

SUPeRB

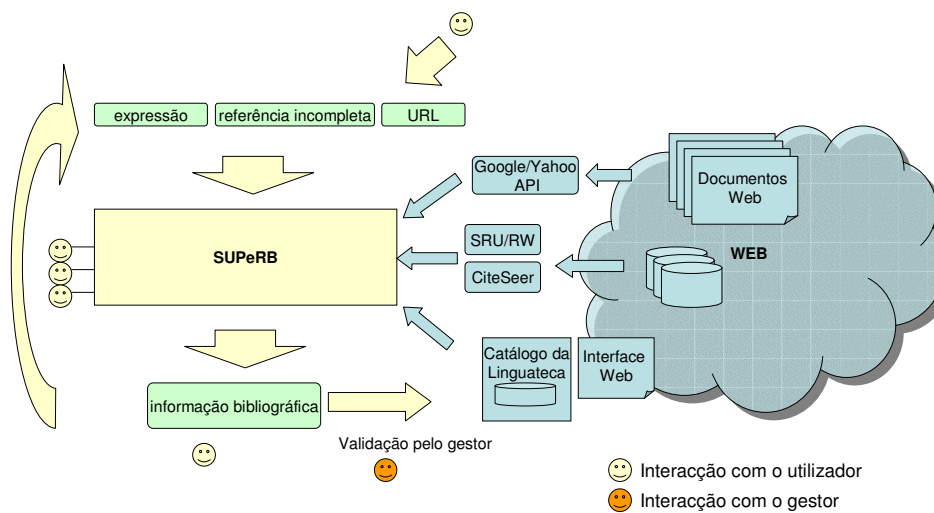
Especificação e exemplos

Luís Miguel Cabral
Linguateca

Ultima versão: 15 de Dezembro de 2006

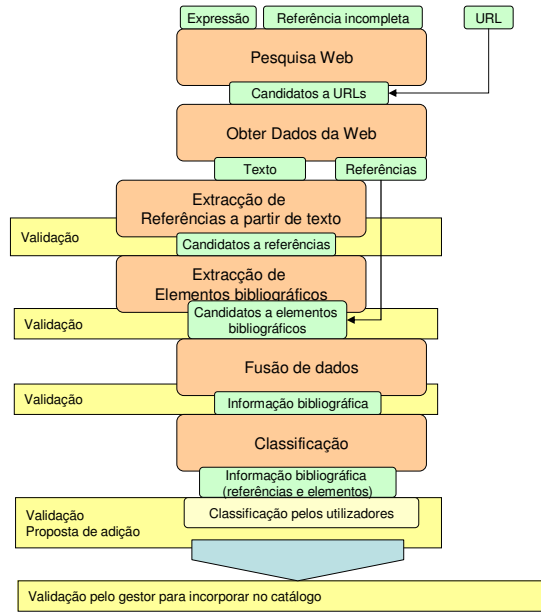
1

SUPeRB numa figura só



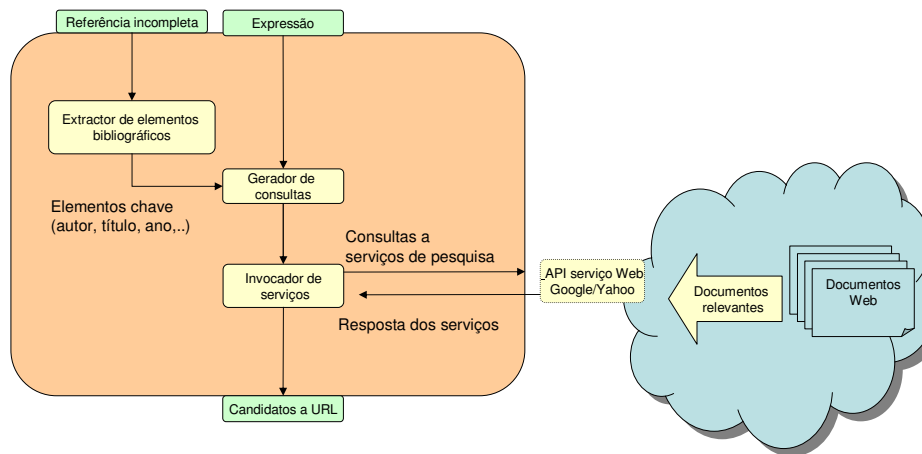
2

SUPeRB



3

Pesquisa na Web



4

Pesquisa na Web

Parâmetros de entrada:

Expressão ou referência bibliográfica incompleta

Saída:

Lista de URL com possível informação relevante

Exemplos:

Parâmetro dado: Dizer automatic discourse analyser

Expressões geradas:

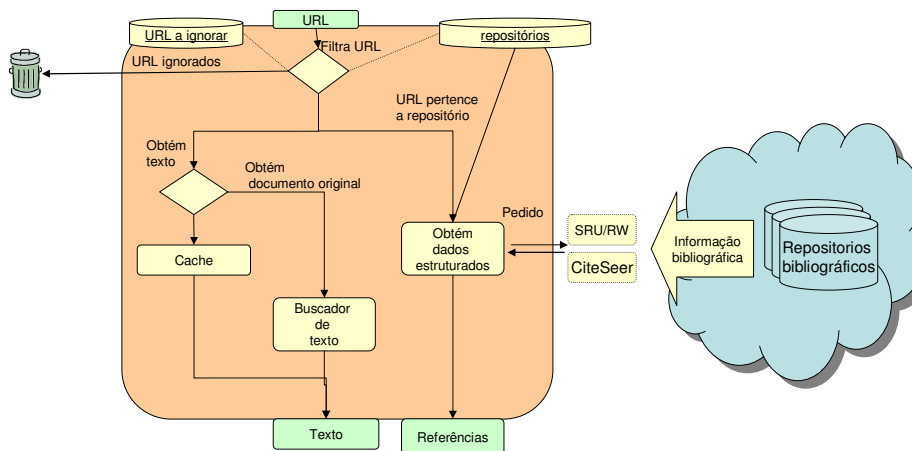
dizer automatic discourse analyser
dizer analyser automatic académico
discourse dizer automatic documents
discourse analyser dizer references
discourse dizer automatic academic

URL devolvidos:

<http://wotan.liu.edu/docis/dbl/sbiasb/index.html>
<http://www.portalpsicologia.org/busqueda.jsp?idProfesion=1>
<http://citeseer.ist.psu.edu/machadorino96discourse.html>
<http://br.groups.yahoo.com/group/ta-l/messages/30?xm=1&m=e>
http://www.shakespeare.uk.net/journal/1_3/cardoso-desousa1_3.html
<http://citeseer.ist.psu.edu/salgueiropardo04dizer.html>
<http://www.inesc-id.pt/pt/indicadores/Ficheiros/2167.pdf>
<http://csdl.tamu.edu/~crln/1986bib.html>
<http://www.csdl.tamu.edu/~crln/1991set.html>

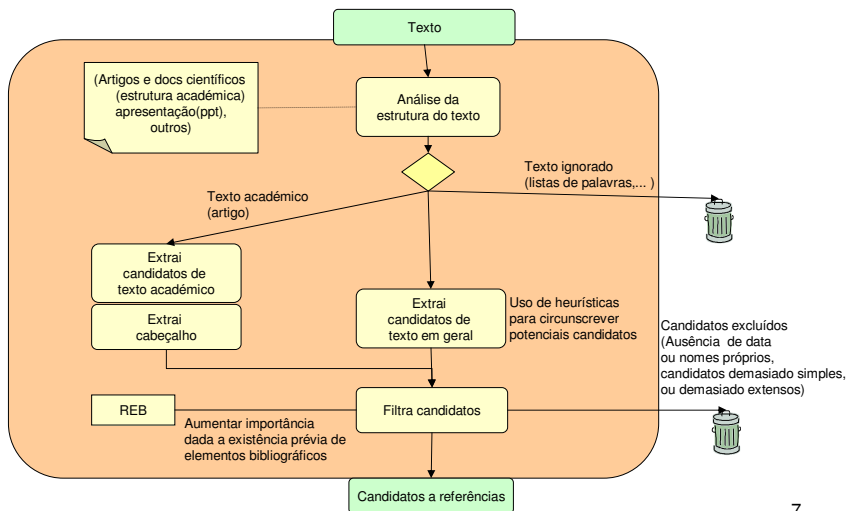
5

Acesso à Web para obter texto e informação bibliográfica



6

Extracção de referências de texto



7

Extracção de referências de texto

Parâmetros de entrada:

Texto extraído de um documento

Exemplo:

164
Data Driven Ontology Evaluation
Christopher Brewster, Harith Alani†, Srinandan Dasmahapatra‡, Yorick Wilks.
 .Department of Computer Science, University of Shef.eld,
 211 Portebello Street, Shef.eld, S1 4DP, U.K.
 {C.Brewster/Y.Wilks}@dcs.shef.ac.uk
 † School of Electronics and Computer Science,
 University of Southampton, SO17 1BJ, United Kingdom
 {ha/sd}@ecs.soton.ac.uk
Abstract
 The evaluation of ontologies is vital for the growth of the Semantic Web. We consider a number of problems in evaluating a knowledge artifact like an ontology. We propose in this paper that one approach to ontology evaluation should be corpus or data driven. A corpus is the most accessible form of knowledge and its use allows a measure to be derived of the 'i' between an ontology and a domain of ...
8. References
 AKT, 2003. AKT reference ontology. Available at www.aktors.org/publications/ontology. An ontology of the academic domain.
 Alani, Harith, Sanghee Kim, David E. Millard, Mark J. Weal, Wendy Hall, Paul H. Lewis, and Nigel R. Shadbolt, 2003. Automatic ontology-based knowledge extraction and tailored biography generation from the web. *IEEE Intelligent Systems*, 18(1).
 Brewster, Christopher, Fabio Ciravegna, and Yorick Wilks, 2001. Knowledge acquisition for knowledge management: Position paper. In *Proceeding of the IJCAI-2001 Workshop on Ontology Learning*. Seattle, WA: IJCAI.CIDOC, 2003. The CIDOC conceptual reference model.

8

Extracção de referências de texto

Dois tipos de saída:

1. A informação recolhida do cabeçalho está já semi-estruturada e, se a informação estiver disponível (fornecida pelas tarefas anteriores), com URL
2. Lista de referências candidatas, obtidas do texto.

Exemplo:

1. **Ttulo**=Data Driven Ontology Evaluation
autor=Christopher Brewster
autor=Harith Alani
autor=Srinandan
autor=Dasmahapatra
autor=Yorick Wilks
resumo=The evaluation of ontologies is ...
lingua=en

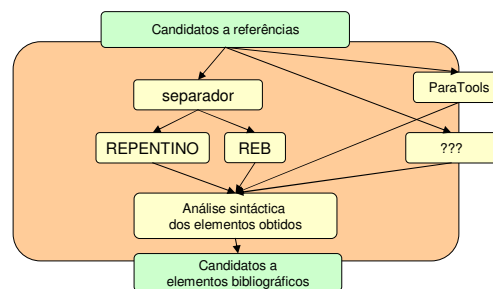
2. AKT, 2003. AKT reference ontology. Available at www.aktors.org/publications/ontology. An ontology of the academic domain.

Alani, Harith, Sanghee Kim, David E. Millard, Mark J. Weal, Wendy Hall, Paul H. Lewis, and Nigel R. Shadbolt, 2003. Automatic ontology-based knowledge extraction and tailored biography generation from the web. *IEEE Intelligent Systems*, 18(1).

Brewster, Christopher, Fabio Ciravegna, and Yorick Wilks, 2001. Knowledge acquisition for knowledge management: Position paper. In *Proceeding of the IJCAI-2001 Workshop on Ontology Learning*. Seattle, WA: IJCAI.CIDOC, 2003. The CIDOC conceptual reference model.

9

Extracção de elementos bibliográficos



10

Extracção de elementos bibliográficos

Parâmetros de entrada:

Referências em texto

Saída:

Informação estruturada, mantendo o texto fonte

Exemplos (1)

Luís Sarmento. "O SIEMÊS e a sua participação no HAREM e no Mini-HAREM". In Diana Santos & Nuno Cardoso (eds.), HAREM, a primeira avaliação conjunta de sistemas de reconhecimento de entidades mencionadas para português: documentação e actas do encontro. Linguateca., 2006.



autor=Luís Sarmento
titulo=O SIEMÊS e a sua participação no HAREM e no Mini-HAREM
editor=Diana Santos
editor=Nuno Cardoso
colecacao=HAREM, a primeira avaliação conjunta de sistemas de reconhecimento de entidades mencionadas para português: documentação e actas do encontro
ano=2006.
texto=Luís Sarmento. "O SIEMÊS e a sua participação no HAREM e no Mini-HAREM". In Diana Santos & Nuno Cardoso (eds.), HAREM, a primeira avaliação conjunta de sistemas de reconhecimento de entidades mencionadas para português: documentação e actas do encontro. Linguateca., 2006.

