

Onto.PT: Towards the Automatic Construction of a Lexical Ontology for Portuguese

PhD Thesis by: Hugo Gonalo Oliveira¹

`hroliv@dei.uc.pt`

Supervised by: Paulo Gomes

Cognitive & Media Systems Group
CISUC, Universidade de Coimbra
Portugal

May 26, 2013

¹ supported by the FCT scholarship grant SFRH/BD/44955/2008, co-funded by FSE



- 1 Introduction
- 2 Approach
- 3 Onto.PT – the resource
- 4 Conclusion

Lexical Ontologies

- Knowledge bases structured on:



Lexical Ontologies

- Knowledge bases structured on:
 - ▶ Words
 - ★ **Lexicon**



Lexical Ontologies

- Knowledge bases structured on:
 - ▶ Words
 - ★ **Lexicon**
 - ▶ Meanings → concepts
 - ★ **Ontology**



Lexical Ontologies

- Knowledge bases structured on:
 - ▶ Words
 - ★ **Lexicon**
 - ▶ Meanings → concepts
 - ★ **Ontology**
- Cover the **whole language**, not a specific domain



Lexical Ontologies

- Knowledge bases structured on:
 - ▶ Words
 - ★ **Lexicon**
 - ▶ Meanings → concepts
 - ★ **Ontology**
- Cover the **whole language**, not a specific domain
- Key in the development of NLP tools for a language
 - ▶ See Princeton WordNet for English!



Princeton WordNet

Synsets & semantic relations

- **S: (n) bank** (sloping land (especially the slope beside a body of water)) *"they pulled the canoe up on the bank"; "he sat on the bank of the river and watched the currents"*
 - [direct hyponym](#) / [full hyponym](#)
 - [direct hypernym](#) / [inherited hypernym](#) / [sister term](#)
 - **S: (n) slope, incline, side** (an elevated geological formation) *"he climbed the steep slope"; "the house was built on the side of a mountain"*
 - [derivationally related form](#)
- **S: (n) depository financial institution, bank, banking concern, banking company** (a financial institution that accepts deposits and channels the money into lending activities) *"he cashed a check at the bank"; "that bank holds the mortgage on my home"*
 - [direct hyponym](#) / [full hyponym](#)
 - [member holonym](#)
 - [direct hypernym](#) / [inherited hypernym](#) / [sister term](#)
 - **S: (n) financial institution, financial organization, financial organisation** (an institution (public or private) that collects funds (from the public or other institutions) and invests them in financial assets)
 - [derivationally related form](#)
- **S: (n) bank** (a long ridge or pile) *"a huge bank of earth"*
- **S: (n) bank** (an arrangement of similar objects in a row or in tiers) *"he operated a bank of switches"*
- **S: (n) bank** (a supply or stock held in reserve for future use (especially in emergencies))
- **S: (n) bank** (the funds held by a gambling house or the dealer in some gambling games) *"he tried to break the bank at Monte Carlo"*
- **S: (n) bank, cant, camber** (a slope in the turn of a road or track; the outside is higher than the inside in order to reduce the effects of centrifugal force)
- **S: (n) savings bank, coin bank, money box, bank** (a container (usually with a slot in the top) for keeping money at home) *"the coin bank was empty"*
- **S: (n) bank, bank building** (a building in which the business of banking transacted) *"the bank is on the corner of Nassau and Witherspoon"*



Princeton WordNet

Synsets & semantic relations

- **S: (n) bank** (sloping land (especially the slope beside a body of water)) *"they pulled the canoe up on the bank"; "he sat on the bank of the river and watched the currents"*
 - [direct hyponym](#) / [full hyponym](#)
 - [direct hypernym](#) / [inherited hypernym](#) / [sister term](#)
 - **S: (n) slope, incline, side** (an elevated geological formation) *"he climbed the steep slope"; "the house was built on the side of a mountain"*
 - [derivationally related form](#)
- **S: (n) depository financial institution, bank, banking concern, banking company** (a financial institution that accepts deposits and channels the money into lending activities) *"he cashed a check at the bank"; "that bank holds the mortgage on my home"*
 - [direct hyponym](#) / [full hyponym](#)
 - [member holonym](#)
 - [direct hypernym](#) / [inherited hypernym](#) / [sister term](#)
 - **S: (n) financial institution, financial organization, financial organisation** (an institution (public or private) that collects funds (from the public or other institutions) and invests them in financial assets)
 - [derivationally related form](#)
- **S: (n) bank** (a long ridge or pile) *"a huge bank of earth"*
- **S: (n) bank** (an arrangement of similar objects in a row or in tiers) *"he operated a bank of switches"*
- **S: (n) bank** (a supply or stock held in reserve for future use (especially in emergencies))
- **S: (n) bank** (the funds held by a gambling house or the dealer in some gambling games) *"he tried to break the bank at Monte Carlo"*
- **S: (n) bank, cant, camber** (a slope in the turn of a road or track; the outside is higher than the inside in order to reduce the effects of centrifugal force)
- **S: (n) savings bank, coin bank, money box, bank** (a container (usually with a slot in the top) for keeping money at home) *"the coin bank was empty"*
- **S: (n) bank, bank building** (a building in which the business of banking transacted) *"the bank is on the corner of Nassau and Witherspoon"*

