Ontologising Relational Triples into a Portuguese Thesaurus

Hugo Gonçalo Oliveira[1] Paulo Gomes

{hroliv,pgomes}@dei.uc.pt

Cognitive & Media Systems Group
CISUC, Universidade de Coimbra

October 11, 2011

[1]supported by FCT scholarship grant SFRH/BD/44955/2008

Information extraction from text

1. *He takes care of animals, such as cats and dogs.*

Information extraction from text

1. *He takes care of animals, such as cats and dogs.*

- Concepts, denoted by terms
  - *animals*
  - *cats*
  - *dogs*

Information extraction from text

1. *He takes care of animals, such as cats and dogs.*

- Concepts, denoted by terms
  - *animals*
  - *cats*
  - *dogs*
- Binary relations, between *terms*
  - $t_2 = animal, hypernym\_of, cat$
  - $t_2 = animal, hypernym\_of, dog$

Ontologising [Pantel, 2005]

- Language is **ambiguous**

Ontologising [Pantel, 2005]

- Language is **ambiguous**
  - The same word might have different meanings
  - The same concept might be denoted by different words

Ontologising [Pantel, 2005]

- Language is **ambiguous**
  - The same word might have different meanings
  - The same concept might be denoted by different words
- Kind of word-sense **disambiguation**

Ontologising [Pantel, 2005]

- Language is **ambiguous**
  - The same word might have different meanings
  - The same concept might be denoted by different words
- Kind of word-sense **disambiguation**
- Associate terms with their **meanings** in an ontology

Ontologising without representing extraction context

- Our view: two modules in an IE system
  1. **Extract** term-based triples (tb-triples)

Ontologising without representing extraction context

- Our view: two modules in an IE system
  1. **Extract** term-based triples (tb-triples)
  2. **Ontologise** tb-triples

Ontologising without representing extraction context

- Our view: two modules in an IE system
  1. **Extract** term-based triples (tb-triples)
  2. **Ontologise** tb-triples
     - ★ Attachment of each term in a triple to a concept (synset in a wordnet)

Ontologising without representing extraction context

- Our view: two modules in an IE system
  1. **Extract** term-based triples (tb-triples)
  2. **Ontologise** tb-triples
     - ★ Attachment of each term in a triple to a concept (synset in a wordnet)
     - ★ Extraction from heterogeneous sources
     - ★ Harmoniously integration in a common knowledge base

Ontologising without representing extraction context

- Our view: two modules in an IE system
  1. **Extract** term-based triples (tb-triples)
  2. **Ontologise** tb-triples
     - ★ Attachment of each term in a triple to a concept (synset in a wordnet)
     - ★ Extraction from heterogeneous sources
     - ★ Harmoniously integration in a common knowledge base
- For instance
  1. tb-triple = *animal* hypernym-of *dog*

Ontologising without representing extraction context

- Our view: two modules in an IE system
  1. **Extract** term-based triples (tb-triples)
  2. **Ontologise** tb-triples
     - ★ Attachment of each term in a triple to a concept (synset in a wordnet)
     - ★ Extraction from heterogeneous sources
     - ★ Harmoniously integration in a common knowledge base
- For instance
  1. tb-triple = *animal* hypernym-of *dog*
  2. sb-triple = (*animal*, *animate_being*, *creature*, *fauna*) hypernym-of (*dog*, *domestic_dog*, *Canis_familiaris*)

Ontologising without representing extraction context

- Our view: two modules in an IE system
  1. **Extract** term-based triples (tb-triples)
  2. **Ontologise** tb-triples
     - ★ Attachment of each term in a triple to a concept (synset in a wordnet)
     - ★ Extraction from heterogeneous sources
     - ★ Harmoniously integration in a common knowledge base
- For instance
  1. tb-triple = *animal* hypernym-of *dog*
  2. sb-triple = (*animal*, *animate_being*, *creature*, *fauna*) hypernym-of (*dog*, *domestic_dog*, *Canis_familiaris*)
- In this work: **two methods for ontologising tb-triples**

Purpose of the methods

- Input:
  - Thesaurus T, with synsets

Purpose of the methods

- Input:
  - Thesaurus T, with synsets
  - Lexical network N

Purpose of the methods

- Input:
  - Thesaurus T, with synsets
  - Lexical network N
    - ★ Nodes are terms
    - ★ Edges are relations of type R

