

# Ontologies for NLP

# NLP for Ontologies

FOIS 2014 - LogOnto  
**Workshop on Logics and Ontologies  
for Natural Language**

# Overview

---

- NLP for Ontologies
- Ontologies for NLP
- Portuguese resources
- Research at PUCRS

# Introduction

---

We think and we talk

We put thoughts out of the head in the world

We write, store and share

A lot more things to think about (and much to read )

We think about the way we think and talk

We build machines to help us communicating

# NLP x Ontologies

- How do they converge, need/influence each other?
  - NLP for building ontologies from textual knowledge
  - Ontologies to make more semantically oriented NLP

# NLP for Ontologies

Ontology extraction/learning from texts

# Ontology learning from text

- Ontology components - NLP
  - Concepts – term extraction
  - Hierarchies – is-a relation
  - Properties – other relations
  - Instances – named entities
- Basic NLP needed for ontology learning
  - POS tagging (word classes: verbs, nouns, adjectives, etc.)
  - Parsing (word groups: noun phrases, verb phrases, etc.)
  - PLUS – statistical processing and machine learning

# POS and Parsing

Ronaldo Lemos, diretor do Creative Commons aprovou ontem ....

POS:

Ronaldo Lemos PROP

diretor N

de PRP

o DET

Creative Commons PROP

PARSING:

Noun Phrase



# NLP for Ontologies

Related research at PUCRS



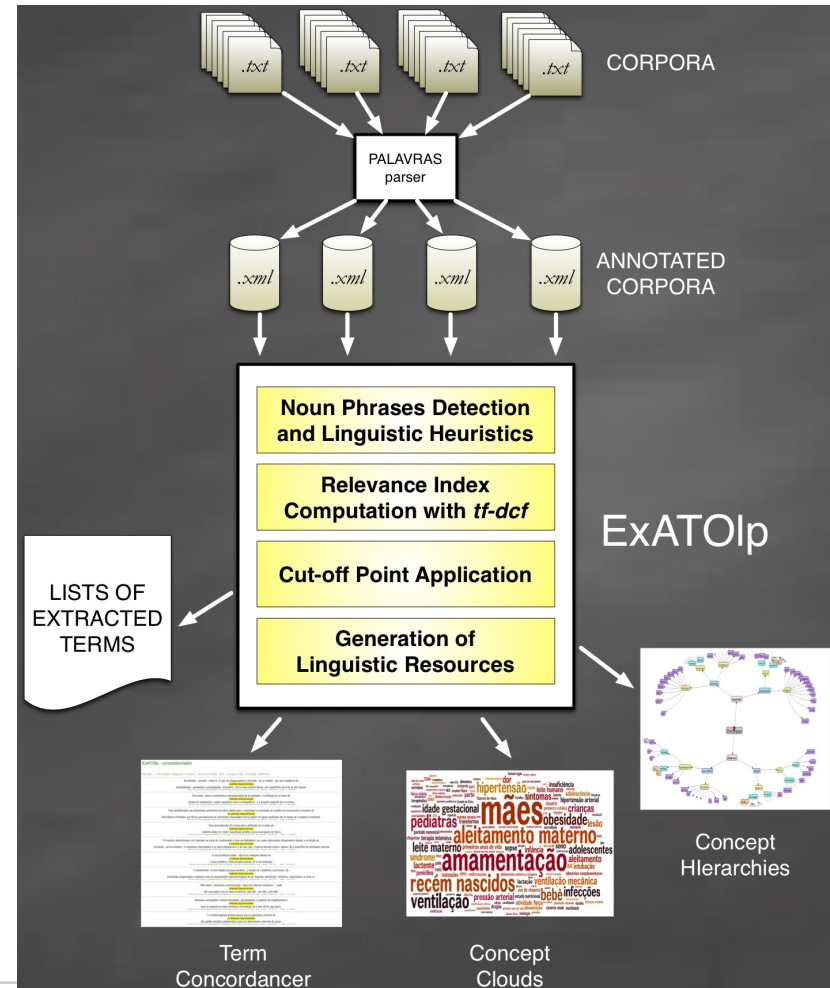
# NLP for Ontologies

- Ontology learning layer by layer
  - **Concepts (Lucelene Lopes, PosDoc)**
  - Hierarchies
  - Properties
  - Instances

# Concept Extraction

## PosDoc Lucelene Lopes

- Input: Parsed Corpora
- Term Extraction
  - (NP + filters)
- Relevance Computation
  - Concept Identification
- Concepts visualization
  - Lists
  - Concordancer
  - Clouds
  - Hierarchies



# Term Extraction Heuristics

Geology corpus:

“Nosso petróleo é uma riqueza mineral e abundante, considerando depósitos marinhos.”

SN tagged by PALAVRAS	Heuristic	SN extracted by ExATOlP
nosso petróleo	<b>pronoun removal</b>	petróleo
uma riqueza mineral	<b>article removal</b> <b>adjective removal</b> <b>adjective conjunction</b>	riqueza mineral riqueza riqueza abundante
depósitos marinhos	<b>lemma</b> <b>adjective removal</b>	depósito marinho depósito

# Relevance computation

Statistically chosen relevant terms according to *tf-dcf* index  
(using contrastive corpora)

term	lemma	head	sem	tf	tfdcf
arenitos	arenito	arenito	mat	516	516.00
granito	granito	granito	anbo mat cc drink	374	374.00
fácies	fácies	fácies	percep-f	817	272.33
Grupo Itararé	grupo itararé	Grupo Itararé	inst	236	236.00
crosta	crosta	crosta	Lcover	212	212.00
sistemas deposicionais	sistema deposicional	sistema	ac	202	202.00
feições	feição	feição	f-an percep-f	377	188.50
lagoa	lagoa	lagoa	Lwater	188	188.00
biotita	biotita	biotita	Hfam	182	182.00
planícies	planície	planície	Ltop	174	174.00
estratificação cruzada	estratificação cruzado	estratificação	event	151	151.00
folhelhos	folhelho	folhelho	anbo con	150	150.00
Bacia de Campos	bacia de campos	Bacia de Campos	?	150	150.00
feldspato	feldspato	feldspato	cm-chem	142	142.00
margem continental	margem continental	margem	am Ltop Labs cc-r	141	141.00
litologias	litologia	litologia	?	361	139.65

# Evaluation of the proposed relevance index *tf-dcf*

Pediatric corpus and reference lists - 15% of the extracted terms

	P	R	F	P	R	F
<b>PALAVRAS SN Detection</b>	12%	13%	13%	13%	7%	9%
<b>ExATOlp Linguistic Heuristics</b>	60%	68%	64%	68%	40%	50%
<b>ExATOlp Relevance Index - <i>tf-dcf</i></b>	72%	93%	81%	77%	93%	84%
	<b>bigrams</b>			<b>trigrams</b>		

# Proposed Index – *tf-dcf*

Top ranked bigrams for Pediatrics corpus

rank	according to <i>tf</i>	according to <i>tf-dcf</i>
1	aleitamento materno	aleitamento materno
2	recém nascido	recém nascido
3	faixa etária	leite materno
4	presente estudo	idade gestacional
5	leite materno	ventilação mecânica
6	idade gestacional	via aérea
7	ventilação mecânica	pressão arterial
8	via aérea	leite humano
9	pressão arterial	hipertensão arterial
10	sexo masculino	terapia intensiva

