

# Troubleshooting and Optimizing Named Entity Resolution Systems in the Industry

*Panos Alexopoulos*

*Semantic Applications Research Manager*

*Expert System Iberia*

*ESWC 2015*

*Portoroz, Slovenia*



# Named Entity Resolution

- An Information Extraction task where:
  - We detect mentions of named entities in texts (e.g. people, organizations, locations, etc.).
  - We map these mentions to the entities they refer to **in an unambiguous way.**

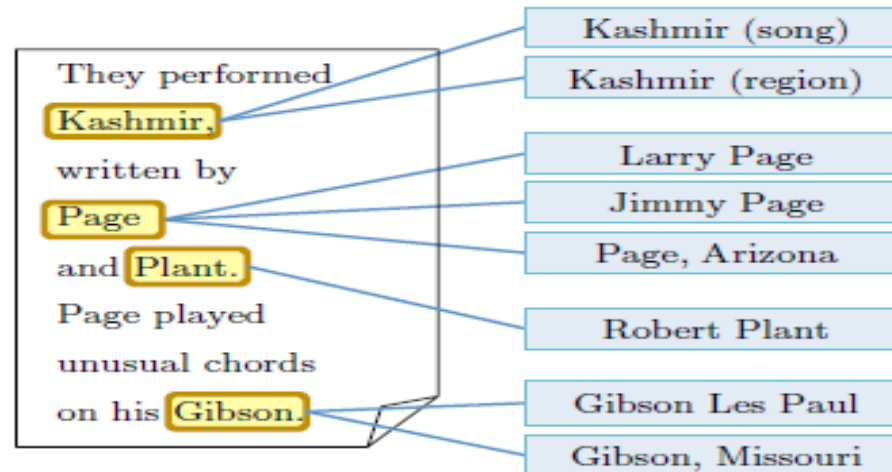


Figure from Hoffart et al: "Robust Disambiguation of Named Entities in Text"

# NER Tools and Frameworks

## Document #388760



Select one extractor

Grid of extractor options:

- AlchemyAPI
- DBpedia Spotlight
- EXTRACTIV
- ontotext
- OpenNLP
- saplo
- TextRazor
- Wikimeta
- Yahoo! Developer Network
- Zemanta
- N·E·R·D

Screenshot of the AIDA Web Interface showing the DBpedia Spotlight configuration and results. The interface includes a browser window with the URL <https://d5gate.ag5.mpi-sb.mpg.de/webaida/>. The main content area displays the "Disambiguation Method" (prior, prior+sim, prior+sim+coherence), "Entities Type Filters", and "Mention Extraction" (Standard NER, Manual). The input text is "When [Jimmy Page] played [Kashmir] at [Knebworth Festival 1979] his Les Paul was uniquely tuned." The results show the entities: Jimmy Page, Kashmir, Knebworth Festival, 1979, Les Paul, and uniquely tuned.

Screenshot of the DBpedia Spotlight interface showing the search results for "Berlin Cathedral". The interface includes a search bar, a "Confidence" slider, a "Contextual score" slider, and a "Prominence (support)" slider. The results show the text: "Berlin Cathedral (German: Berliner Dom) is the colloquial name for the Evangelical (i.e. Protestant) Oberpfarr- und Domkirche (English: analogously: Supreme Parish and Collegiate Church; literally: Supreme Parish and Cathedral Church) in Berlin, Germany. It is the parish church of the Evangelical congregation Gemeinde der Oberpfarr- und Domkirche zu Berlin, a member of the umbrella organisation Evangelical Church of Berlin - Brandenburg - Silesian - Upper Lusatia. Its present building is located on Museum Island in the Mitte borough." The interface also includes a "BACK TO TEXT" button.

You should know:

- This interface has been tested with Firefox 5.0.2 and Chromium 12.0.
  - We have a [code sandbox](#) that you should try out!
  - DBpedia Spotlight is currently only available for English. However, it could be adapted for any language for which there is a DBpedia. [Contact us](#) if you'd like to help!
- This demonstration uses the [DBpedia Spotlight Query Plugin v2.3](#) and the Web Service for [DBpedia Spotlight v2.3](#). For the latest versions, please visit: <http://www.spotlight.dbpedia.org>

# NER Tools and Frameworks

## Document #388760



Select one extract

AlchemyAPI™

TextRazor.

Different features (background knowledge, algorithms, customization, etc.)



Disambiguation Method:  
prior prior+sim prior+sim+coherence

Entities Type Filters:  
Enter the types her

Mention Extraction:  
Standard NER Manual  
You can manually tag the mentions by putting them between [ ] and [ ]. HTML Tables are automatically disambiguated in the manual mode.