Purpose of the methods

- Input:
  - Thesaurus T, with synsets
  - Lexical network N
    - ★ Nodes are terms
    - ★ Edges are relations of type R
- **Goal:** $w_a\ R\ w_b \in N \rightarrow A_i\ R\ B_j, (A_i, B_j) \in T$

Purpose of the methods

- Input:
  - Thesaurus T, with synsets
  - Lexical network N
    - ★ Nodes are terms
    - ★ Edges are relations of type R
- **Goal:** $w_a\ R\ w_b \in N \rightarrow A_i\ R\ B_j, (A_i, B_j) \in T$
  - No information on the extraction context!
  - Take advantage of information in N

Purpose of the methods

- Input:
  - Thesaurus T, with synsets
  - Lexical network N
    - ⋆ Nodes are terms
    - ⋆ Edges are relations of type R
- **Goal:** $w_a\ R\ w_b \in N \rightarrow A_i\ R\ B_j, (A_i, B_j) \in T$
  - No information on the extraction context!
  - Take advantage of information in N
- Output:
  - Semantic network/wordnet W

Purpose of the methods

- Input:
  - Thesaurus T, with synsets
  - Lexical network N
    - ★ Nodes are terms
    - ★ Edges are relations of type R
- **Goal:** $w_a\ R\ w_b \in N \rightarrow A_i\ R\ B_j, (A_i, B_j) \in T$
  - No information on the extraction context!
  - Take advantage of information in N
- Output:
  - Semantic network/wordnet W
    - ★ Nodes are synsets
    - ★ Edges are relations of type R

Methods

- Strategies for selecting the most suitable synsets from:
  - $A : \forall(A_i \in A) w_a \in A_i$
  - $B : \forall(B_j \in B) w_b \in B_j$

Methods

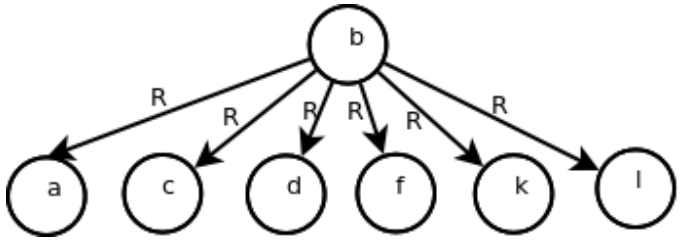
- Strategies for selecting the most suitable synsets from:
  - $A : \forall(A_i \in A) w_a \in A_i$
  - $B : \forall(B_j \in B) w_b \in B_j$
- Baselines:
  - **Random**
    - ★ attach w_a to a random synset of A
    - ★ attach w_b to a random synset of B

Methods

- Strategies for selecting the most suitable synsets from:
  - $A : \forall(A_i \in A) w_a \in A_i$
  - $B : \forall(B_j \in B) w_b \in B_j$
- Baselines:
  - **Random**
    - ★ attach w_a to a random synset of A
    - ★ attach w_b to a random synset of B
  - **Average frequency** (AF)
    - ★ compute the average frequency of the terms of each synset in Google
    - ★ attach w_a to the *most frequent* synset of A
    - ★ attach w_b to the *most frequent* synset of B

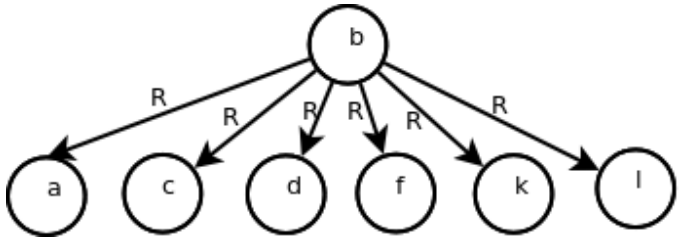
Related proportion (RP)

- For attaching w_a ($\theta = 0.5$)
  1. Fix w_b



Related proportion (RP)