# Concordancer

## Terms occurrences with context information

*ExATOLp v. 2.0 - concordanciador*

Termo **arenito** encontrado 516 vezes nas frases abaixo

Clique em um termo para ver detalhes da sua ocorrência na frase

---

Em contraste com as rochas sedimentares de a Formação\_Águas\_Claras , constituídas predominantemente por rochas textural e composicionalmente maduras , formadas por quartzo arenitos ,

**arenitos**

sublíticos e argilitos depositados em ambiente de plataforma marinha a litorânea e fluvial entrelaçado , ocorrem arenitos com baixa maturidade composicional e textural .

função gramatical: CJT - núcleo: **arenito** - etiqueta sintática: n - etiqueta semântica: mat

Ocorrência 3 (rank: 1) retirada do arquivo

/Users/lucelene/Documents/POS\_DOC/corpora/geo/Txts71.txt.xml (frase: 135)

---

Coefficientes de a função discriminante utilizado para a identificação de

**os arenitos**

e carbonatos de o Campo\_de\_Namorado , Bacia\_de\_Campos , Brasil .

função gramatical: CJT - núcleo: **arenito** - etiqueta sintática: n - etiqueta semântica: mat

Ocorrência 4 (rank: 1) retirada do arquivo

/Users/lucelene/Documents/POS\_DOC/corpora/geo/Txts85.txt.xml (frase: 75)

---

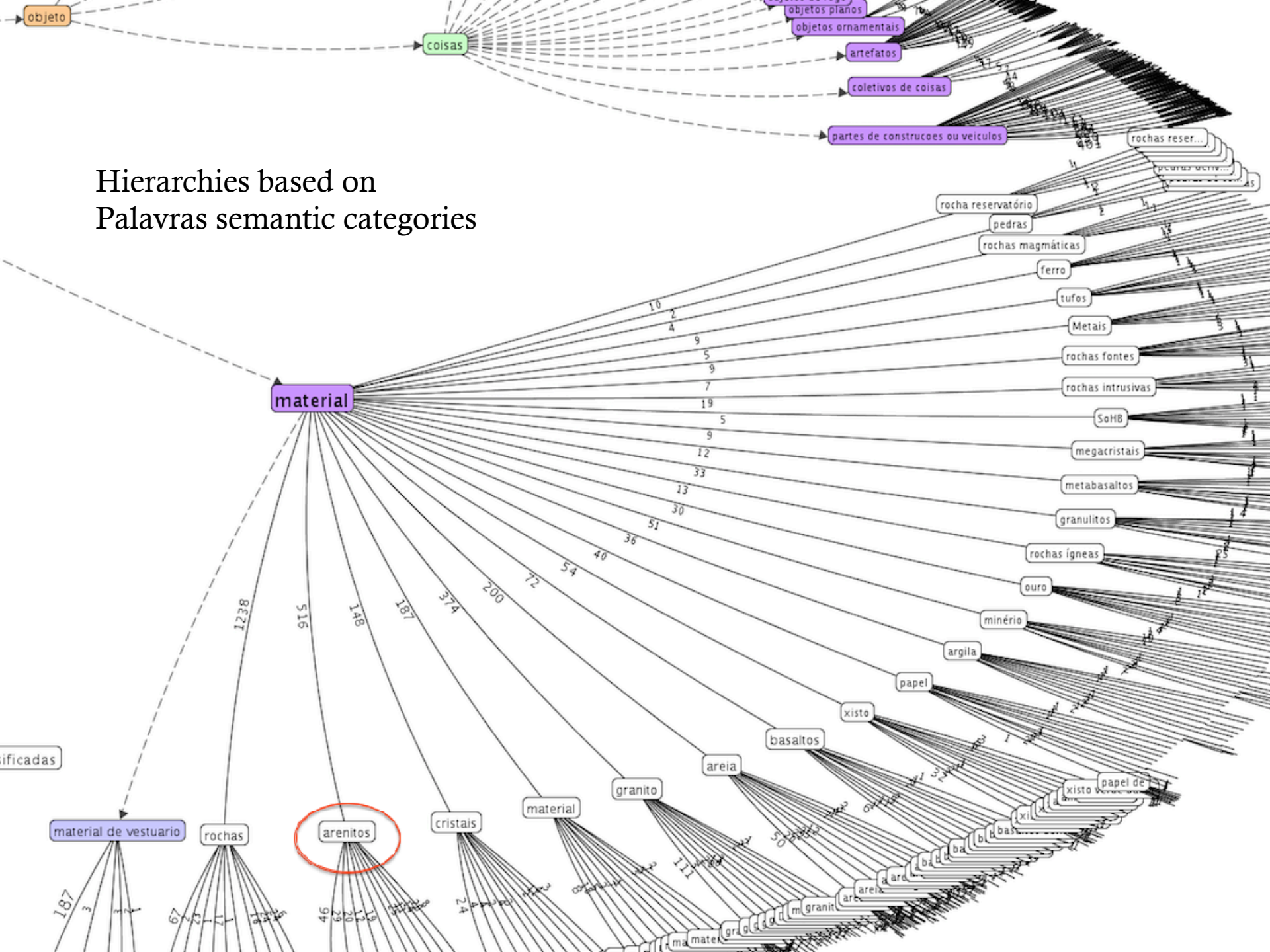




# Hierarchies

- Some hierarchical relations are also given by the tool
  - Semantic classes (parser)
  - Noun phrase structure
    - Arenito
      - Arenito maciço

# Hierarchies based on Palavras semantic categories



# References

**Lucelene Lopes.** Extração Automática de Conceitos a partir de Textos em Língua Portuguesa - Tese de Doutorado. Porto Alegre: PUCRS, 2012. v. 1. 156p .

**Lucelene Lopes, Renata Vieira.** Aplicando Pontos de Corte para Listas de Termos Extraídos. In: STIL 2013 The 9th Brazilian Symposium in Information and Human Language Technology, 2013, Fortaleza. Proceedings of STIL 2013, 2013. p. 1-6.

**Lucelene Lopes, Paulo Fernandes, Renata Vieira.** Domain term relevance through tf-dcf. In: ICAI - International Conference in Artificial Intelligence, 2012, Las Vegas, EUA. Proceedings of ICAI'12. Las Vegas, USA: Worldcomp, 2012. p. 1-7.

**Lucelene Lopes, Renata Vieira.** Improving Portuguese Term Extraction. In: International Conference on Computational Processing of the Portuguese Language - PROPOR, 2012, Coimbra. Lecture Notes in Computer Science - Proceedings of PROPOR 2012. Heidelberg: Springer, 2012. v. 7243. p. 85-92.

**Lucelene Lopes, Paulo Fernandes, Renata Vieira, Guilherme Fedrezzi.** ExATO lp -- An Automatic Tool for Term Extraction from Portuguese Language Corpora.. In: LTC'09 - 4th Language and Technology Conference, 2009, Poznan, 2009, Poznan. Proceedings of the Fourth Language and Technology Conference. Poznan: Adam Mickiewicz University, 2009. p. 427-431.