Public, widely used



Princeton WordNet

Synsets & semantic relations

- **S: (n) bank** (sloping land (especially the slope beside a body of water)) *"they pulled the canoe up on the bank"; "he sat on the bank of the river and watched the currents"*
 - [direct hyponym](#) / [full hyponym](#)
 - [direct hypernym](#) / [inherited hypernym](#) / [sister term](#)
 - **S: (n) slope, incline, side** (an elevated geological formation) *"he climbed the steep slope"; "the house was built on the side of a mountain"*
 - [derivationally related form](#)
- **S: (n) depository financial institution, bank, banking concern, banking company** (a financial institution that accepts deposits and channels the money into lending activities) *"he cashed a check at the bank"; "that bank holds the mortgage on my home"*
 - [direct hyponym](#) / [full hyponym](#)
 - [member holonym](#)
 - [direct hypernym](#) / [inherited hypernym](#) / [sister term](#)
 - **S: (n) financial institution, financial organization, financial organisation** (an institution (public or private) that collects funds (from the public or other institutions) and invests them in financial assets)
 - [derivationally related form](#)
- **S: (n) bank** (a long ridge or pile) *"a huge bank of earth"*
- **S: (n) bank** (an arrangement of similar objects in a row or in tiers) *"he operated a bank of switches"*
- **S: (n) bank** (a supply or stock held in reserve for future use (especially in emergencies))
- **S: (n) bank** (the funds held by a gambling house or the dealer in some gambling games) *"he tried to break the bank at Monte Carlo"*
- **S: (n) bank, cant, camber** (a slope in the turn of a road or track; the outside is higher than the inside in order to reduce the effects of centrifugal force)
- **S: (n) savings bank, coin bank, money box, bank** (a container (usually with a slot in the top) for keeping money at home) *"the coin bank was empty"*
- **S: (n) bank, bank building** (a building in which the business of banking transacted) *"the bank is on the corner of Nassau and Witherspoon"*

Public, widely used

- Writing aids



Princeton WordNet

Synsets & semantic relations

- **S: (n) bank** (sloping land (especially the slope beside a body of water)) *"they pulled the canoe up on the bank"; "he sat on the bank of the river and watched the currents"*
 - [direct hyponym](#) / [full hyponym](#)
 - [direct hypernym](#) / [inherited hypernym](#) / [sister term](#)
 - **S: (n) slope, incline, side** (an elevated geological formation) *"he climbed the steep slope"; "the house was built on the side of a mountain"*
 - [derivationally related form](#)
- **S: (n) depository financial institution, bank, banking concern, banking company** (a financial institution that accepts deposits and channels the money into lending activities) *"he cashed a check at the bank"; "that bank holds the mortgage on my home"*
 - [direct hyponym](#) / [full hyponym](#)
 - [member holonym](#)
 - [direct hypernym](#) / [inherited hypernym](#) / [sister term](#)
 - **S: (n) financial institution, financial organization, financial organisation** (an institution (public or private) that collects funds (from the public or other institutions) and invests them in financial assets)
 - [derivationally related form](#)
- **S: (n) bank** (a long ridge or pile) *"a huge bank of earth"*
- **S: (n) bank** (an arrangement of similar objects in a row or in tiers) *"he operated a bank of switches"*
- **S: (n) bank** (a supply or stock held in reserve for future use (especially in emergencies))
- **S: (n) bank** (the funds held by a gambling house or the dealer in some gambling games) *"he tried to break the bank at Monte Carlo"*
- **S: (n) bank, cant, camber** (a slope in the turn of a road or track; the outside is higher than the inside in order to reduce the effects of centrifugal force)
- **S: (n) savings bank, coin bank, money box, bank** (a container (usually with a slot in the top) for keeping money at home) *"the coin bank was empty"*
- **S: (n) bank, bank building** (a building in which the business of banking transacted) *"the bank is on the corner of Nassau and Witherspoon"*

Public, widely used

- Writing aids
- Determining similarities



Princeton WordNet

Synsets & semantic relations

- **S: (n) bank** (sloping land (especially the slope beside a body of water)) *"they pulled the canoe up on the bank"; "he sat on the bank of the river and watched the currents"*
 - [direct hyponym](#) / [full hyponym](#)
 - [direct hypernym](#) / [inherited hypernym](#) / [sister term](#)
 - **S: (n) slope, incline, side** (an elevated geological formation) *"he climbed the steep slope"; "the house was built on the side of a mountain"*
 - [derivationally related form](#)
- **S: (n) depository financial institution, bank, banking concern, banking company** (a financial institution that accepts deposits and channels the money into lending activities) *"he cashed a check at the bank"; "that bank holds the mortgage on my home"*
 - [direct hyponym](#) / [full hyponym](#)
 - [member holonym](#)
 - [direct hypernym](#) / [inherited hypernym](#) / [sister term](#)
 - **S: (n) financial institution, financial organization, financial organisation** (an institution (public or private) that collects funds (from the public or other institutions) and invests them in financial assets)
 - [derivationally related form](#)
- **S: (n) bank** (a long ridge or pile) *"a huge bank of earth"*
- **S: (n) bank** (an arrangement of similar objects in a row or in tiers) *"he operated a bank of switches"*
- **S: (n) bank** (a supply or stock held in reserve for future use (especially in emergencies))
- **S: (n) bank** (the funds held by a gambling house or the dealer in some gambling games) *"he tried to break the bank at Monte Carlo"*
- **S: (n) bank, cant, camber** (a slope in the turn of a road or track; the outside is higher than the inside in order to reduce the effects of centrifugal force)
- **S: (n) savings bank, coin bank, money box, bank** (a container (usually with a slot in the top) for keeping money at home) *"the coin bank was empty"*
- **S: (n) bank, bank building** (a building in which the business of banking transacted) *"the bank is on the corner of Nassau and Witherspoon"*

Public, widely used

- Writing aids
- Determining similarities
- Word sense disambiguation
- Natural language generation
- Question answering
- Automatic summarization
- Machine translation
- ...