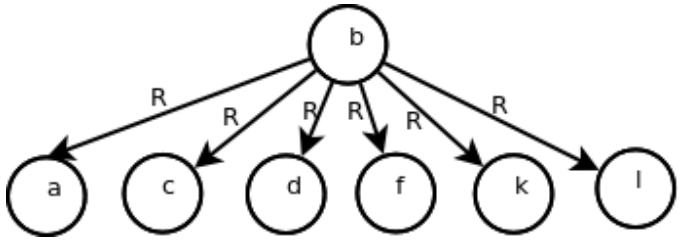
- For attaching w_a ($\theta = 0.5$)
  1. Fix w_b



  2. For each $A_i \in A$

Related proportion (RP)

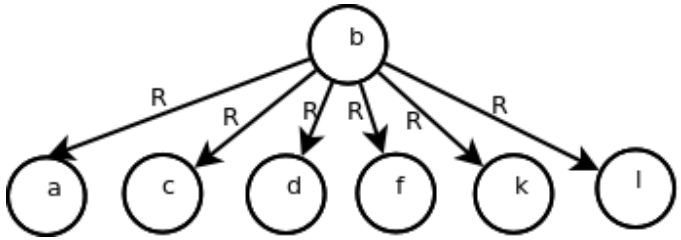
- For attaching w_a ($\theta = 0.5$)
  1. Fix w_b



  2. For each $A_i \in A$
     - ★ $A_1 = (a, c, d, e)$, $p_{a1} = \frac{3}{4}$
     - ★ $A_2 = (a, f, g)$, $p_{a2} = \frac{2}{3}$
     - ★ $A_3 = (a, h, i, j)$, $p_{a3} = \frac{1}{4}$

Related proportion (RP)

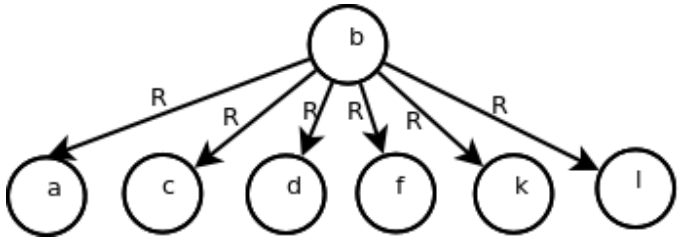
- For attaching w_a ($\theta = 0.5$)
  1. Fix w_b



  2. For each $A_i \in A$
     - ★ $A_1 = (a, c, d, e)$, $p_{a1} = \frac{3}{4}$
     - ★ $A_2 = (a, f, g)$, $p_{a2} = \frac{2}{3}$
     - ★ $A_3 = (a, h, i, j)$, $p_{a3} = \frac{1}{4}$
  3. $p_{a1} \geq \theta$, $w_a \rightarrow A_1$

Related proportion (RP)

- For attaching w_a ($\theta = 0.5$)
  1. Fix w_b



  2. For each $A_i \in A$
     - ★ $A_1 = (a, c, d, e)$, $p_{a1} = \frac{3}{4}$
     - ★ $A_2 = (a, f, g)$, $p_{a2} = \frac{2}{3}$
     - ★ $A_3 = (a, h, i, j)$, $p_{a3} = \frac{1}{4}$
  3. $p_{a1} \geq \theta$, $w_a \rightarrow A_1$
- For attaching w_b
  - Fix w_a ...

Average cosine (AC)

1. M = *term-term* matrix with the adjacencies of N

Average cosine (AC)

1. M = *term-term* matrix with the adjacencies of N
   - Synset matrix $M_{Ai} = ([M_{Ai0}], ..., [M_{Ain}]), n = |A_i|$
   - Synset matrix $M_{Bj} = ([M_{Bj0}], ..., [M_{Bjn}]), n = |B_j|$

Average cosine (AC)

1. M = *term-term* matrix with the adjacencies of N
   - Synset matrix $M_{Ai} = ([M_{Ai0}], ..., [M_{Ain}]), n = |A_i|$
   - Synset matrix $M_{Bj} = ([M_{Bj0}], ..., [M_{Bjn}]), n = |B_j|$
2. For each $A_i \in A$ and $B_j \in B$:

$$sim(A_i, B_j) = \frac{\sum_{k=1}^{|A_i|} \sum_{l=1}^{|B_j|} \cos([M_{Aik}], [M_{Bjl}])}{|A_i||B_j|}$$

Average cosine (AC)