# NLP for Ontologies

- Ontology learning
  - Concepts
  - **Hierarchies (Roger Granada, PhD student)**
  - Properties
  - Instances

# Hierarchies

PhD Student Roger Granada

- Comparison of several methods of hierarchy extraction from texts
  - 2 Rule-based methods
  - 2 Statistical-based methods

# Hierarchy extraction methods

## Lexico-syntactic patterns

“...os vários ambientes que compõem os rios, **tais como** planícies de inundação, canais, macroformas e depósitos de transbordamento.”

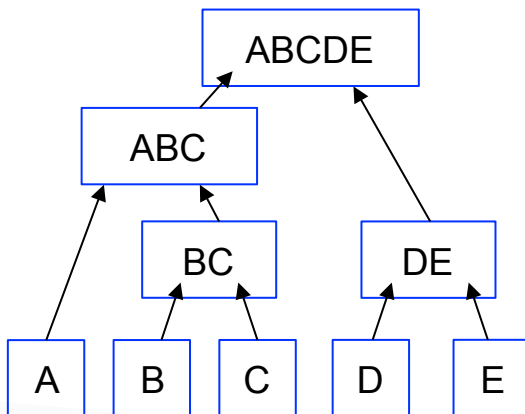
## Head modifier

Arenito

arenito eólico

arenito maciço

## Hierarchical clustering



Clusters are generated based on the **contexts** of each **word**

## Co-occurrence analysis

A term **x** **subsumes** **y** if the documents in which **y** occurs are a subset of the documents in which **x** occurs.

$P(x|y) > P(y|x)$  and  $P(x|y) >$  **threshold**

# Hierarchy extraction methods

## Lexico-syntactic patterns

Only extracts relations inside the same phrase.

High precision, low recall

## Hierarchical clustering

Uses contexts to extract relations.  
May generate other semantic relations, like synonymy, meronymy, etc.

Low precision, high recall

## Head modifier

Only extracts relations inside a noun phrase.

High precision, low recall

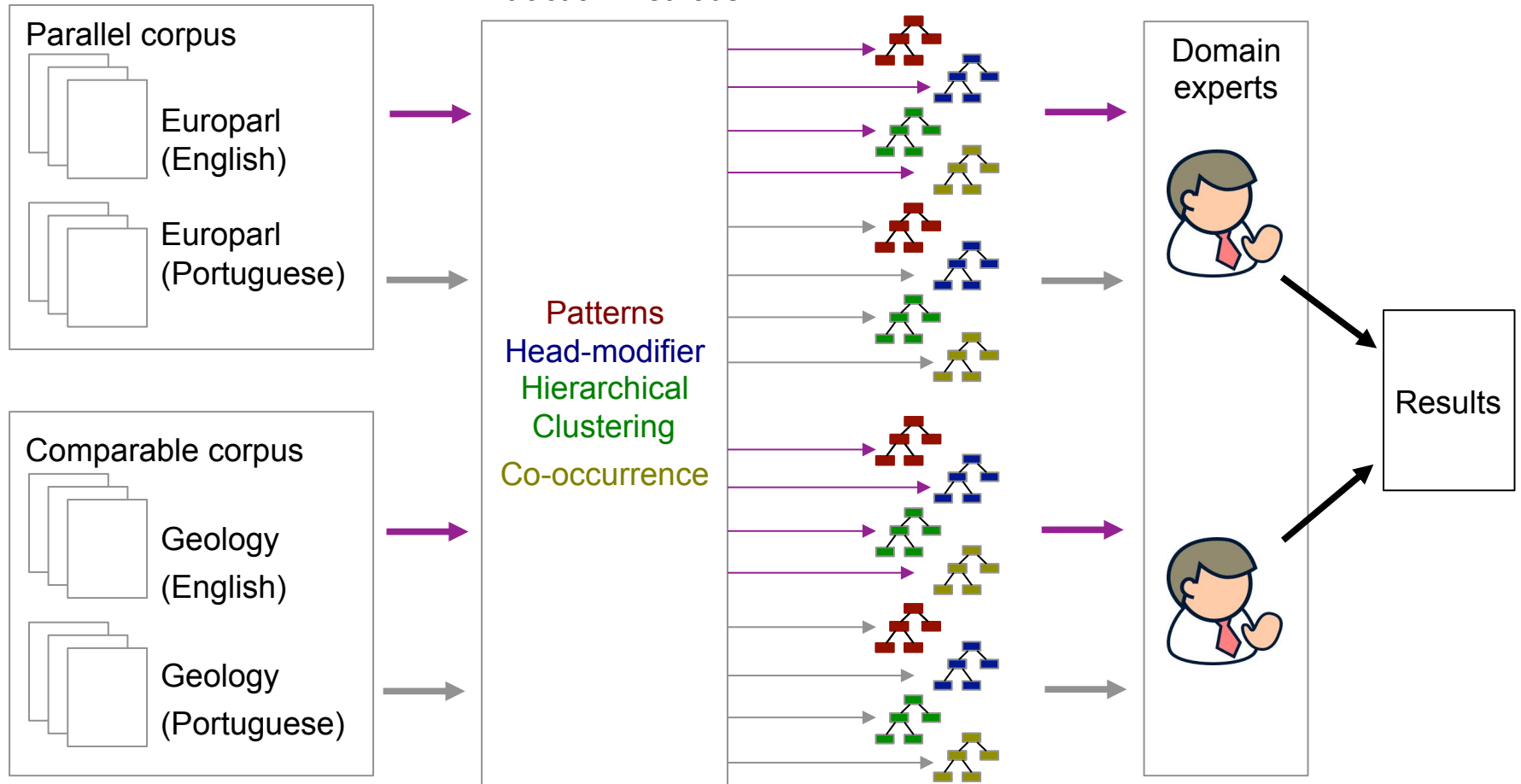
## Co-occurrence analysis

Uses the co-occurrence of terms in documents, generates relations even if the terms are not semantic related.

Low precision, high recall

# Evaluation

## Extraction Methods





# References

---

Roger Granada, Lucelene Lopes, Cassia Trojahn, Renata Vieira. A Survey of Automatic Concept Hierarchy Construction. *Artificial Intelligence Review* (submitted).