Princeton WordNet

Synsets & semantic relations

- **S: (n) bank** (sloping land (especially the slope beside a body of water)) *"they pulled the canoe up on the bank"; "he sat on the bank of the river and watched the currents"*
 - [direct hyponym](#) / [full hyponym](#)
 - [direct hypernym](#) / [inherited hypernym](#) / [sister term](#)
 - **S: (n) slope, incline, side** (an elevated geological formation) *"he climbed the steep slope"; "the house was built on the side of a mountain"*
 - [derivationally related form](#)
- **S: (n) depository financial institution, bank, banking concern, banking company** (a financial institution that accepts deposits and channels the money into lending activities) *"he cashed a check at the bank"; "that bank holds the mortgage on my home"*
 - [direct hyponym](#) / [full hyponym](#)
 - [member holonym](#)
 - [direct hypernym](#) / [inherited hypernym](#) / [sister term](#)
 - **S: (n) financial institution, financial organization, financial organisation** (an institution (public or private) that collects funds (from the public or other institutions) and invests them in financial assets)
 - [derivationally related form](#)
- **S: (n) bank** (a long ridge or pile) *"a huge bank of earth"*
- **S: (n) bank** (an arrangement of similar objects in a row or in tiers) *"he operated a bank of switches"*
- **S: (n) bank** (a supply or stock held in reserve for future use (especially in emergencies))
- **S: (n) bank** (the funds held by a gambling house or the dealer in some gambling games) *"he tried to break the bank at Monte Carlo"*
- **S: (n) bank, cant, camber** (a slope in the turn of a road or track; the outside is higher than the inside in order to reduce the effects of centrifugal force)
- **S: (n) savings bank, coin bank, money box, bank** (a container (usually with a slot in the top) for keeping money at home) *"the coin bank was empty"*
- **S: (n) bank, bank building** (a building in which the business of banking transacted) *"the bank is on the corner of Nassau and Witherspoon"*

Public, widely used

- Writing aids
- Determining similarities
- Word sense disambiguation
- Natural language generation
- Question answering
- Automatic summarization
- Machine translation
- ...


But...

- Created manually ☹️



Portuguese Lexical Knowledge Bases

Wordnets

-  **WordNet.PT Global**
Rede Léxico-Conceptual das variedades do Português
-  **MWN.PT**
MultiWordnet of Portuguese
- WordNet.Br
- OpenWN-PT

Portuguese Lexical Knowledge Bases

Wordnets

-  **WordNet.PT Global**
Rede Léxico-Conceptual das variedades do Português
-  **MWN.PT**
MultiWordnet of Portuguese
- WordNet.Br
- OpenWN-PT

Thesauri

-  **TeP**
-  **OpenThesaurusPT**

Portuguese Lexical Knowledge Bases


Wordnets

-  **WordNet.PT Global**
Rede Léxico-Conceptual das variedades do Português
-  **MWN.PT**
MultiWordnet of Portuguese
- WordNet.Br
- OpenWN-PT

Term-based lexical-semantic network

- **PAPEL:**
Palavras Associadas Porto
Editora-Linguateca

Thesauri

-  **TeP**
-  **OpenThesaurusPT**

Portuguese Lexical Knowledge Bases

Wordnets

-  **WordNet.PT Global**
Rede Léxico-Conceptual das variedades do Português
-  **MWN.PT**
MultiWordnet of Portuguese
- WordNet.Br
- OpenWN-PT

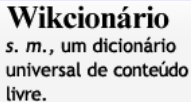
Thesauri

-  **TeP**
-  **OpenThesaurusPT**

Term-based lexical-semantic network

- **PAPEL:**
Palavras Associadas Porto
Editora-Linguateca

Electronic dictionaries

-  **Wikcionário**
s. m., um dicionário
universal de conteúdo
livre.

-  **Dicionário
Aberto**

Main Goal: Onto.PT

- **Public** lexical ontology for Portuguese



Main Goal: Onto.PT

- **Public** lexical ontology for Portuguese
- **Wordnet** model
 - ▶ Synsets: groups of synonym words → concepts
 - ▶ Connected by semantic relations



Main Goal: Onto.PT

- **Public** lexical ontology for Portuguese
- **Wordnet** model
 - ▶ Synsets: groups of synonym words → concepts
 - ▶ Connected by semantic relations
- Created **automatically**



Main Goal: Onto.PT

- **Public** lexical ontology for Portuguese
- **Wordnet** model
 - ▶ Synsets: groups of synonym words → concepts
 - ▶ Connected by semantic relations
- Created **automatically**
 - ▶ Exploitation and integration of **Portuguese** public resources



Main Goal: Onto.PT

- **Public** lexical ontology for Portuguese
- **Wordnet** model
 - ▶ Synsets: groups of synonym words → concepts
 - ▶ Connected by semantic relations
- Created **automatically**
 - ▶ Exploitation and integration of **Portuguese** public resources
- **Alternative/complement** to Portuguese LKBs



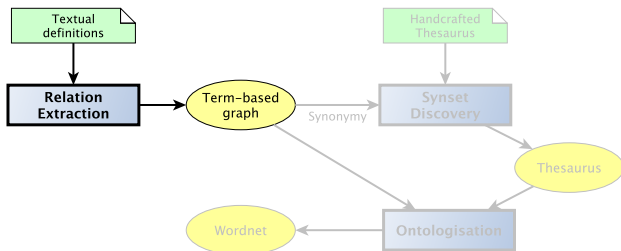
ECO: From textual definitions to a wordnet in three steps

- Combination of several **information extraction** techniques



ECO: From textual definitions to a wordnet in three steps

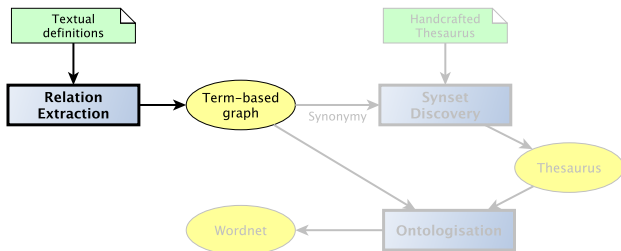
- Combination of several **information extraction** techniques



- gado s.m. conjunto de animais criados para diversos fins; rebanho
(cattle noun set of animals raised for various purposes; flock)