1. M = *term-term* matrix with the adjacencies of N
   - Synset matrix $M_{Ai} = ([M_{Ai0}], ..., [M_{Ain}]), n = |A_i|$
   - Synset matrix $M_{Bj} = ([M_{Bj0}], ..., [M_{Bjn}]), n = |B_j|$
2. For each $A_i \in A$ and $B_j \in B$:

$$sim(A_i, B_j) = \frac{\sum_{k=1}^{|A_i|} \sum_{l=1}^{|B_j|} \cos([M_{Aik}], [M_{Bjl}])}{|A_i||B_j|}$$

3. Select the most similar pair

Thesaurus: TePOT

- Noun **synsets** from TeP 2.0[2] + OpenThesaurus.PT[3]
  - 18,501 nouns, organised in 8,293 synsets

[2] http://www.nilc.icmc.usp.br/tep2/
[3] http://openthesaurus.caixamagica.pt/

Thesaurus: TePOT

- Noun **synsets** from TeP 2.0[2] + OpenThesaurus.PT[3]
  - 18,501 nouns, organised in 8,293 synsets
  - Average synset size = 3.84 terms
  - On average one term is in 1.72 synsets

[2] http://www.nilc.icmc.usp.br/tep2/
[3] http://openthesaurus.caixamagica.pt/

Thesaurus: TePOT

- Noun **synsets** from TeP 2.0[2] + OpenThesaurus.PT[3]
  - 18,501 nouns, organised in 8,293 synsets
  - Average synset size = 3.84 terms
  - On average one term is in 1.72 synsets
  - Synsets ranked according to frequency in Google

$$Freq(S) = \frac{\sum_{i=1}^{|S|} Hits(w_i)}{|S|}, w_i \in S$$

[2] http://www.nilc.icmc.usp.br/tep2/
[3] http://openthesaurus.caixamagica.pt/

Tb-triples: PAPEL 2.0[4]

- **Lexical network** extracted automatically from a Portuguese dictionary [Gonçalo Oliveira et al., 2010]

[4] http://www.linguateca.pt/PAPEL

Tb-triples: PAPEL 2.0[4]

- **Lexical network** extracted automatically from a Portuguese dictionary [Gonçalo Oliveira et al., 2010]
  - ≈ 200k tb-triples of different kinds

[4] http://www.linguateca.pt/PAPEL

Tb-triples: PAPEL 2.0[4]

- **Lexical network** extracted automatically from a Portuguese dictionary [Gonçalo Oliveira et al., 2010]
  - ≈ 200k tb-triples of different kinds
- Sample for ontologisation:
  - Triples supported by CETEMPúblico [Rocha and Santos, 2000]

[4] http://www.linguateca.pt/PAPEL

Tb-triples: PAPEL 2.0[4]

- **Lexical network** extracted automatically from a Portuguese dictionary [Gonçalo Oliveira et al., 2010]
  - $\approx$ 200k tb-triples of different kinds
- Sample for ontologisation:
  - Triples supported by CETEMPúblico [Rocha and Santos, 2000]
  - Triples with arguments CETEMPúblico frequency ≥ 25

[4] http://www.linguateca.pt/PAPEL

Tb-triples: PAPEL 2.0[4]

- **Lexical network** extracted automatically from a Portuguese dictionary [Gonçalo Oliveira et al., 2010]
  - $\approx$ 200k tb-triples of different kinds
- Sample for ontologisation:
  - Triples supported by CETEMPúblico [Rocha and Santos, 2000]
  - Triples with arguments CETEMPúblico frequency ≥ 25
  - No triples with very frequent/abstract arguments (eg. *acto*, *coisa*)

[4] http://www.linguateca.pt/PAPEL

Tb-triples: PAPEL 2.0[4]

- **Lexical network** extracted automatically from a Portuguese dictionary [Gonçalo Oliveira et al., 2010]
  - $\approx$ 200k tb-triples of different kinds
- Sample for ontologisation:
  - Triples supported by CETEMPúblico [Rocha and Santos, 2000]
  - Triples with arguments CETEMPúblico frequency ≥ 25
  - No triples with very frequent/abstract arguments (eg. *acto*, *coisa*)
    - ★ 500 hypernymy tb-triples
    - ★ 199 part-of tb-triples
    - ★ 436 member-of tb-triples
    - ★ 125 purpose-of tb-triples