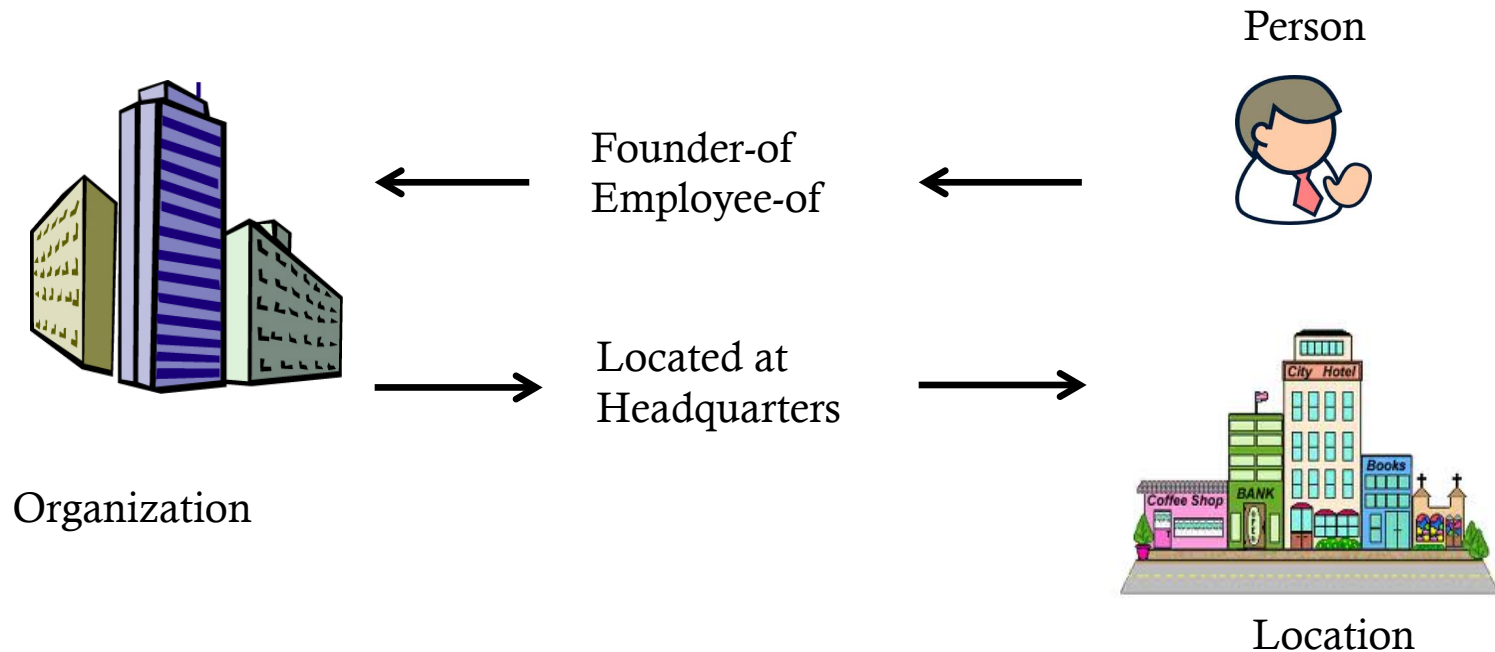
# NLP for Ontologies

- Ontology learning
  - Concepts
  - Hierarchies
  - **Properties/Relations (Sandra Collovini, PosDoc)**
  - Instances

# Relation Extraction

PosDoc Sandra Collovini

Explicit relations between entities:  
restricted by relation type; **by entity type**; open



# Relation Extraction

## ORG-PES

Relation Descriptors

**Fernando Gomes**, presidente da **Câmara Municipal do Porto**

*Fernando Gomes, **president of the** Câmara Municipal do Porto*

**A Legião da Boa Vontade**, instituição educacional, cultural e beneficente, **foi fundada pelo** jornalista **Alziro Zarur**

*Legião da Boa Vontade, an educational, cultural and beneficent institution, **was founded by** journalist **Alziro Zarur***

# Relation Extraction

## ORG-LOCAL

Relation Descriptors

Hospital de São João, **no** Porto

*Hospital de São João, at Porto*

Departamento Municipal de Limpeza Urbana **de** Porto Alegre

*Departamento Municipal de Limpeza Urbana of Porto Alegre*

# Relation Extraction

- Resources
  - Palavras parser
  - HAREM's Golden Collections for NER
  - Manual annotation of the Relations between NE

# Relation Extraction

HAREM's Golden Collections<sup>1</sup> for Entities Recognition

Ronaldo Lemos, diretor do Creative Commons

<EM ID="ric-13" CATEG="PESSOA" >Ronaldo Lemos<EM>, diretor do  
<EM ID="ric-14" CATEG="ORGANIZACAO" >Creative Commons<EM>

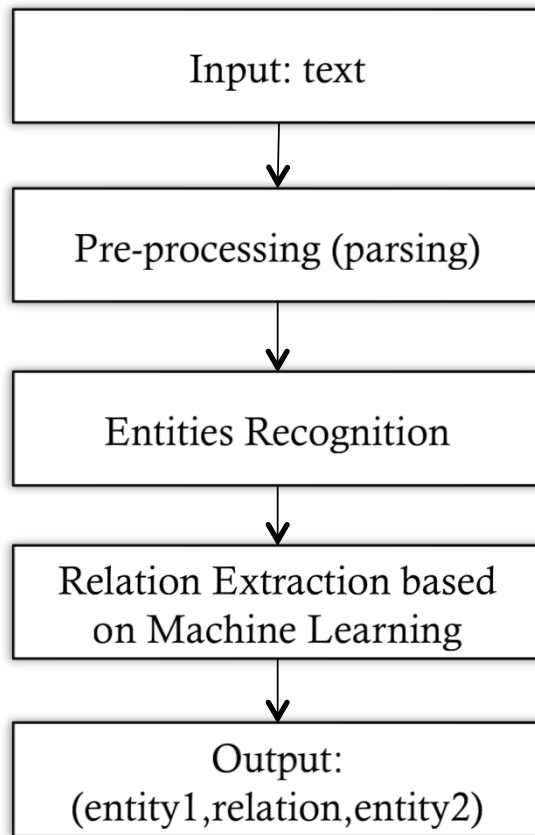
# Relation Extraction

Manual annotation of the relations between NEs

Ronaldo\_Lemos , diretor do Creative\_Commons  
[ O O REL REL O ]



# Relation Extraction



Ronaldo Lemos, diretor do Creative Commons

Ronaldo Lemos <hum> PROP @SUBJ>  
diretor <Hprof> N @N<PRED  
de PRP @N<  
o ART @>N  
Creative Commons <org> PROP @P<

Ronaldo\_Lemos <PROP, PER>  
Creative\_Commons<PROP, ORG>

Annotated corpus with Features

(Ronaldo\_Lemos, diretor-de, Creative\_Common)

# References

Sandra Collovini de Abreu, Tiago L. Bonamigo, and Renata Vieira. *A review on relation extraction with an eye on portuguese*. Journal of the Brazilian Computer Society, pages 1–19, 2013.

Sandra Collovin, Lucas Pugens, Aline A. Vanin, and Renata Vieira. *Extraction of Relation Descriptors for Portuguese using Conditional Random Fields*. In: 4th edition of the Ibero-American Conference on Artificial Intelligence - IBERAMIA 2014, Santiago, Chile, 2014.

# NLP for Ontologies

- Ontology learning
  - Concepts
  - Hierarchies
  - Properties
  - **Instances**
    - **Named entities/Daniela Amaral, PhD student**
    - **Co-reference Evandro Fonseca, PhD student**

# Named entities

## PhD Student Daniela Amaral

A opinião é do agrônomo **Miguel Guerra**, da **UFSC** (**Universidade Federal de Santa Catarina**). **Guerra** participou do debate "Biotecnologia para uma Agricultura Sustentável", realizado ontem durante a 52ª Reunião Anual da **SBPC** (**Sociedade Brasileira para o Progresso da Ciência**), sobre as biotecnologias apropriadas ao desenvolvimento do país. **Guerra** citou a micropropagação de vegetais (produção de mudas em laboratório, feita para evitar doenças e selecionar vegetais saudáveis) como exemplo de biotecnologia de baixo custo. Com ela, aumentou-se a produção de moranguinho, no sul do país, de 3,2 kg para 60 kg por hectare. Para o agrônomo...

# Named Entity Recognition

- The input/output vector

- “A opinião é do agrônomo Miguel Guerra da UFSC...”