ECO: From textual definitions to a wordnet in three steps

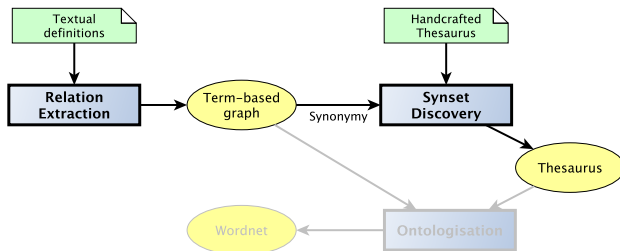
- Combination of several **information extraction** techniques



- 1 gado s.m. conjunto de animais criados para diversos fins; rebanho
 (cattle noun set of animals raised for various purposes; flock)
 - ▶ tb_triple_2 = animal MEMBRO_DE gado (animal MEMBER.OF cattle)
 - ▶ tb_triple_1 = rebanho SINONIMO_DE gado (flock SYNONYM.OF cattle)

ECO: From textual definitions to a wordnet in three steps

- Combination of several **information extraction** techniques



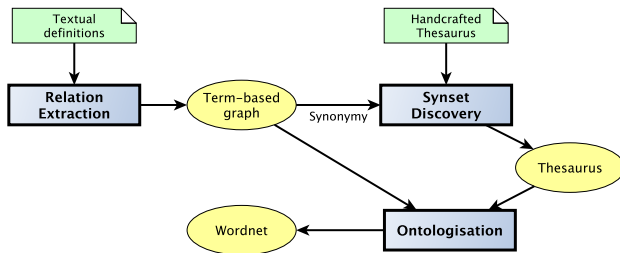
- gado s.m. conjunto de animais criados para diversos fins; rebanho
 (*cattle noun set of animals raised for various purposes; flock*)

 - ▶ $tb_triple_2 = \text{animal MEMBRO_DE gado}$ (*animal MEMBER.OF cattle*)
 - ▶ $tb_triple_1 = \text{rebanho SINONIMO_DE gado}$ (*flock SYNONYM.OF cattle*)
- $synset_1 = (\text{manada, rebanho, mancheia, boiada})$

 - ▶ $+tb_triple_1 = (\text{manada, rebanho, mancheia, boiada, gado})$

ECO: From textual definitions to a wordnet in three steps

- Combination of several **information extraction** techniques



- gado s.m. conjunto de animais criados para diversos fins; rebanho
 (*cattle noun set of animals raised for various purposes; flock*)

 - $tb_triple_2 = \text{animal MEMBRO_DE gado}$ (*animal MEMBER.OF cattle*)
 - $tb_triple_1 = \text{rebanho SINONIMO_DE gado}$ (*flock SYNONYM.OF cattle*)
- $synset_1 = (\text{manada, rebanho, mancheia, boiada})$

 - $+tb_triple_1 = (\text{manada, rebanho, mancheia, boiada, gado})$
- $synset_2 = (\text{bicho, animal, alimal, béstia, minante})$

 - $sb_triple_1 = \text{synset}_2 \text{ MEMBRO_DE } synset_1$

Step 1: relation extraction from dictionaries

- Semantic relations extracted from **three** dictionaries
 - ▶ Dicionário PRO da Língua Portuguesa (DLP), through PAPEL
 - ▶ Dicionário Aberto (DA)
 - ▶ Wiktionary.PT



Step 1: relation extraction from dictionaries

- Semantic relations extracted from **three** dictionaries
 - ▶ Dicionário PRO da Língua Portuguesa (DLP), through PAPEL
 - ▶ Dicionário Aberto (DA)
 - ▶ Wiktionary.PT
- Extraction grammars of PAPEL:
 - ▶ synonymy, hypernymy, part-of, causation, purpose-of, manner-of...

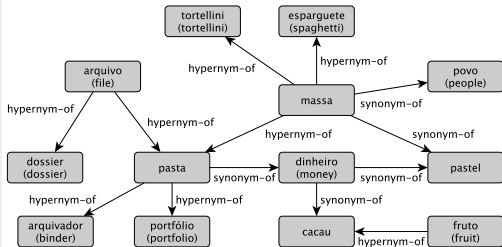


Step 1: relation extraction from dictionaries

- Semantic relations extracted from **three** dictionaries
 - ▶ Dicionário PRO da Língua Portuguesa (DLP), through PAPEL
 - ▶ Dicionário Aberto (DA)
 - ▶ Wiktionary.PT
- Extraction grammars of PAPEL:
 - ▶ synonymy, hypernymy, part-of, causation, purpose-of, manner-of...

CARTÃO

- $\approx 146,000$ lexical items
 - ▶ PAPEL 3.2 has $\approx 125,000$
- $\approx 328,000$ relational instances
 - ▶ PAPEL 3.2 has $\approx 203,000$

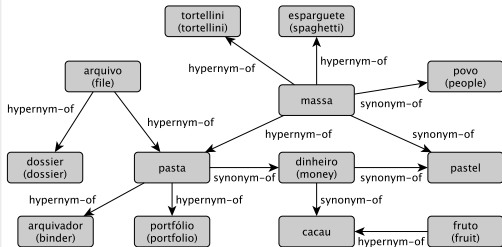


Step 1: relation extraction from dictionaries

- Semantic relations extracted from **three** dictionaries
 - ▶ Dicionário PRO da Língua Portuguesa (DLP), through PAPEL
 - ▶ Dicionário Aberto (DA)
 - ▶ Wiktionary.PT
- Extraction grammars of PAPEL:
 - ▶ synonymy, hypernymy, part-of, causation, purpose-of, manner-of...