When [Jimmy Page] played [Kashmir] at [Knebworth Festival 1979] his [Les Paul] was uniquely tuned.

Disambiguate

Examples YAGOTypes

DBpedia Spotlight

Confidence: 0

Contextual score: 0

Prominence (support): 0

No common words

Document-centric

Show n-best candidates

SELECT TYPES... ANNOTATE

Berlin Cathedral ( German : Berliner Dom ) is the colloquial name for the Evangelical ( i.e. Protestant ) Oberpfarr- und Domkirche ( English - analogously: Supreme Parish and Collegiate Church ; literally: Supreme Parish and Cathedral Church ) in Berlin, Germany . It is the parish church of the Evangelical congregation Gemeinde der Oberpfarr- und Domkirche zu Berlin , a member of the umbrella organisation Evangelical Church of Berlin - Brandenburg - Silesian - Upper Lusatia . Its present building is located on Museum Island in the Mitte borough .

The Berlin Cathedral had never been a cathedral in the actual sense of that term since it has never been the seat of a bishop . The bishop of the Evangelical Church in Berlin - Brandenburg ( under this name 1945-2003 ) is based in St. Mary's Church, Berlin , and Kaiser Wilhelm Memorial Church . St. Hedwig's Cathedral serves as seat of Berlin 's Roman Catholic metropolitan bishop .

BACK TO TEXT

You should know:

- This interface has been tested with Firefox 5.0.2 and Chromium 12.0.
- We have a [FAQ](#) that you should try out!
- DBpedia Spotlight is currently only available for English.
- However, it could be adapted for any language for which there is a DBpedia. [Contact us](#) if you'd like to help!

This demonstration uses the [DBpedia Spotlight Query Plugin v0.2](#) and the [Web Service for DBpedia Spotlight v0.2](#). For the latest versions, please visit: [http://spotlight.dbpedia.org](#)

# NER Tools and Frameworks

## Document #388760



Select one extract

AlchemyAPI™

TextRazor.

Effectiveness empirically measured in various experiments with several datasets

Disambiguation Method: prior prior+sim prior+sim+coherence

Entities Type Filters: Enter the types her

Mention Extraction: Stanford NER Manual

You can manually tag the mentions by putting them between [ ] and [ ]. HTML Tables are automatically disambiguated in the manual mode.

When [[Page]] played Kashmir at Knebworth Festival 1979 Knebworth, his [Les Paul] Les Paul was uniquely tuned.

Path: Disambiguate

Examples YAGOTypes

DBpedia Spotlight

Confidence: 0 Contextual score: 0 Prominence (support): 0

No 'common words' Document-centric Show n-best candidates SELECT TYPES... ANNOTATE

Berlin Cathedral<sup>en</sup> ( German<sup>en</sup> : Berliner Dom<sup>en</sup> ) is the colloquial name for the Evangelical<sup>en</sup> ( i.e. Protestant<sup>en</sup> ) Oberpfarr- und Domkirche ( English<sup>en</sup> : analogously: Supreme Parish and Collegiate Church ; literally: Supreme Parish and Cathedral Church<sup>en</sup> ) in Berlin, Germany . It is the parish church<sup>en</sup> of the Evangelical<sup>en</sup> congregation Gemeinde der Oberpfarr- und Domkirche zu - Berlin<sup>en</sup> , a member of the umbrella organisation Evangelical Church<sup>en</sup> of Berlin<sup>en</sup> - Brandenburg<sup>en</sup> - Silesian<sup>en</sup> - Upper Lusatia<sup>en</sup> . Its present building is located on Museum Island<sup>en</sup> in the Mitte<sup>en</sup> borough<sup>en</sup> .

The Berlin Cathedral<sup>en</sup> had never been a cathedral<sup>en</sup> in the actual sense of that term<sup>en</sup> since it has never been the seat of a bishop<sup>en</sup> . The bishop<sup>en</sup> of the Evangelical Church<sup>en</sup> in Berlin<sup>en</sup> - Brandenburg<sup>en</sup> ( under this name 1945-2003 ) is based in St. Mary's Church, Berlin<sup>en</sup> , and Kaiser Wilhelm Memorial Church<sup>en</sup> . St. Hedwig's Cathedral<sup>en</sup> serves as seat of Berlin<sup>en</sup>'s Roman Catholic<sup>en</sup> metropolitan bishop<sup>en</sup> .