‘O’, ‘O’, ‘O’, ‘O’, ‘O’, ‘PESS’ ‘PESS’, ‘O’, ‘LOCAL’, ...

# Named Entity Recognition

## Features

- 1) 'word': the word itself;
- 2) 'tag': POS of each word;
- 3) 'ini': the word begins contains lowercase or uppercase;
- 4) 'prevCap/nextCap': the previous/next word contains lowercase or uppercase;
- 5) ...

{'nextCap': 'min', 'word': 'opinião', 'prevCap': 'max', 'tag': 'n', 'ini': 'min'}

{'nextCap': 'min', 'word': 'é', 'prevCap': 'min', 'tag': 'v', 'ini': 'min'}

{'nextCap': 'min', 'word': 'do', 'prevCap': 'min', 'tag': 'prp', 'ini': 'min'}

....

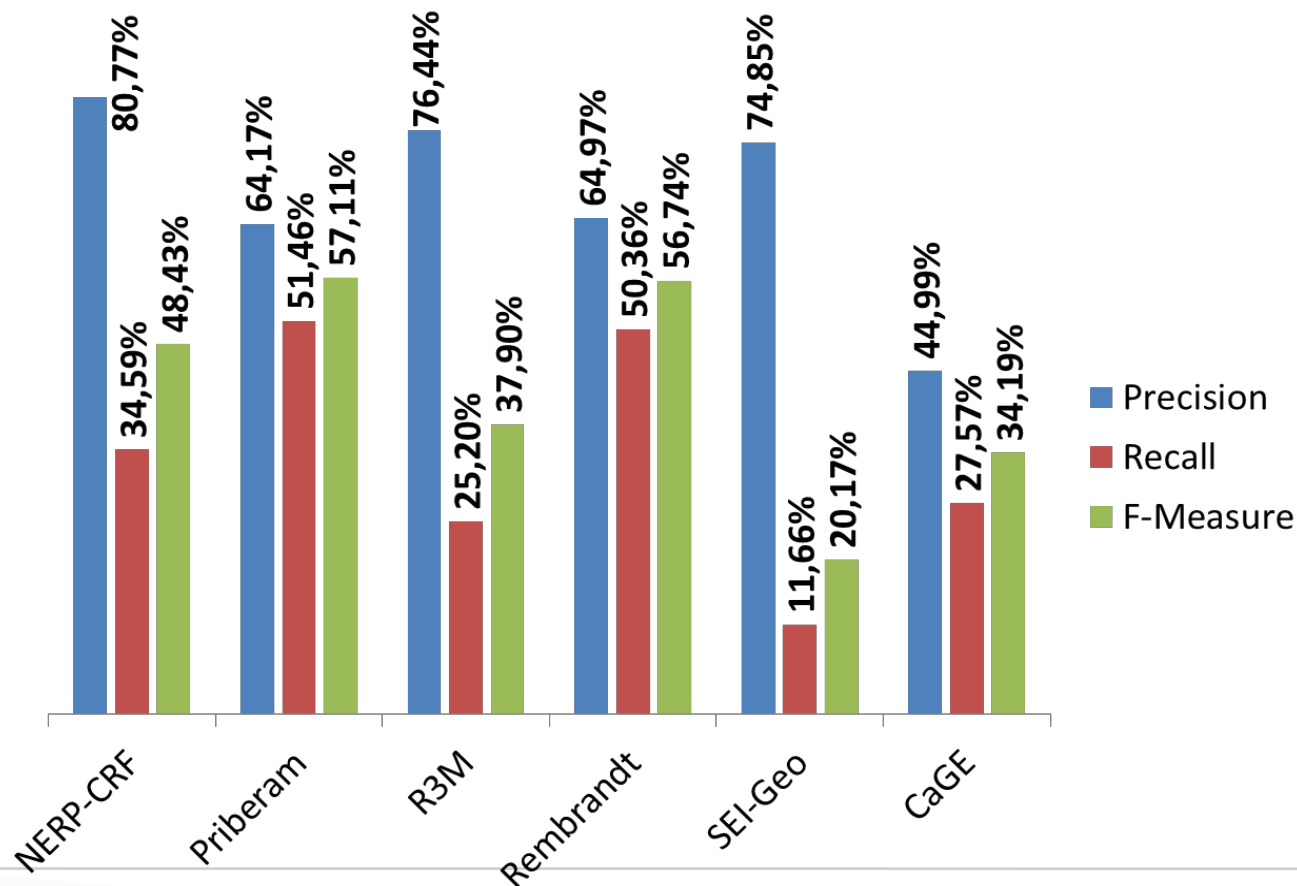
# Named Entity Recognition

- Evaluation
  - Evaluation tool: SAHARA
  - Reference: Golden Collection of Second HAREM
- Criteria:
  - Training corpus: Golden Collection of First HAREM
  - Test corpus: Golden Collection of Second HAREM
  - Ten categories: Person, Place, Organization, Value, Abstraction, Time, Work, Event, Thing, Other.



# Named Entity Recognition

- Results of CRF-NERP compared with systems



# Co-reference resolution

## PhD Student Evandro Fonseca

A opinião é do agrônomo Miguel Guerra, da UFSC (Universidade Federal de Santa Catarina). Guerra participou do debate "Biotecnologia para uma Agricultura Sustentável", realizado ontem durante a 52ª Reunião Anual da SBPC (Sociedade Brasileira para o Progresso da Ciência), sobre as biotecnologias apropriadas ao desenvolvimento do país. Guerra citou a micropropagação de vegetais (produção de mudas em laboratório, feita para evitar doenças e selecionar vegetais saudáveis) como exemplo de biotecnologia de baixo custo. Com ela, aumentou-se a produção de moranguinho, no sul do país, de 3,2 kg para 60 kg por hectare. Para o agrônomo...

# Co-reference resolution

A opinião é **do agrônomo Miguel Guerra**, da UFSC (Universidade Federal de Santa Catarina).

Parser output (CoGrOO):

[NP: A opinião ]

[VP: é ]

[PP: de ]

[NP: o agrônomo ]

[NP: Miguel\_Guerra ]

[PP: de ]

[NP: a UFSC ]

[NP:Universidade\_Federal\_de\_Santa\_Catarina ]

# Co-reference resolution

Guerra participou do debate "Biotecnologia para uma Agricultura Sustentável", realizado ontem Para o agrônomo...

Parser output (CoGrOO):

[NP: Guerra ]

[VP: participou ]

[PP: de ]

[NP: o debate ]

[NP: Biotecnologia ]

[PP: para ]

[NP: uma Agricultura\_Sustentável " ]

[PP: Para ]

[NP: o agrônomo ]

# Co-reference resolution

Same entity:

[NP: Guerra ]

[NP: o agrônomo ]

[NP: Miguel\_Guerra ]

[NP: o agrônomo ]

# Co-reference resolution

- Portuguese corpora and tool:
  - Harem corpus
  - Summ-it corpus
  - CoGroo parser

# References

Amaral, D. O. F., Fonseca, E. B. , Lopes, L., Vieira, R., Comparing NERP-CRF with Publicly Available Portuguese Named Entities Recognition Tools. In: Proceedings of International Conference on Computational Processing of Portuguese, PROPOR, São Paulo, 2014.