CARTÃO

- $\approx 146,000$ lexical items
 - ▶ PAPEL 3.2 has $\approx 125,000$
- $\approx 328,000$ relational instances
 - ▶ PAPEL 3.2 has $\approx 203,000$
- Accuracy:
 - ▶ From 71% (property-of) to 99% (synonymy)
 - ▶ Hypernymy between 88-90%



Step 2a: take advantage of public handcrafted thesauri

- 1 Integrate synonymy of CARTÃO in TeP



Step 2a: take advantage of public handcrafted thesauri

- 1 Integrate synonymy of CARTÃO in TeP
 - ▶ CARTÃO synonymy as a network N
 - ▶ TeP as a thesaurus T , with synsets



Step 2a: take advantage of public handcrafted thesauri

- 1 Integrate synonymy of CARTÃO in TeP
 - ▶ CARTÃO synonymy as a network N
 - ▶ TeP as a thesaurus T , with synsets
 - ▶ Exploit network similarity for:

Synonymy pair (\vec{p})

alimentação, manutenção

(feeding, maintenance)

escravizar, servilizar

(enslave, render_servile)

Step 2a: take advantage of public handcrafted thesauri

1 Integrate synonymy of CARTÃO in TeP

- ▶ CARTÃO synonymy as a network N
- ▶ TeP as a thesaurus T , with synsets
- ▶ Exploit network similarity for:

Synonymy pair (\vec{p})	Synset (\vec{S})
<i>alimentação, manança</i> (feeding, maintenance)	→ { <i>sustento, alimento, mantimento, alimentação, manança</i> } (livelihood, sustenance, food, feeding, maintenance)
<i>escravizar, servilizar</i> (enslave, render_servile)	→ { <i>oprimir, tiranizar, escravizar, servilizar</i> } (oppress, tyrannize, enslave, render_servile)

Step 2a: take advantage of public handcrafted thesauri

1 Integrate synonymy of CARTÃO in TeP

- ▶ CARTÃO synonymy as a network N
- ▶ TeP as a thesaurus T , with synsets
- ▶ Exploit network similarity for:

Synonymy pair (\vec{p})	Synset (\vec{S})
<i>alimentação, manança</i> (feeding, maintenance)	→ { <i>sustento, alimento, mantimento, alimentação, manança</i> } (livelihood, sustenance, food, feeding, maintenance)
<i>escravizar, servilizar</i> (enslave, render_servile)	→ { <i>oprimir, tiranizar, escravizar, servilizar</i> } (oppress, tyrannize, enslave, render_servile)

- ▶ Best results ($\cos(\vec{p}, \vec{S}) \geq 0.15$):
 - ★ Precision = 74%-82%, $F_{0.5}$ = 73-83%

Step 2a: take advantage of public handcrafted thesauri

1 Integrate synonymy of CARTÃO in TeP

- ▶ CARTÃO synonymy as a network N
- ▶ TeP as a thesaurus T , with synsets
- ▶ Exploit network similarity for:

Synonymy pair (\vec{p})	Synset (\vec{S})
<i>alimentação, manança</i> (feeding, maintenance)	→ { <i>sustento, alimento, mantimento, alimentação, manança</i> } (livelihood, sustenance, food, feeding, maintenance)
<i>escravizar, servilizar</i> (enslave, render_servile)	→ { <i>oprimir, tiranizar, escravizar, servilizar</i> } (oppress, tyrannize, enslave, render_servile)

- ▶ Best results ($\cos(\vec{p}, \vec{S}) \geq 0.15$):
 - ★ Precision = 74%-82%, $F_{0.5}$ = 73-83%

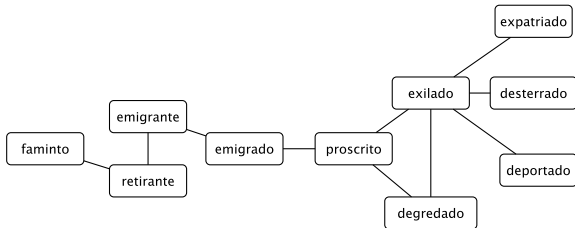
2 Discover clusters in remaining synonymy pairs

- ▶ Use remaining pairs as a synonymy network N'



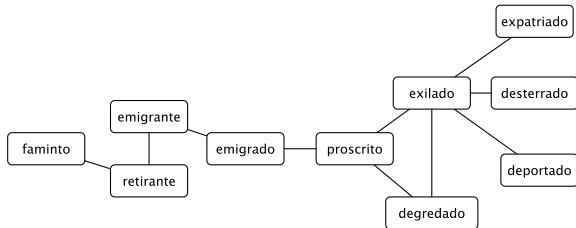
Step 2b: clustering for new synsets

- Synonymy network



Step 2b: clustering for new synsets

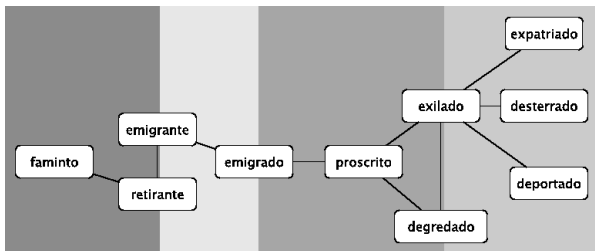
- Synonymy network



- Each node and its neighbourhood define a potential cluster (synset)

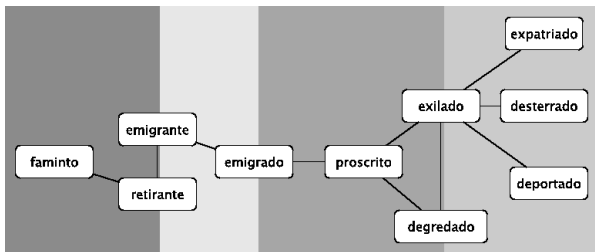
Step 2b: clustering for new synsets

- Synonymy network
- Each node and its neighbourhood define a potential cluster (synset)



Step 2b: clustering for new synsets

- Synonymy network
- Each node and its neighbourhood define a potential cluster (synset)



- Clustering accuracy:

Situation	Nouns	Verbs	Adjectives
Whole network	75%	-	-
After pair assignment	89%	83-89%	94-95%

Step 3: from term to synset relations

Synsets

- **fundação**, instituição, instituto
- **fundação**, base, alicerce
- **fundação, inauguração**, implantação, instalação, estabelecimento
- **fundação**, abertura, inauguração
- edificação, **construção**
- produção, **construção**, fabricação