[4] http://www.linguateca.pt/PAPEL

Coverage

Table: Hypernym-of triples

Method	Ontologised triples	Coverage (%) Total	TS
Random	499/500	**99.80**	**69.00**
AF	496/500	99.60	68.80
RP 0.5	315/500	63.00	39.00
RP 0.2	382/500	76.40	49.00
AC	458/500	91.60	61.20

Table: Part-of triples

Method	Ontologised triples	Coverage (%) Total	TS
Random	199/199	**100.00**	**54.27**
AF	199/199	**100.00**	**54.27**
RP 0.5	87/199	17.40	6.60
RP 0.2	129/199	64.82	31.16
AC	188/199	94.47	49.75

Table: Member-of triples

Method	Ontologised triples	Coverage (%) Total	TS
Random	436/436	**100.00**	**74.77**
AF	427/436	97.94	72.71
RP 0.5	156/436	35.78	22.48
RP 0.2	266/436	61.01	54.59
AC	400/436	91.74	67.20

Table: Purpose-of triples

Method	Ontologised triples	Coverage (%) Total	TS
Random	125/125	**100.00**	**69.60**
AF	125/125	**100.00**	**69.60**
RP 0.5	44/125	35.20	20.80
RP 0.2	73/125	58.40	37.60
AC	121/125	96.80	66.40

Corpus-based validation

- Support in CETEMPúblico for synset-based triples

Corpus-based validation

- Support in CETEMPúblico for synset-based triples
- For synsets S_a and S_b in (S_a, R, S_b):

Corpus-based validation

- Support in CETEMPúblico for synset-based triples
- For synsets S_a and S_b in (S_a, R, S_b):
  - Search for occurrences of $a_i \in S_a$ and $b_j \in S_b$
  - Connected by a **discriminating pattern** of R

Corpus-based validation

- Support in CETEMPúblico for synset-based triples
- For synsets S_a and S_b in (S_a, R, S_b):
  - Search for occurrences of $a_i \in S_a$ and $b_j \in S_b$
  - Connected by a **discriminating pattern** of R
- If support is found, $found(a_i, b_j, R) = 1$, otherwise $found(a_i, b_j, R) = 0$

Corpus-based validation

- Support in CETEMPúblico for synset-based triples
- For synsets S_a and S_b in (S_a, R, S_b):
  - Search for occurrences of $a_i \in S_a$ and $b_j \in S_b$
  - Connected by a **discriminating pattern** of R
- If support is found, $found(a_i, b_j, R) = 1$, otherwise $found(a_i, b_j, R) = 0$

$$Precision = \frac{\sum_{i=1}^{|S_a|} \sum_{j=1}^{|S_b|} found(a_i, b_j, R)}{|S_a| \times |S_b|}$$

Corpus-based validation

- Support in CETEMPúblico for synset-based triples
- For synsets S_a and S_b in (S_a, R, S_b):
  - Search for occurrences of $a_i \in S_a$ and $b_j \in S_b$
  - Connected by a **discriminating pattern** of R
- If support is found, $found(a_i, b_j, R) = 1$, otherwise $found(a_i, b_j, R) = 0$

$$Precision = \frac{\sum_{i=1}^{|S_a|} \sum_{j=1}^{|S_b|} found(a_i, b_j, R)}{|S_a| \times |S_b|}$$

$$Recall = \frac{Precision \times |OntologisedTriples|}{|TriplesInSample|} \qquad F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Scores of ontologised triples

Tb-triples	Sb-triples	Score
aparelho **Hypernym-of** *televisor*	*apresto,utensílio,petrechos,instrumento,apetrechos,aparelho* **Hypernym-of** *tv,televisão,televisor,têvê*	1.0
ângulo **Part-of** *triângulo*	*ângulo,face,lado* **Part-of** *triângulo,trilateral*	1.0
técnica **Member-of** *marketing*	*arte,técnica* **Member-of** *marketing*	1.0
edição **Purpose-of** *programa*	*edição,lançamento* **Purpose-of** *programa,aplicativo*	1.0
arte **Hypernym-of** *escultura*	*arte,obra* **Hypernym-of** *escultura,vulto,imagem,estátua*	0.75
província **Part-of** *país*	*província,distrito,circunscrição,região* **Part-of** *território,pátria,região,nação*	0.55
pessoa **Member-of** *raça*	*ser,pessoa,criatura,indivíduo* **Member-of** *raça,gente*	0.5
calor **Purpose-of** *combustível*	*calor,aquecimento,animação,entusiasmo* **Purpose-of** *combustível*	0.5