Amaral, D. O. F., Vieira, R., NERP-CRF: uma ferramenta para o reconhecimento de entidades nomeadas por meio de Conditional Random Fields. In: Linguamática, V.6(1): 41-49, 2014.

Amaral, D. O. F., Fonseca, E. B., Lopes, L., Vieira, R., Comparative Analysis of Portuguese Named Entities Recognition Tools. In: Proceedings of IX International Conference on Language Resources and Evaluation - LREC, 1: 2554-2558, Iceland, 2014.

Amaral, D. O. F., Vieira, R., O Reconhecimento de Entidades Nomeadas por meio de Conditional Random Fields para a Língua Portuguesa. In: Proceedings of Brazilian Conference on Intelligent Systems - STIL. , 1-10, Fortaleza, 2013.

# References

---

Fonseca E. B., Resolução de Correferência em Língua Portuguesa: Pessoa, Local e Organização, Dissertação de mestrado, Pontifícia Universidade Católica do Rio Grande do Sul, 2014.

Collovini S., Carbonel T., Fuchs J., Coelho J., Rino L., Vieira R., Summ-it: Um corpus anotado com informações discursivas visando à sumarização automática. In: V Workshop em Tecnologia da Informação e da Linguagem Humana – TIL. Proceedings of XXVII Congresso da SBC, Rio de Janeiro, 2007.



# Common problems in all levels

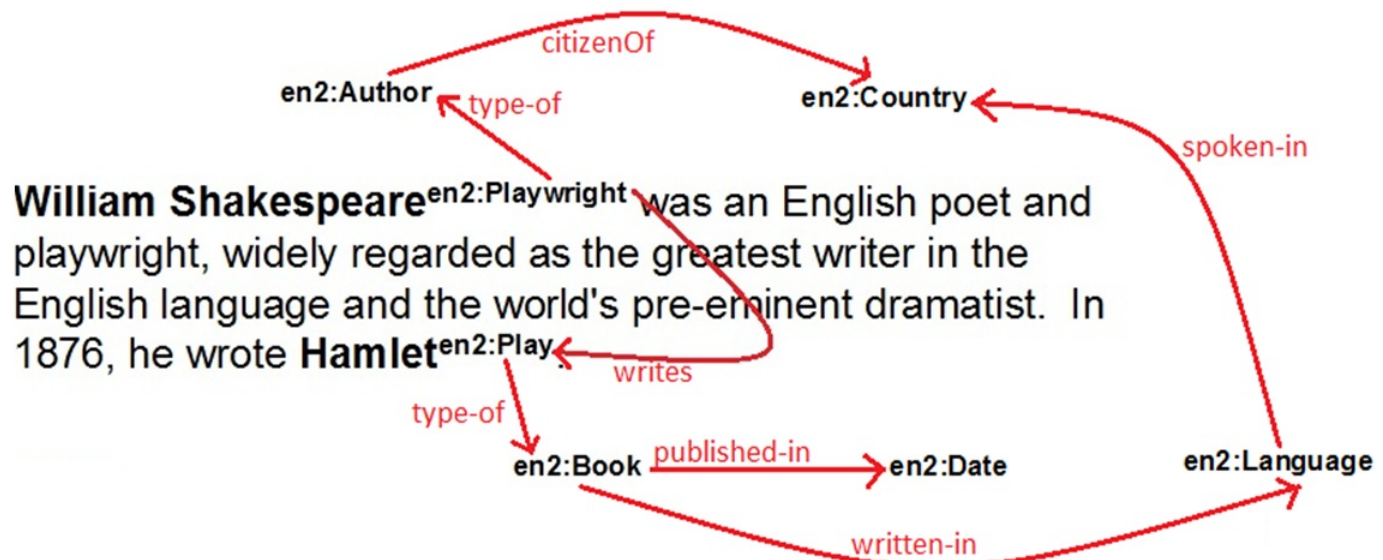
- Both ML and rule based systems require some basic linguistic pre-processing
  - POS, parsing
- For machine learning – define a relevant set of features
- Usually annotated corpus is considered as input or/and as output for evaluation
  - sometimes they are available, others not

# Ontologies for NLP

Improving NLP with richer semantics

# Ontologies for NLP

- Semantics
- A play is a type of book, has an author, has a language



# Google x entity recognition

excalibur porto **alegre**  
excalibur porto **velho**  
**motel** excalibur porto **alegre**  
excalibur **pub** porto **velho**  
Aproximadamente 52.600 resultados (0,50 segundos)

**Excalibur Forneria - Pizzaria e Forneria em Porto Alegre.**  
[www.excaliburforneria.com.br/](http://www.excaliburforneria.com.br/) ▼  
Bem-vindos à **Excalibur Forneria**. Encontre-nos. Galeria. roadtrip: imagem 1 de 4 thumb. Twitter. Tweets by @fornoExcalibur. Facebook. © 2013 **Excalibur**.  
[3 comentários no Google](#) · [Comentar](#)

R. Des. Espiridião de Lima Medeiros, 238 - Tres Figueiras, RS, 91330-020, Brasil  
+55 51 3332-0404

<b>Cardápio</b> PIZZAS SALGADAS. COM OU SEM BORDA( Catupiry ou ...	<b>Contato</b> Rua Espiridião de Lima Medeiros, 238 - Três Figueiras Porto ...
---	---

**A Forneria**  
Pode acreditar, a gente conseguiu juntar tudo isto num forno só e ...  
[Mais resultados de excaliburforneria.com.br >](#)

**Excalibur Forneria & Lounge, Três Figueiras, Porto Alegre ...**  
[www.hagah.com.br](http://www.hagah.com.br) > [Resultado da busca](#) ▼  
★★★★★ Avaliação: 4 - 9 comentários  
Encontre telefone, endereço e saiba como chegar até **Excalibur Forneria & Lounge**.  
Veja também pizzarias em **Porto Alegre, RS**.

**Excalibur Forneria, Porto Alegre - TripAdvisor**  
[www.tripadvisor.com.br](http://www.tripadvisor.com.br) > ... > [Porto Alegre](#) > [Porto Alegre Restaurantes](#) ▼  
★★★★★ Avaliação: 4,5 - 26 comentários  
**Excalibur Forneria, Porto Alegre:** Veja 26 dicas e avaliações imparciais de **Excalibur Forneria**, com classificação Nº 4,5 de 5 no TripAdvisor e classificado como ...

**Excalibur Forneria - RESTAURANTES - VEJA Porto Alegre**  
[veja.brasil.abril.com.br/porto-alegre/restaurantes/excalibur-forneria-52702](http://veja.brasil.abril.com.br/porto-alegre/restaurantes/excalibur-forneria-52702) ▼  
**Excalibur Forneria.** Endereço: Rua Desembargador Espiridião de Lima Medeiros, 238 - Três Figueiras - **Porto Alegre**. Telefone: 3332-0404 e 4100-2446.

**Excalibur Forneria - Porto Alegre, Brazil - Pizzaria | Facebook**  
<https://pt-br.facebook.com/Excaliburforno> ▼  
**Excalibur Forneria, Porto Alegre, Brazil.** 2.338 curtidas · 76 falando sobre isso · 2.967 estiveram aqui. Pizza à la carte em forno à lenha, no melhor...