11

Extracção de elementos bibliográficos

Exemplos (2)

Diana Santos & Paulo Rocha. "CHAVE: topics and questions on the Portuguese participation in CLEF". In Carol Peters & Francesca Borri (eds.), Cross Language Evaluation Forum: Working Notes for the CLEF 2004 Workshop (CLEF 2004) (Bath, UK, 15-17 September 2004), Pisa, Italy: IST-CNR, pp. 639-648. Revised as Santos & Rocha (2005) <http://www.linguateca.pt/documentos/SantosRochaCLEF2004WN.pdf>



autor=Diana Santos
autor=Paulo Rocha
titulo=CHAVE: topics and questions on the Portuguese participation in CLEF
editor=Carol Peters
editor=Francesca Borri
colecacao=Cross Language Evaluation Forum: Working Notes for the CLEF 2004 Workshop (CLEF 2004)
local=Bath, UK
data= 15-17 September 2004
morada=Pisa, Italy
editora=IST-CNR
pagina=639-648
url= <http://www.linguateca.pt/documentos/SantosRochaCLEF2004WN.pdf>
texto=Diana Santos & Paulo Rocha. "CHAVE: topics and questions on the Portuguese participation in CLEF". In Carol Peters & Francesca Borri (eds.), Cross Language Evaluation Forum: Working Notes for the CLEF 2004 Workshop (CLEF 2004) (Bath, UK, 15-17 September 2004), Pisa, Italy: IST-CNR, pp. 639-648. Revised as Santos & Rocha (2005) <http://www.linguateca.pt/documentos/SantosRochaCLEF2004WN.pdf>

12

Extracção de elementos bibliográficos

Exemplos (3)

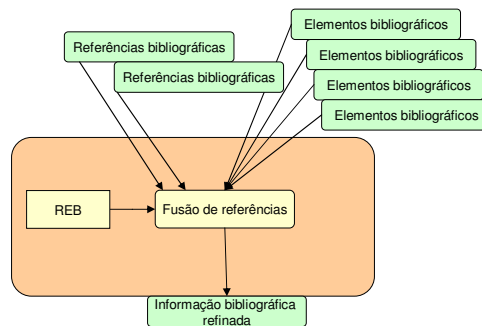
Alani, Harith, Sanghee Kim, David E. Millard, Mark J. Weal, Wendy Hall, Paul H. Lewis, and Nigel R. Shadbolt, 2003. Automatic ontology-based knowledge extraction and tailored biography generation from the web. *IEEE Intelligent Systems*, 18(1).



```
autor= Harith Alani
autor=Sanghee Kim
autor=David E. Millard
autor=Mark J. Weal
author= Wendy Hall
autor=Paul H. Lewis
autor=Nigel R. Shadbolt
ano=2003
titulo=Automatic ontology-based knowledge extraction and tailored biography generation from the web
coleccion=IEEE Intelligent Systems
volume=18
numero=1
texto=Alani, Harith, Sanghee Kim, David E. Millard, Mark J. Weal, Wendy Hall, Paul H. Lewis, and Nigel R. Shadbolt, 2003. Automatic ontology-based knowledge extraction and tailored biography generation from the web. IEEE Intelligent Systems, 18(1).
```

13

Fusão da informação bibliográfica



14

Fusão da informação bibliográfica

Exemplo:

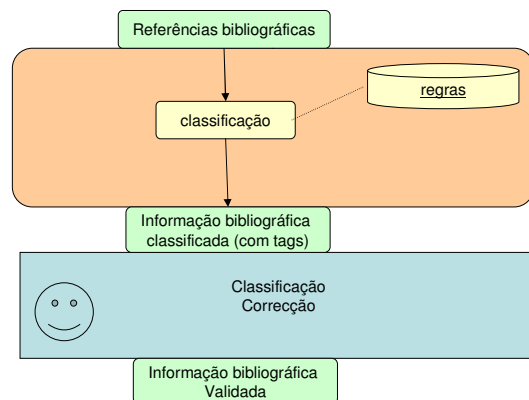
tipo=artigo
titulo=Applying a Lexical Similarity Measure to Compare Portuguese Term Collections
autor=Chaves, M.S.
autor=Lima, V.L.S.
revista=LECTURE NOTES IN COMPUTER SCIENCE
ano=2004
editora=Springer

tipo=livro
revista=Lecture Notes in Computer Science : Advances in Artificial Intelligence – SBIA 2004
titulo=Applying a Lexical Similarity Measure to Compare Portuguese Term Collections
páginas=194 - 203
ano=2004
autor=Marcirio Silveira Chaves
autor=Vera Lúcia Strube de Lima

tipo=artigo
titulo=Applying a Lexical Similarity Measure to Compare Portuguese Term Collections
autor=**Marcirio Silveira Chaves**
autor=**Vera Lúcia Strube de Lima**
revista=**Lecture Notes in Computer Science : Advances in Artificial Intelligence – SBIA 2004**
páginas=**194-203**
ano=2004
editora=**Springer**

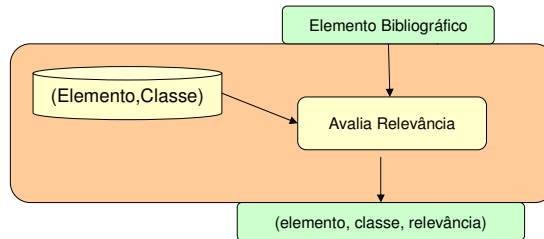
15

Classificação das Referências



16

REB- Repositório de Elementos Bibliográficos

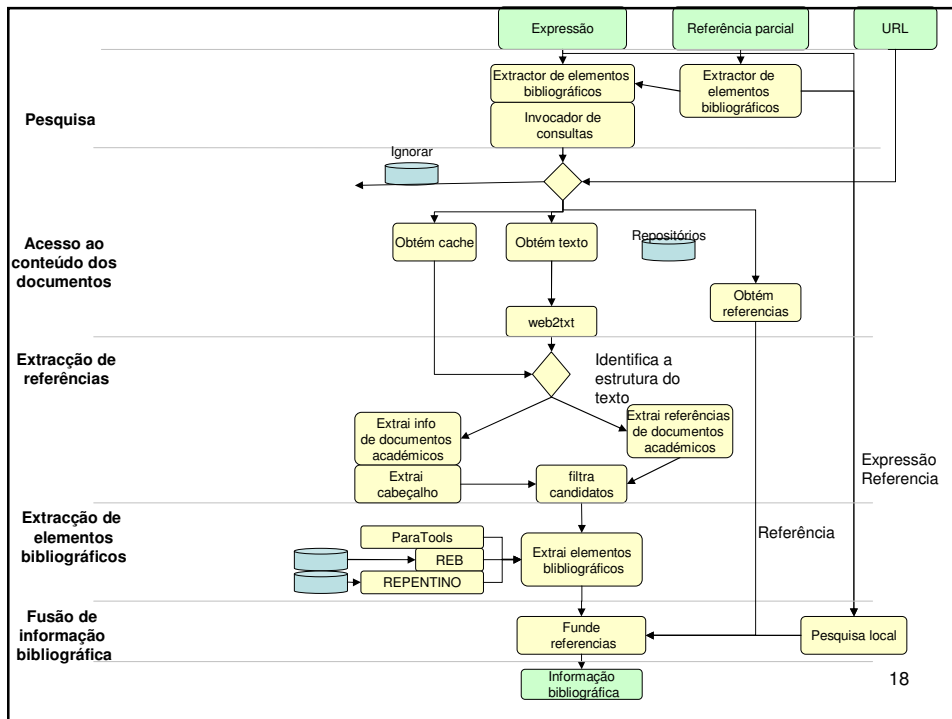


Exemplo:

Parâmetro: "M. Chaves"

Saída: [{"Marcirio Silveira Chaves", "pessoa", 21.0}, {"Marcirio Chaves", "pessoa", 7.3}]

17



18