Supporting sentences

- *...as máquinas fotográficas digitais, os* **televisores**, *os videogravadores e outros* **aparelhos** *da electrónica de consumo...*
- *...também não consigo reparar* **televisões** *e outros* **aparelhos**.
- *O* **ângulo** *externo do* **triângulo** *($n = 3$) se obtém dividindo 360 por 3.*
- *E não estão suficientemente precavidos contras as* **técnicas** *de* **marketing**.
- *A guerra entre* **seres** *da mesma* **raça** *tornou-se no maior cataclismo jamais conhecido da história do Universo.*
- *...é um disparate queimar* **combustível** *para produzir* **calor**.

Validation results

Table: Hypernym-of triples

Method	Precision (%)		F_1 (%)	
	Total	TS	Total	TS
Random	33.46	27.37	33.43	22.35
AF	35.76	29.88	35.69	**24.36**
RP 0.5	**43.22**	**33.10**	33.41	18.58
RP 0.2	41.50	31.90	**35.95**	20.98
AC	36.45	29.16	34.85	22.14

Table: Part-of triples

Method	Precision (%)		F_1 (%)	
	Total	TS	Total	TS
Random	36.54	24.05	36.54	16.92
AF	37.30	25.05	37.30	17.63
RP 0.5	**52.99**	**28.55**	15.71	3.54
RP 0.2	47.71	28.43	37.53	13.51
AC	42.03	29.36	**40.83**	**19.51**

Table: Member-of triples

Method	Precision (%)		F_1 (%)	
	Total	TS	Total	TS
Random	32.17	27.35	32.17	23.41
AF	36.50	29.76	**36.12**	**25.06**
RP 0.5	**43.64**	**33.04**	23.00	12.13
RP 0.2	41.98	32.66	31.82	23.06
AC	37.05	30.69	35.45	24.67

Table: Purpose-of triples

Method	Precision (%)		F_1 (%)	
	Total	TS	Total	TS
Random	29.40	18.30	29.40	**15.02**
AF	29.14	17.54	29.14	14.40
RP 0.5	**49.73**	**23.63**	25.90	8.14
RP 0.2	40.69	21.62	30.01	11.81
AC	31.98	18.21	**31.36**	14.53

Gold resource creation

- Previous samples used as a starting point

Gold resource creation

- Previous samples used as a starting point
- Only tb-triples whose attachment to TePOT sets was *acceptable*

Gold resource creation

- Previous samples used as a starting point
- Only tb-triples whose attachment to TePOT sets was *acceptable*

Relation	$\overline{Possibilities}$	$\overline{Correct}$	$\overline{Correct\%}$
Hypernym-of	14.09	3.92	42.05%
Part-of	9.54	2.72	45.49%
Member-of	12.61	4.58	52.40%
Purpose-of	12.53	4.95	52.47%

Table: Matching possibilities in the gold resource.

Gold resource creation

- Previous samples used as a starting point
- Only tb-triples whose attachment to TePOT sets was *acceptable*

Relation	$\overline{Possibilities}$	$\overline{Correct}$	$\overline{Correct\%}$
Hypernym-of	14.09	3.92	42.05%
Part-of	9.54	2.72	45.49%
Member-of	12.61	4.58	52.40%
Purpose-of	12.53	4.95	52.47%

Table: Matching possibilities in the gold resource.

Method	**Hypernym-of** (135 tb-triples)		**Part-of** (72 tb-triples)		**Member-of** (105 tb-triples)		**Purpose-of** (60 tb-triples)	
	P (%)	F_1(%)	P (%)	F_1 (%)	P (%)	F_1 (%)	P (%)	F_1 (%)
Random	42.54	17.19	49.30	26.22	56.19	20.14	52.54	17.42
AF	43.18	17.25	45.83	24.63	52.48	18.21	46.55	15.21
RP 0.5	60.81	14.93	**66.67**	14.54	53.13	6.63	**68.75**	7.03
RP 0.2	57.45	17.34	58.70	22.31	56.25	13.21	59.38	11.55
AC	**61.16**	**22.77**	62.50	**30.77**	**70.00**	**22.07**	59.26	**18.23**

Table: Gold resource evaluation results.