BACK TO TEXT

You should know:

- This interface has been tested with Firefox 6.0.2 and Chromium 12.0.
- We have a nice [FAQ](#) that you should try out!
- DBpedia Spotlight is currently only available for English.
- However, it could be adapted for any language for which there is a DBpedia. [Contact us](#) if you'd like to help!

This demonstration uses the [DBpedia Spotlight Query Plugin v0.2](#) and the Web Service for [DBpedia Spotlight v0.2](#). For the latest versions, please visit: [http://spotlight.dbpedia.org](#)

# NER Systems Evaluations (F1 Scores)

- AIDA
  - 83% on the AIDA-YAGO2 dataset
  - 62% on Reuters-21578
- DBPedia Spotlight
  - 81% on a set of 155,000 wikilink samples
  - 56% on a set of 35 paragraphs from New York
  - 34% on the AIDA/CO-NLL-TestB dataset
- AGDISTIS
  - 76% on the AQUAINT dataset
  - 60% on the AIDA/CO-NLLTestB dataset
  - 31% on the IITB dataset

## What the evaluations show

*A NER system's satisfactory performance in a given scenario does not constitute a trustworthy predictor of its performance in a different scenario.*





# Our goal is to avoid the situation below

**Our system has normally 85% precision**

**Yes, but for this client, it only achieves 25%!**

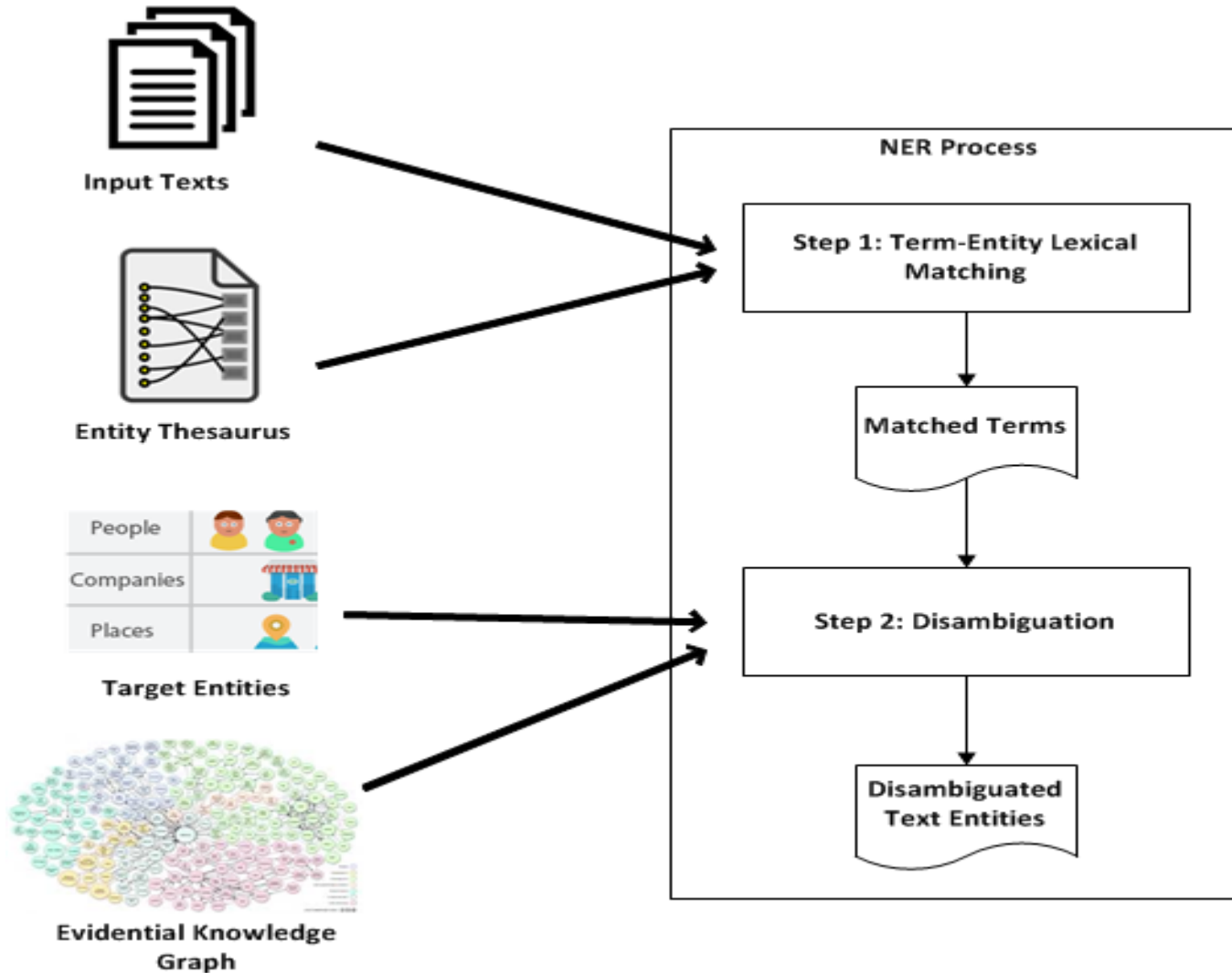




## What we have done for that

1. We analyzed the typical way a NER system works and identified (some) potential causes of low NER effectiveness.
2. We defined a set of metrics for determining in more depth the causes for low NER effectiveness in a given scenario.
3. We mapped the potential values of these metrics to actions that may increase NER effectiveness.

# How a NER system typically works



# When things may go wrong

- **Low Precision:** The text does not really contain the system-assigned entities.
- **Potential Reasons:**
  - High ambiguity in the domain and/or texts.
  - Evidence applied not enough/appropriate for the texts we have.
- **Low Recall:** System fails to detect entities in the text that are actually there.
- **Potential Reasons:**
  - The thesaurus is incomplete
  - The system requires a certain minimum amount of evidence per entity but cannot find it, either in the background knowledge or in the texts.

# How do we know what happens in our scenario

- We calculate two set of metrics:
  - **Ambiguity Metrics:** Measure the level of ambiguity in our domain knowledge and texts.
  - **Evidence Adequacy Metrics:** Measure how appropriate the domain knowledge that we apply as evidence is.
- To do that we first perform:
  - **Manual annotation** to a representative sample of the input texts with target and non-target entities from the knowledge graph.
  - **Automatic annotation** to the same texts without any disambiguation (i.e. term matching only!)

# Ambiguity Types

- **Lexical Ambiguity** of entity names, i.e. ambiguity between the target entities and common non-entity terms
  - E.g., “Page” – the piece of paper or a person?
- **Target Entity Ambiguity**, i.e., ambiguity between the target entities