**Excalibur Forneria & Lounge - Porto Alegre, RS - Foursquare**

**Excalibur Forneria & Lounge**

[Como chegar](#) [Comentar](#)  
[Seguir](#)

**Endereço:** R. Des. Espiridião de Lima Medeiros, 238 - Tres Figueiras, RS, 91330-020, Brasil  
**Telefone:** +55 51 3332-0404  
**Horário:** Aberto hoje · 11:00–14:30, 19:00–23:00

**Comentários**  
[3 comentários no Google](#)

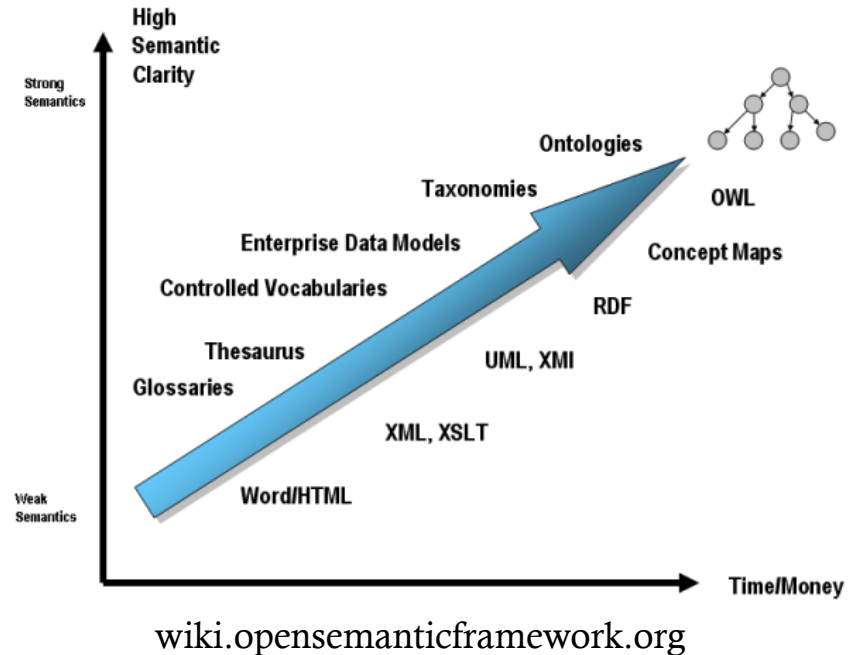
**As pessoas também pesquisaram**

<b>O Barão Pizzaria</b>	<b>Bianca Pizzaria</b>	<b>Nono Ludovico</b>
-------------------------	------------------------	----------------------

[Comentários](#)

# Ontologies spectrum

- Spectrum
  - Terms
  - Glossary
  - Thesaurus (narrower term)
  - Is-a hierarchies
  - Properties
  - Instances
  - Logical constraints
  - Axioms



**NLP: LEXICON**

# Lexicon x Ontologies

- NLP: has always be based on lexicons
  - dictionary of language tokens
  - conventional inventory of words
- Ontology formalizes concepts and their logical relations
- Computational linguists must be able to accurately map the relations between words and the concepts that they can be linked to
- Integration between lexical and semantic resources

<http://www.cambridge.org/us/academic/subjects/languages-linguistics/computational-linguistics/ontology-and-lexicon-natural-language-processing-perspective?format=AR>

# Lexicon x Ontologies

- WordNet
  - Semantic lexical database widely used in NLP

# WordNet

- Projects for linking upper level ontologies and WordNet
  - SUMO
  - DOLCE

<http://www.cambridge.org/us/academic/subjects/languages-linguistics/computational-linguistics/ontology-and-lexicon-natural-language-processing-perspective?format=AR>



# Lexicon x Ontologies

## Portuguese

- Open WN-PT
- Onto.PT
- Onto.LP
- Can we link them to SUMO and DOLCE?

# Ontologies for NLP

Related research at PUCRS

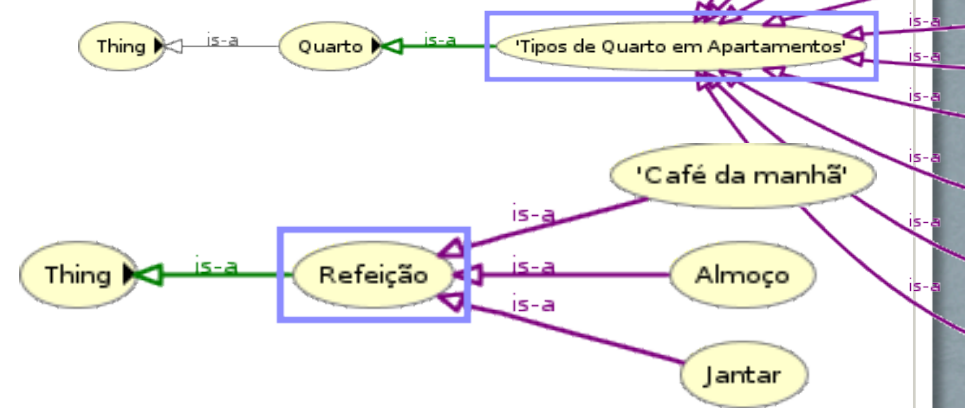
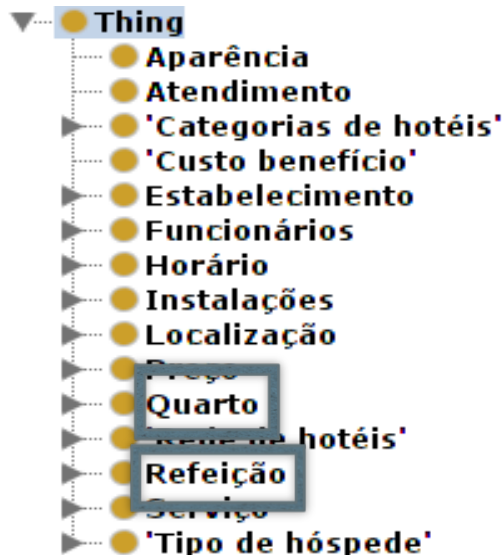
# Ontologies for Sentiment Analysis

Phd Student Larissa Freitas

- Sentiment Analysis is the field of study that analyzes people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities.
- The term aspect or feature is used to denote parts and attributes of an entity.
- An ontology formally represents knowledge (entities and aspects) as a hierarchy of concepts within a domain.

# HOntology

- HOntology<sup>1</sup> is first **multilingual** (English, Portuguese, Spanish and French) ontology for hotel domain in OWL format.



<sup>1</sup><http://ontolp.inf.pucrs.br/Recursos/downloads-Hontology.php>

# Ontology based Sentiment Analysis in Aspect Level

- Explicit and Implicit Aspects

## Explicit - Ontology Concepts

“Os quartos e banheiros são bons”

## Implicit – Ontology relations

“ As camas são novas”: Room

“Apesar da taxa de estacionamento ser salgada”: Value

### Rating summary

Location	
Sleep Quality	
Rooms	
Service	
Value	
Cleanliness	

# Ontology based Sentiment Analysis in Aspect Level

- Manual Annotation
  - Polarity of explicit aspects
  - Implicit aspects
  - Polarity of implicit aspect

# FERRAMENTA DE ANOTAÇÃO MANUAL AUTOMATIZADA

Estava de mudança para POA e reservei pela internet pois gostei das fotos do site, foi muito gratificante me hospedar neste hotel, apesar da taxa de estacionamento ser salgada devido a localização ser no Centro Histórico. Os quartos e banheiros são bons e o café da manhã é aceitável.