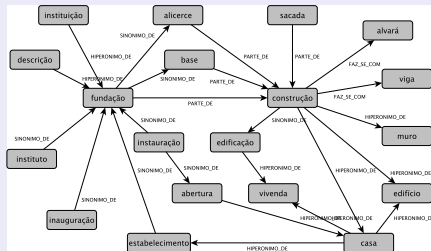
Step 3: from term to synset relations

Synsets

- **fundação**, instituição, instituto
- **fundação**, base, alicerce
- **fundação, instauração**, implantação, instalação, estabelecimento
- **fundação**, abertura, inauguração
- **edificação**, **construção**
- **produção**, **construção**, fabricação

+

Term lexical-semantic network



- Find suitable synsets for term arguments:
 - ▶ *fundação* PARTE_DE *construção*

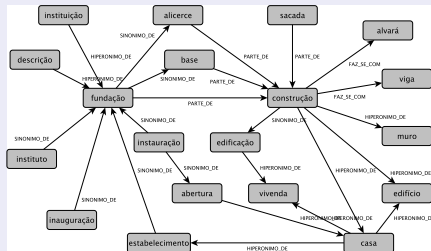
Step 3: from term to synset relations

Synsets

- **fundação**, instituição, instituto
- **fundação**, base, alicerce
- **fundação, inauguração**, implantação, instalação, estabelecimento
- **fundação**, abertura, inauguração
- **edificação**, **construção**
- **produção**, **construção**, fabricação

+

Term lexical-semantic network



- Find suitable synsets for term arguments:
 - ▶ *fundação* PARTE_DE *construção*
 - ★ { *fundação,alicerce,base* } PARTE_DE { *edificação,construção* }

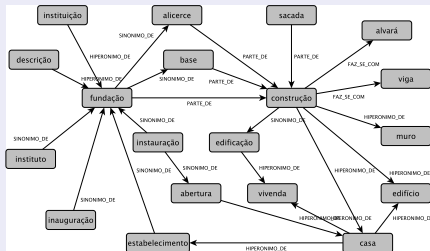
Step 3: from term to synset relations

Synsets

- **fundação**, instituição, instituto
- **fundação**, base, alicerce
- **fundação**, **instauração**, implantação, instalação, estabelecimento
- **fundação**, abertura, inauguração
- **edificação**, **construção**
- **produção**, **construção**, fabricação

+

Term lexical-semantic network



- Find suitable synsets for term arguments:
 - ▶ *fundação* PARTE_DE *construção*
 - ★ { *fundação*, *alicerce*, *base* } PARTE_DE { *edificação*, *construção* }

Relation	Sample	Algorithm	Precision	F _{0.5}
hyponymy	210	AC	60.1%	38.4%
part-of	175	RP+AC	63.3%	40.7%
purpose-of	67	RP+AC	63.4%	36.5%
antonymy	800	RP	99.4%	77.2%

- Best results:



Current version: Onto.PT v.0.4.1

- Available as RDF/OWL model



Current version: Onto.PT v.0.4.1

- Available as RDF/OWL model
- Online queries through OntoBusca

OntoBusca

[Onto.PT](#) v.0.3.1

Introduza a sua pesquisa:

Busca: *hospital*



lentamente palavra termo teste confusão casa pasta investigar

massa dinheiro ontologia borracheira cão zabumbar planta ciência

fado bebida envolvente hospital

- **S:** (n) **hospital**, [nosocômio](#), [nosocômio](#), [espiritual](#)
 - [hiperonimoDe](#)
 - **S:** (n) [manicômio](#), [rihafoles](#)
 - **S:** (n) [lazareto](#)
 - **S:** (n) [leprosaria](#), [gafaria](#)
 - **S:** (n) [leprocômio](#)
 - [hiponimoDe](#)
 - **S:** (n) [empresa](#), [casa](#), [estabelecimento](#), [firma](#)
 - **S:** (n) [obra](#), [casa](#), [edifício](#), [prédio](#), [bloco](#), [edificação](#), [herdade](#), [edificamento](#), [imóvel](#)
 - **S:** (n) [edifício](#), [prédio](#), [aranha-céu](#)
 - [referidoPeloQueÉ](#)
 - **S:** (adj) [hospitalário](#), [hospitalar](#), [nosocômico](#), [nosocomial](#), [nosocômico](#)
 - [temParte](#)
 - **S:** (n) [enfermaria](#)
 - [parteDe](#)
 - [hiperonimoDe](#)
 - **S:** (n) [ambulatório](#)
 - [servePara](#)
 - **S:** (n) [paciente](#), [enfermo](#), [doente](#)
 - [hiponimoDe](#)
- **S:** (adj) **hospital**, [benfeitor](#), [virtuoso](#), [humano](#), [santo](#), [bom](#), [favorável](#), [bendito](#), [piedoso](#), [misericordioso](#), [propenso](#), [benévolo](#), [patriarcal](#), [indulgente](#), [clemente](#), [benigno](#), [benéfico](#), [benevolente](#), [generoso](#), [bondoso](#), [obsequioso](#), [caridoso](#), [almo](#), [caritativo](#), [bondadoso](#), [benfazeite](#), [agatoide](#), [bondoso](#)
- **S:** (adj) **hospital**, [esmoler](#), [esmolador](#), [caridoso](#), [caritativo](#)

In numbers

- About 110,000 synsets

POS	Synsets		
	size > 1	size = 1	Total
Nouns	17,559	45,233	62,792
Verbs	4,422	21,638	26,060
Adjectives	8,077	11,013	19,090
Adverbs	824	1,301	2,125



In numbers

- About 110,000 synsets
- About 175,000 synset relations
 - Rich set of semantic relations

POS	Synsets		
	size > 1	size = 1	Total
Nouns	17,559	45,233	62,792
Verbs	4,422	21,638	26,060
Adjectives	8,077	11,013	19,090
Adverbs	824	1,301	2,125