  - E.g, “Tripoli” – the one in Greece or the one in Libya?
- **Knowledge Graph Ambiguity**, i.e., ambiguity between the target entities and other entities in the ontology.
  - E.g., “Barcelona” - the team or the city?
- **Global Ambiguity**, i.e. ambiguity between the target entities and entities from other domains, not covered by our thesaurus and knowledge graph.
  - E.g., “orange” - the company or the fruit?

# Ambiguity Metrics

- **Lexical Ambiguity:** The percentage of text terms which:
  - Are common lexical terms rather than entities in the text.
  - Have been wrongly mapped by the system to one or more target entities.
- **Target Entity Ambiguity:** The percentage of text terms which:
  - Correspond to a target entity.
  - Have been mapped by the system to this target entity but also to other target entities.
- **Global Ambiguity:** The percentage of text terms which
  - Are not common lexical terms but actual entities in the texts.
  - Do not correspond to any knowledge graph entity.
  - Have been wrongly mapped by the system to one or more target entities.

# Ambiguity Metrics

- **Knowledge Graph Ambiguity:** Two metrics:
  - **KGA1:** The percentage of terms which:
    - Correspond to a target entity
    - Have been mapped by the system to this entity but also to other non-target entities.
  - **KGA2:** The percentage of terms which:
    - Correspond to a non-target entity
    - Have been mapped by the system to this entity but also to other target entities.
- KGA1 shows how noisy our Knowledge Graph is with respect to the texts!
- KGA2 shows how noisy our texts are with respect to the Knowledge Graph!



# Evidence Adequacy Metrics

- **Knowledge Graph Richness**
  - Percentage of target entities with no related entities in the graph.
  - Average number of entities a target entity is related to (overall and per relation) in the graph.
- **Knowledge Graph Prevalence in Texts**
  - Percentage of target entities for which there is at least one evidential entity in the texts (overall and per relation).
  - Average number of evidential entities a target entity is related to in the texts (overall and per relation).

# Interpreting and acting on the metrics

Metric Values	Diagnosis	Action
<ul style="list-style-type: none"><li>High Lexical Ambiguity</li></ul>	The NER system cannot perform well enough Word Sense Disambiguation	Improve the linguistic analysis of the NER system.
<ul style="list-style-type: none"><li>High Global Ambiguity</li></ul>	Many of the input texts are not really related to the domain of the target entities	Use a domain/topic classifier to filter out the non-relevant texts and apply the NER process only to the relevant ones.
<ul style="list-style-type: none"><li>High KGA1</li><li>Low KGA2</li></ul>	The evidence knowledge graph contains several non-target entities that hamper the disambiguation process rather than helping it.	Prune the evidence knowledge graph and keep the most prevalent entities.

