, and have it produce the corresponding definientia $\langle w^i_1 \ \ldots \ w^i_n \rangle$.

- You can see that as a basic sequence-to-sequence task, mapping $\langle w^i \rangle$ to $\langle w^i_1 \ \ldots \ w^i_n \rangle$

Example applications of dictionaries

Definition Modeling

(Noraset et al., 2017): can you evaluate word embeddings content using a dictionary? You can see it as the inverse function of the previous crossword solver

- Given a set of embeddings $E = \vec{w}^1 \ \dots \ \vec{w}^n$, we want to assert that the vector representations adequately capture the meaning of the words they are purported to represent

- If $\vec{w}^i$ is a good representation of w^i, then the dictionary definition of w^i can be reconstructed from $\vec{w}^i$

- How can you test that? Initialize a recurrent neural network with $\vec{w}^i$, and have it produce the corresponding definientia $\langle w^i_1 \ \dots \ w^i_n \rangle$.

- You can see that as a basic sequence-to-sequence task, mapping $\langle w^i \rangle$ to $\langle w^i_1 \ \dots \ w^i_n \rangle$

- From this basic formulation, you can derive representation for words in context, simply by extending the set containing the definiendum with its context words.

Outline

Working with Wiki

A closer look at Wiktionary

- Wiktionary is a collaboratively edited multilingual web-based project.

Working with Wiki

A closer look at Wiktionary

- Wiktionary is a collaboratively edited multilingual web-based project.
- The aim is to produce dictionaries for all the world's languages, currently it covers $\approx$ 171 languages

Working with Wiki

A closer look at Wiktionary

- Wiktionary is a collaboratively edited multilingual web-based project.
- The aim is to produce dictionaries for all the world's languages, currently it covers $\approx$ 171 languages
- Wiktionary data is frequently used in NLP, both in multilingual and in monolingual contexts

Working with Wiki

A closer look at Wiktionary

- Wiktionary is a collaboratively edited multilingual web-based project.
- The aim is to produce dictionaries for all the world's languages, currently it covers ≈ 171 languages
- Wiktionary data is frequently used in NLP, both in multilingual and in monolingual contexts
- As a consequence of the collaborative nature of the project, Wiktionary is generally deemed to have broad coverage, but unsystematic definitions.

Working with Wiki

"It's complicated"

- Wiki-markup is notably messy. As a potpourri to give you a taste: includes recursive markups, implicit section and subsection handling, template and boilerplate elements...

Working with Wiki

“It’s complicated”

- Wiki-markup is notably messy. As a potpourri to give you a taste: includes recursive markups, implicit section and subsection handling, template and boilerplate elements...

- Wiki-markup is doubly nightmarish for Wiktionaries across languages, which generally do not have the same conventions

Working with Wiki

"It's complicated"

- Wiki-markup is notably messy. As a potpourri to give you a taste: includes recursive markups, implicit section and subsection handling, template and boilerplate elements...
- Wiki-markup is doubly nightmarish for Wiktionaries across languages, which generally do not have the same conventions
- The raw data can be found here: `https://dumps.wikimedia.org/`

Working with Wiki

“It’s complicated”

- Wiki-markup is notably messy. As a potpourri to give you a taste: includes recursive markups, implicit section and subsection handling, template and boilerplate elements...

- Wiki-markup is doubly nightmarish for Wiktionaries across languages, which generally do not have the same conventions

- The raw data can be found here: `https://dumps.wikimedia.org/`

- But parsing it correctly takes time...

Working with Wiki

Solutions exist.

- Some languages have NLP-friendly wiktionary, or wiktionary ports

Working with Wiki

Solutions exist.

- Some languages have NLP-friendly wiktionary, or wiktionary ports
- A very good example is GLAWI, a French wiktionary dump parsed in a XML format, developed by CLLE-ERSS (Toulouse)

Working with Wiki

Solutions exist.

- Some languages have NLP-friendly wiktionary, or wiktionary ports
- A very good example is GLAWI, a French wiktionary dump parsed in a XML format, developed by CLLE-ERSS (Toulouse)
- Another such example is DBnary, which has coverage for multiple languages (mostly European, but also Indonesian, Japanese, Turkish); intended as a SPARQL endpoint for a webservice, it requires a bit of elbow grease to handle

Working with Wiki

Solutions exist.

- Some languages have NLP-friendly wiktionary, or wiktionary ports
- A very good example is GLAWI, a French wiktionary dump parsed in a XML format, developed by CLLE-ERSS (Toulouse)
- Another such example is DBnary, which has coverage for multiple languages (mostly European, but also Indonesian, Japanese, Turkish); intended as a SPARQL endpoint for a webservice, it requires a bit of elbow grease to handle
- There are some libraries dedicated to wiki-markup handling, so some (or a lot) of dedication using these may suffice to retrieve the information you care about.

References I

Chang, Ting-Yun and Yun-Nung Chen (Nov. 2019). "What Does This Word Mean? Explaining Contextualized Embeddings with Natural Language Definition". In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Hong Kong, China: Association for Computational Linguistics, pp. 6066–6072. DOI: `10.18653/v1/D19-1627`. URL: `https://www.aclweb.org/anthology/D19-1627`.

Gaume, Bruno, Nabil Hathout, and Philippe Muller (2004). "Word Sense Disambiguation Using a Dictionary for Sense Similarity Measure". In: *Proceedings of the 20th International Conference on Computational Linguistics*. COLING '04. Geneva, Switzerland: Association for Computational Linguistics. DOI: `10.3115/1220355.1220528`. URL: `https://doi.org/10.3115/1220355.1220528`.

Hill, Felix et al. (2016). "Learning to Understand Phrases by Embedding the Dictionary". In: *Transactions of the Association for Computational Linguistics* 4, pp. 17–30. ISSN: 2307-387X. URL: `https://transacl.org/ojs/index.php/tacl/article/view/711`.

Noraset, Thanapon et al. (2017). "Definition Modeling: Learning to define word embeddings in natural language". In: *AAAI*.