Relations	Predicates	Instances
Hypernym	<i>n hiperonimoDe n</i>	82,106
Part	<i>n parteDe n</i>	3,754
	<i>n parteDeAlgoComProp adj</i>	5,011
Member	<i>n membroDe n</i>	5,957
	<i>n membroDeAlgoComProp adj</i>	115
	<i>adj propDeAlgoMembroDe n</i>	918
Contains	<i>n contidoEm n</i>	361
	<i>n contidoEmAlgoComProp adj</i>	263
Material	<i>n materialDe n</i>	847
Causation	<i>n causadorDe n</i>	1,405
	<i>n causadorDeAlgoComProp adj</i>	30
	<i>adj propDeAlgoQueCausa n</i>	638
	<i>n causadorDaAccao v</i>	79
	<i>v accaoQueCausa n</i>	7,822
Producer	<i>n produtorDe n</i>	1,649
	<i>n produtorDeAlgoComProp adj</i>	78
	<i>adj propDeAlgoProdutorDe n</i>	458
Place	<i>n localOrigemDe n</i>	1,456

Relations	Predicates	Instances
Purpose	<i>n fazSeCom n</i>	6,601
	<i>n fazSeComAlgoComProp adj</i>	84
	<i>v finalidadeDe n</i>	8,398
	<i>v finalidadeDeAlgoComProp adj</i>	333
Antonym	<i>n antonimoNDe n</i>	2,208
	<i>v antonimoVDe v</i>	1,900
	<i>adj antonimoAdjDe adj</i>	2,335
	<i>adv antonimoAdvDe adv</i>	119
Quality	<i>n temQualidade n</i>	972
	<i>n devidoAQualidade adj</i>	1,102
State	<i>n temEstado n</i>	331
	<i>n devidoAEstado adj</i>	198
Manner	<i>adv maneiraPorMeioDe n</i>	1,888
	<i>adv maneiraComProp adj</i>	1,604
Manner without	<i>adv maneiraSem n</i>	227
	<i>adv maneiraSemAccao v</i>	20
Property	<i>adj dizSeSobre n</i>	9,915
	<i>adj dizSeDoQue v</i>	24,866



Comparison with other wordnets

Resource	Lexical items	Synsets	Synset relations
Onto.PT	160,000	110,000	175,000 (×2)
WordNet.PT	11,000	12,000	40,000
WordNet.Br	44,000	18,000	N/A
MWN.PT	17,000	17,000	69,000
WordNet 3.0	155,000	118,000	285,000



Comparison with other wordnets

Resource	Lexical items	Synsets	Synset relations
Onto.PT	160,000	110,000	175,000 (×2)
WordNet.PT	11,000	12,000	40,000
WordNet.Br	44,000	18,000	N/A
MWN.PT	17,000	17,000	69,000
WordNet 3.0	155,000	118,000	285,000

- Benefits of an automatic creation approach!

Evaluation

- 600 relation instances, revised by two human judges



Evaluation

- 600 relation instances, revised by two human judges
 - ▶ Just synsets:
 - ★ 774 with size > 1
 - ★ 73.9% correct, 7.5% incorrect by both judges.
 - ★ No agreement for remaining



Evaluation

- 600 relation instances, revised by two human judges
 - ▶ Just synsets:
 - ★ 774 with size > 1
 - ★ 73.9% correct, 7.5% incorrect by both judges.
 - ★ No agreement for remaining
 - ▶ Relation instances with **two correct synsets**:
 - ★ hypernymy (247): $\approx 79\%$ accurate



Evaluation

- 600 relation instances, revised by two human judges
 - ▶ Just synsets:
 - ★ 774 with size > 1
 - ★ 73.9% correct, 7.5% incorrect by both judges.
 - ★ No agreement for remaining
 - ▶ Relation instances with **two correct synsets**:
 - ★ hypernymy (247): $\approx 79\%$ accurate
 - ★ other relations (267): 88-92% accurate



Evaluation

- 600 relation instances, revised by two human judges
 - ▶ Just synsets:
 - ★ 774 with size > 1
 - ★ 73.9% correct, 7.5% incorrect by both judges.
 - ★ No agreement for remaining
 - ▶ Relation instances with **two correct synsets**:
 - ★ hypernymy (247): $\approx 79\%$ accurate
 - ★ other relations (267): 88-92% accurate
- Coverage
 - ▶ Rough matches with 153/164 EuroWordNet base concepts



Evaluation

- 600 relation instances, revised by two human judges
 - ▶ Just synsets:
 - ★ 774 with size > 1
 - ★ 73.9% correct, 7.5% incorrect by both judges.
 - ★ No agreement for remaining
 - ▶ Relation instances with **two correct synsets**:
 - ★ hypernymy (247): $\approx 79\%$ accurate
 - ★ other relations (267): 88-92% accurate
- Coverage
 - ▶ Rough matches with 153/164 EuroWordNet base concepts
- External utilisation
 - ▶ Query expansion (Págico)
 - ▶ Cloze question answering



Main contributions

- A total of **16** publications in national and international venues

Scientific

Resources



Main contributions

- A total of **16** publications in national and international venues

Scientific

- Comparison of the structure and contents of Portuguese dictionaries (Gonçalo Oliveira et al., 2011)@**Linguamática**

Resources

- **CARTÃO**



Main contributions

- A total of **16** publications in national and international venues

Scientific

- Comparison of the structure and contents of Portuguese dictionaries (Gonçalo Oliveira et al., 2011)@**Linguamática**
- Discovery of (fuzzy) synsets from dictionary definitions (Gonçalo Oliveira and Gomes, 2011)@**IJCAI 2011**