## Rating summary



Escolher arquivo hotel0\_0.txt

## FEATURES EXPLÍCITAS

localização	Selecione uma Opção	SubSubFeature	0	Explicito	Remover linha
quarto	Selecione uma Opção	SubSubFeature	0	Explicito	Remover linha
estabelecimento	hotel	Selecione uma Opção	0	Explicito	Remover linha
refeição	café da manhã	Selecione uma Opção	0	Explicito	Remover linha
instalação	instalação interna	estacionamento	0	Explicito	Remover linha
instalação	instalação do quarto	internet	0	Explicito	Remover linha

## FEATURES IMPLÍCITAS

custo benefício	Sem Opção	SubSubFeature	0	Implicito	ser salgada	Remover linha
-----------------	-----------	---------------	---	-----------	-------------	---------------

Adicionar linha

Guia de Anotação

Salvar

# References

CHAVES, M. S. ; TROJAHN, C. . Towards a multi-lingual ontology for ontology-driven content mining in social web sites. In: 1st International Workshop on Cross-Cultural and Cross-Lingual Aspects of the Semantic Web, Shanguai, China. ISWC, 2010.

CHAVES, M. S. ; FREITAS, L. A. ; VIEIRA, R. . HOntology: a multilingual ontology for the accommodation sector in the tourism industry. In: 4th International Conference on Knowledge Engineering and Ontology Development, 2012, Barcelona. 4th International Conference on Knowledge Engineering and Ontology Development, p. 149-154, 2012.

FREITAS, L. A. ; VIEIRA, R. . Ontology based feature level opinion mining for portuguese reviews. In: 22nd International Conference on World Wide Web - Doctoral Consortium, 2013, Rio de Janeiro, Brasil. 22nd International Conference on World Wide Web Companion, p. 367-370, 2013.

FREITAS, L. A. ; VIEIRA, R. . Comparing Portuguese Opinion Lexicons in Feature-Based Sentiment Analysis. International Journal of Computational Linguistics and Applications, v. 4, p. 147-158, 2012.

BOCHERNITSAN, M. ; FREITAS, L. A. ; VANIN, A. A. ; VIEIRA, R. . Análise de Sentimento: Descrição de uma Ferramenta de Anotação de Textos Opinativos. In: III Student Workshop on Information and Human Language Technology, 2013, Fortaleza. 2nd Brazilian Conference on Intelligent Systems, 2013.



# Portuguese

- Tools
  - LX-Center, FreeLing, Cogroo, Palavras
  - Much better than when I started in 1998
- Semantic resources ?
  - Some (as Hontology)
  - Onto.PT, Open WN-PT, Onto.LP
  - Multilinguality for ontology x lexicon linking?
- Annotated corpora
  - Harem, Summit, CSTNews
- How to build more resources from what is available?

Language resources collaboratively constructed:

Wikipedia (TorPorEsp 2014)

MsC Student Cristofer Weber

# Brasil

---

Origem: Wikipédia, a enciclopédia livre.

🔗 *Nota: Para outras definições de Brasil ou Brazil, veja [Brasil \(desambiguação\)](#).*

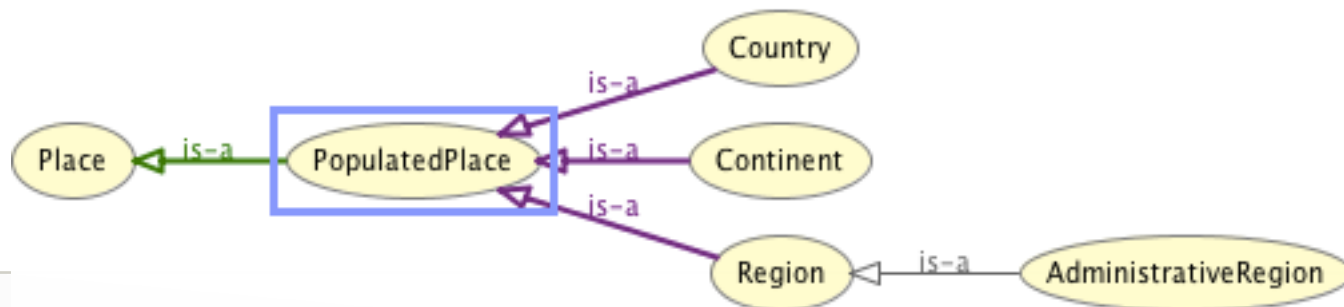
**Brasil** (pronuncia-se localmente AFI: [braˈziw]<sup>9</sup>), oficialmente **República Federativa do Brasil** (🔊 [escutar](#) (ajuda·info)),<sup>10</sup> é o maior país da América do Sul e da região da América Latina, sendo o quinto maior do mundo em área territorial (equivalente a 47% do território sul-americano)<sup>1</sup> e população (com mais de 201 milhões de habitantes).<sup>11</sup>

- Title: Brasil (pt.wikipedia.org/wiki/Brasil)
- *Hyperlinks:* país, América do Sul, América Latina, quinto maior do mundo em área territorial, população
- Named entities: América do Sul, América Latina

# Language resources collaboratively constructed: Wikipedia

<i>Wikipedia URI</i>	<i>DBpedia Instance IRI</i>
<a href="http://pt.wikipedia.org/wiki/Brasil">http://pt.wikipedia.org/wiki/Brasil</a>	<a href="http://pt.dbpedia.org/resource/Brasil">http://pt.dbpedia.org/resource/Brasil</a>
<a href="http://pt.wikipedia.org/wiki/País">http://pt.wikipedia.org/wiki/País</a>	<a href="http://pt.dbpedia.org/resource/País">http://pt.dbpedia.org/resource/País</a>
<a href="http://pt.wikipedia.org/wiki/América_do_Sul">http://pt.wikipedia.org/wiki/América_do_Sul</a>	<a href="http://pt.dbpedia.org/resource/América_do_Sul">http://pt.dbpedia.org/resource/América_do_Sul</a>
<a href="http://pt.wikipedia.org/wiki/América_Latina">http://pt.wikipedia.org/wiki/América_Latina</a>	<a href="http://pt.dbpedia.org/resource/América_Latina">http://pt.dbpedia.org/resource/América_Latina</a>

<i>DBpedia Instance IRI</i>	<i>Instance Class</i>
<a href="http://pt.dbpedia.org/resource/Brasil">http://pt.dbpedia.org/resource/Brasil</a>	Country
<a href="http://pt.dbpedia.org/resource/País">http://pt.dbpedia.org/resource/País</a>	Thing
<a href="http://pt.dbpedia.org/resource/América_do_Sul">http://pt.dbpedia.org/resource/América_do_Sul</a>	AdministrativeRegion
<a href="http://pt.dbpedia.org/resource/América_Latina">http://pt.dbpedia.org/resource/América_Latina</a>	AdministrativeRegion



# Applications

Sentiment analysis

Hotel reviews

Profile generation

University courses

Post-graduation program





# Thanks!

We think and we talk

We put thoughts out of the head in the world

We write, store and share

A lot more things to think about (and much too read)

We think about the way we think and talk

We build machines to help us communicating