# Interpreting and acting on the metrics

Metric Values	Diagnosis	Action
<ul style="list-style-type: none"><li>• Low Knowledge Graph Richness</li></ul>	Knowledge Graph is not adequate as disambiguation evidence.	Enrich the knowledge graph starting from the most prevalent relations
<ul style="list-style-type: none"><li>• High Knowledge Graph Richness</li><li>• Low Text Prevalence</li></ul>	Knowledge Graph is not adequate as disambiguation evidence.	Change or expand the knowledge graph with entities that are more likely to appear in the texts.
<ul style="list-style-type: none"><li>• Low Knowledge Graph Text Prevalence</li><li>• Low Target Entity Ambiguity</li><li>• Low Knowledge Graph Ambiguity</li></ul>	The system's minimum evidence threshold is too high	Decrease the threshold.

# Framework Application Cases

## Case 1: Football

- **Target Texts:** Short textual descriptions of video scenes from football matches.
- **Target Entities:** Football players and teams.
- **NER System:** Knowledge Tagger (in-house)
- **Knowledge Graph:** DBPedia
- **(Initial) NER Effectiveness:**
  - P = 60%
  - R = 55%

## Case 2: Startups

- **Target Texts:** News articles from JSI's Newsfeed
- **Target Entities:** Startup companies
- **NER System:** Knowledge Tagger (in-house)
- **Knowledge Graph:** custom-built containing info about founders, investors, competitors etc.
- **(Initial) NER Effectiveness:**
  - P = 35%
  - R = 50%

# Ambiguity Metrics

Metric / Case	Football	Startups
Lexical Ambiguity	1%	<u>10%</u>
Target Entity Ambiguity	<u>30%</u>	4%
KGA1	<u>56%</u>	4%
KGA2	<u>4%</u>	<u>3%</u>
Global Ambiguity	<u>2%</u>	<u>40%</u>

# Evidence Adequacy Metrics - Knowledge Graph Prevalence

## Football

Relation	Prevalence
Players and their current club	85%
Players and their current co-players	<u>95%</u>
Players and their current managers	<u>75%</u>
Players and their nationality	<u>10%</u>
Players and their place of birth	<u>2%</u>

## Startups

Relation	Prevalence
Companies and their business areas	50%
Companies and their founders	<u>40%</u>
Companies and their competitors	<u>35%</u>
Companies and their CEO	<u>20%</u>
Companies and their investors	<u>15%</u>

# From metrics to actions

## Football

- **Metric Values:**
  - High KGA1
  - Low KGA2
  - Good Evidence Adequacy
- **Actions:**
  - We pruned the knowledge graph by removing non-football related entities and keeping only the 3 most prevalent relations.
- **Achieved NER Effectiveness:**
  - $P = 82\%$
  - $R = 80\%$

## Startups

- **Metric Values:**
  - Considerable Lexical Ambiguity
  - High Global Ambiguity
  - Mediocre Evidence Adequacy
- **Actions:**
  - Applied heuristic rules for company names detection.
  - Applied a classifier to filter out irrelevant news articles
  - Reduced the evidence threshold
- **Achieved NER Effectiveness:**
  - $P = 78\%$
  - $R = 62\%$



# Wrapping Up

- **Key Points:**

- Our NER diagnostics framework is informal and crude but has proved very helpful in optimizing our client's NER deployments.
- The main lesson we've learned is that it's very hard to build one NER solution for all possible scenarios; therefore NER systems must be easily and intuitively customizable.

- **Future Agenda:**

- Implement a comprehensive and intuitive visualization of the metrics.
- Define metrics for measuring the evidential adequacy of textual knowledge resources.
- Automate the interpretation of metric values by means of formal rules.

# Thank you for your attention!



*Dr. Panos Alexopoulos*

*Semantic Applications Research Manager*

**Email:** [palexopoulos@expertsystem.com](mailto:palexopoulos@expertsystem.com)

**Web:** [www.panosalexopoulos.com](http://www.panosalexopoulos.com)

**LinkedIn:** [www.linkedin.com/in/panosalexopoulos](http://www.linkedin.com/in/panosalexopoulos)

**Twitter:** @PAlexop