Resources

- **CARTÃO**
- **CLIP**



Main contributions

- A total of **16** publications in national and international venues

Scientific

- Comparison of the structure and contents of Portuguese dictionaries (Gonçalo Oliveira et al., 2011)@**Linguamática**
- Discovery of (fuzzy) synsets from dictionary definitions (Gonçalo Oliveira and Gomes, 2011)@**IJCAI 2011**
- Enriching a thesaurus with new synonymy relations (Gonçalo Oliveira and Gomes, 2013)@**Expert Systems**

Resources

- **CARTÃO**
- **CLIP**
- **TRIP**



Main contributions

- A total of **16** publications in national and international venues

Scientific

- Comparison of the structure and contents of Portuguese dictionaries (Gonçalo Oliveira et al., 2011)@**Linguamática**
- Discovery of (fuzzy) synsets from dictionary definitions (Gonçalo Oliveira and Gomes, 2011)@**IJCAI 2011**
- Enriching a thesaurus with new synonymy relations (Gonçalo Oliveira and Gomes, 2013)@**Expert Systems**
- Moving from term-based to synset-based relations, without using the extraction context (Gonçalo Oliveira and Gomes, 2012)@**ECAI 2012**

Resources

- **CARTÃO**
- **CLIP**
- **TRIP**



Main contributions

- A total of **16** publications in national and international venues

Scientific

- Comparison of the structure and contents of Portuguese dictionaries (Gonçalo Oliveira et al., 2011)@**Linguamática**
- Discovery of (fuzzy) synsets from dictionary definitions (Gonçalo Oliveira and Gomes, 2011)@**IJCAI 2011**
- Enriching a thesaurus with new synonymy relations (Gonçalo Oliveira and Gomes, 2013)@**Expert Systems**
- Moving from term-based to synset-based relations, without using the extraction context (Gonçalo Oliveira and Gomes, 2012)@**ECAI 2012**
- **ECO**: a flexible approach for creating wordnets automatically (Gonçalo Oliveira et al., 2012)@**NLDB 2012**

Resources

- **CARTÃO**
- **CLIP**
- **TRIP**
- **Onto.PT**



Final remarks

- Available since April 2012 (<http://ontopt.dei.uc.pt>)
 - ▶ >1200 visits to the website (\approx 600 unique)
 - ▶ 300 visits to the download page



Final remarks

- Available since April 2012 (<http://ontopt.dei.uc.pt>)
 - ▶ >1200 visits to the website (\approx 600 unique)
 - ▶ 300 visits to the download page
- Some solid steps... but still a lot to do!



Final remarks

- Available since April 2012 (<http://ontopt.dei.uc.pt>)
 - ▶ >1200 visits to the website (\approx 600 unique)
 - ▶ 300 visits to the download page
- Some solid steps... but still a lot to do!
- Onto.PT is in constant development



Final remarks

- Available since April 2012 (<http://ontopt.dei.uc.pt>)
 - ▶ >1200 visits to the website (\approx 600 unique)
 - ▶ 300 visits to the download page
- Some solid steps... but still a lot to do!
- Onto.PT is in constant development
 - ▶ Association of dictionary definitions with synsets (paper @EPIA 2013)



Final remarks

- Available since April 2012 (<http://ontopt.dei.uc.pt>)
 - ▶ >1200 visits to the website (\approx 600 unique)
 - ▶ 300 visits to the download page
- Some solid steps... but still a lot to do!
- Onto.PT is in constant development
 - ▶ Association of dictionary definitions with synsets (paper @EPIA 2013)
 - ▶ Augmentation by exploiting other resources (e.g. OpenWN-PT, Wikipedia)



Final remarks

- Available since April 2012 (<http://ontopt.dei.uc.pt>)
 - ▶ >1200 visits to the website (\approx 600 unique)
 - ▶ 300 visits to the download page
- Some solid steps... but still a lot to do!
- Onto.PT is in constant development
 - ▶ Association of dictionary definitions with synsets (paper @EPIA 2013)
 - ▶ Augmentation by exploiting other resources (e.g. OpenWN-PT, Wikipedia)
 - ▶ Assign a confidence to relation instances



Final remarks

- Available since April 2012 (<http://ontopt.dei.uc.pt>)
 - ▶ >1200 visits to the website (\approx 600 unique)
 - ▶ 300 visits to the download page
- Some solid steps... but still a lot to do!
- Onto.PT is in constant development
 - ▶ Association of dictionary definitions with synsets (paper @EPIA 2013)
 - ▶ Augmentation by exploiting other resources (e.g. OpenWN-PT, Wikipedia)
 - ▶ Assign a confidence to relation instances
 - ▶ ...



References I

- Gonçalo Oliveira, H., Antón Pérez, L., Costa, H., and Gomes, P. (2011). Uma rede léxico-semântica de grandes dimensões para o português, extraída a partir de dicionários electrónicos. *Linguamática*, 3(2):23–38.
- Gonçalo Oliveira, H., Antón Pérez, L., and Gomes, P. (2012). Integrating lexical-semantic knowledge to build a public lexical ontology for Portuguese. In *Natural Language Processing and Information Systems, Proceedings of 17th International Conference on Applications of Natural Language to Information Systems (NLDB)*, volume 7337 of LNCS, pages 210–215, Groningen, The Netherlands. Springer.
- Gonçalo Oliveira, H. and Gomes, P. (2011). Automatic Discovery of Fuzzy Synsets from Dictionary Definitions. In *Proceedings of 22nd International Joint Conference on Artificial Intelligence, IJCAI 2011*, pages 1801–1806, Barcelona, Spain. IJCAI/AAAI.
- Gonçalo Oliveira, H. and Gomes, P. (2012). Ontologising semantic relations into a relationless thesaurus. In *Proceedings of 20th European Conference on Artificial Intelligence (ECAI 2012)*, pages 915–916, Montpellier, France. IOS Press.
- Gonçalo Oliveira, H. and Gomes, P. (2013). Towards the Automatic Enrichment of a Thesaurus with Information in Dictionaries. *Expert Systems: The Journal of Knowledge Engineering (KDBI special issue)*.



Thank you



Thank you

Questions?