Concluding remarks

- Ontologising without a context is a **challenging** task

Concluding remarks

- Ontologising without a context is a **challenging** task
- AC method performed better than the others

Concluding remarks

- Ontologising without a context is a **challenging** task
- AC method performed better than the others
- It is the method currently used in **Onto.PT**

Concluding remarks

- Ontologising without a context is a **challenging** task
- AC method performed better than the others
- It is the method currently used in **Onto.PT**

- **Onto.PT**: automatic construction of a lexical ontology for Portuguese (`http://ontopt.dei.uc.pt/`)

Concluding remarks

- Ontologising without a context is a **challenging** task
- AC method performed better than the others
- It is the method currently used in **Onto.PT**

- **Onto.PT**: automatic construction of a lexical ontology for Portuguese (`http://ontopt.dei.uc.pt/`)
  - Extraction of semantic relations

Concluding remarks

- Ontologising without a context is a **challenging** task
- AC method performed better than the others
- It is the method currently used in **Onto.PT**

- **Onto.PT**: automatic construction of a lexical ontology for Portuguese (`http://ontopt.dei.uc.pt/`)
  - Extraction of semantic relations
  - Discovery of synsets [Gonçalo Oliveira and Gomes, 2011a]/enrichment of existing thesaurus [Gonçalo Oliveira and Gomes, 2011b]

Concluding remarks

- Ontologising without a context is a **challenging** task
- AC method performed better than the others
- It is the method currently used in **Onto.PT**

- **Onto.PT**: automatic construction of a lexical ontology for Portuguese (`http://ontopt.dei.uc.pt/`)
  - Extraction of semantic relations
  - Discovery of synsets [Gonçalo Oliveira and Gomes, 2011a]/enrichment of existing thesaurus [Gonçalo Oliveira and Gomes, 2011b]
  - Ontologising the extracted relations

Concluding remarks

- Ontologising without a context is a **challenging** task
- AC method performed better than the others
- It is the method currently used in **Onto.PT**

- **Onto.PT**: automatic construction of a lexical ontology for Portuguese (`http://ontopt.dei.uc.pt/`)
  - Extraction of semantic relations
  - Discovery of synsets [Gonçalo Oliveira and Gomes, 2011a]/enrichment of existing thesaurus [Gonçalo Oliveira and Gomes, 2011b]
  - Ontologising the extracted relations
  - Soon freely available!

Thank you!

References I

[Gonçalo Oliveira and Gomes, 2011a] Gonçalo Oliveira, H. and Gomes, P. (2011a).
Automatic discovery of fuzzy synsets from dictionary definitions.
In *Proceedings 22nd International Joint Conference on Artificial Intelligence (IJCAI 2011)*, pages 1801–1806, Barcelona, Spain. AAAI Press.

[Gonçalo Oliveira and Gomes, 2011b] Gonçalo Oliveira, H. and Gomes, P. (2011b).
Automatically enriching a thesaurus with information from dictionaries.
In *Proceedings 15th Portuguese Conference on Artificial Intelligence (EPIA 2011)*, Lisbon, Portugal. APPIA.

[Gonçalo Oliveira et al., 2010] Gonçalo Oliveira, H., Santos, D., and Gomes, P. (2010).
Extracção de relações semânticas entre palavras a partir de um dicionário: o PAPEL e sua avaliação.
Linguamática, 2(1):77–93.
Nova versão, revista e aumentada, da publicação Gonçalo Oliveira et al (2009), no STIL 2009.

[Pantel, 2005] Pantel, P. (2005).
Inducing ontological co-occurrence vectors.
In *Proc. 43rd Annual Meeting of the ACL*, pages 125–132. ACL Press.

[Rocha and Santos, 2000] Rocha, P. A. and Santos, D. (2000).
CETEMPúblico: Um corpus de grandes dimensões de linguagem jornalística portuguesa.
In das Graças Volpe Nunes, M., editor, *V Encontro para o processamento computacional da língua portuguesa escrita e falada (PROPOR 2000)*, pages 131–140, São Paulo. ICMC/USP.