

TP1: Expressões Regulares e Filtros de Texto

Relatório de Desenvolvimento da Questão 4

Grupo 8

Ana Beatriz Ribeiro e Silva (A91678)

Beatriz Fernandes Oliveira (A91640)

Conteúdo

1	Introdução	2
2	Descrição do Problema	3
3	Decisões de Implementação	4
4	Conclusão	5
A	Ficheiro de Input	6
B	Código do Programa	43
B.1	Alínea a e b	43
B.2	Alínea c	44

1 Introdução

Neste relatório, pretende-se expor o processo de resolução do trabalho prático proposto cujo tema é *Expressões Regulares e Filtros de Texto*.

Uma expressão regular pode ser definida como um padrão de caracteres que se associa a uma ou várias sequências de caracteres num determinado texto ou documento, que neste caso será um documento no formato BibTex.

BibTeX é uma ferramenta de formatação de citações bibliográficas em documentos LATEX, criada com o propósito de facilitar a separação da bibliografia com a apresentação do documento.

À vista disso, iremos ser confrontadas com um documento para o qual teremos de criar enumeras expressões regulares que permitam extrair a informação solicitada do mesmo. Para além disso, para concluirmos com sucesso este projeto será também necessário tirar proveito das diferentes funções e funcionalidades do módulo ‘re’ de Python.

2 Descrição do Problema

Tal como já foi referido anteriormente, iremos focar-nos num documento com o formato BibTex (A). É de realçar que, a categoria de uma referência corresponde à primeira palavra antecedida por um '@'.

O problema encontra-se dividido em quatro alíneas. No entanto, como as duas primeiras alíneas complementam-se, iremos proceder à divisão do problema em dois, ou seja, as duas primeiras alíneas correspondem ao primeiro problema e as seguintes alíneas correspondem ao segundo problema.

Assim sendo, inicialmente é nos pedido, que depois de uma análise do documento, façamos um programa que conte as diferentes categorias que ocorrem no mesmo e que produza um documento em HTML que contenha a designação das categorias encontradas, bem como as suas respectivas contagens. É ainda requisitado que, para cada categoria existente, anotemos, no HTML, as respetivas chaves, autores e títulos.

No segundo problema, pretende-se, novamente a partir do ficheiro dado, que criemos um índice de autores e, ainda, um grafo que mostre, para um determinado autor, todos os autores que publicaram com o autor em causa.

3 Decisões de Implementação

Para a segunda parte do primeiro problema, seguimos um processo muito semelhante ao anterior, utilizando o *regex* e o método tentativa-erro, com o propósito de adquirir as diferentes expressões necessárias para obter os autores, os títulos e as chaves. Sendo que, as expressões obtidas foram, respetivamente:

- $\wedge (\backslash s)^*(?i:author)(\backslash s)*=(\backslash s)*((\{|\backslash " \} [^}]|"\backslash }*(\backslash }|\backslash "))$
- $\wedge (\backslash s)^*(?i:title)(\backslash s)*=(\backslash s)*((\{|\backslash " \} [^}]|"\backslash }*(\backslash }|\backslash "))$
- $\backslash w+|(\backslash w+[:]\backslash w+)),\$$

Tal como pedido, foi necessário a criação de um ficheiro em HTML, que segue um formato bastante distinto dos demais ficheiros. Ou seja, para que o ficheiro seja corretamente interpretado pelos browsers, é necessário que siga o formato abaixo.

```
<!DOCTYPE html>
<html>
<head>
<meta charset="UTF-8"/>
<title>Document</title>
</head>
<body>
<!-- Conteúdo -->
</body>
</html>
```

No que diz respeito à sua implementação, de forma muito sucinta, começamos por abrir o BibTex em modo de leitura e, posteriormente, percorremos o ficheiro de forma a obter as diferentes categorias e as respetivas contagens, guardando esses valores num dicionário. Fazendo uma segunda passagem pelo ficheiro, guardamos para uma dada categoria, a sua chave, os seus autores e títulos noutro dicionário. Finalmente, percorremos os dois dicionários e escrevemos toda a informação no ficheiro de output de HTML.

Para resolver o segundo problema, foi necessário reutilizar a expressão regular descrita acima que nos permitiu obter os autores existentes no documento e guardá-los numa lista. Como muitos dos autores correspondem a uma conjunção de dois diferentes autores, foi necessário percorrer essa lista, de forma a separar cada conjunção em dois autores. Com o propósito de obtermos um índice de autores ordenada alfabeticamente, aplicamos, ainda, à lista dos autores o método *sorted*. Foi, ainda, necessário para tratar os nomes dos autores, criar um filtro que filtre os que já estão em abreviatura:

- $(.*)\backslash .\backslash .\backslash .(*)$

4 Conclusão

Durante a resolução deste trabalho, tivemos de superar diversos desafios, entre eles, aprender a trabalhar com novas ferramentas com as quais não estávamos familiarizadas.

Tendo em conta que os principais objetivos do trabalho prático eram: melhorar a nossa capacidade de escrever *expressões regulares (ER)*; desenvolver *filtros de texto (FT)*; e, por último, utilizar o módulo 're' de python, podemos afirmar que as nossas capacidades, nos três campos, melhoraram exponencialmente durante a realização do mesmo.

Desta forma, a resolução do problema foi longa e trabalhosa, mas também recompensadora, visto que, no final desta, adquirimos um novo conjunto de habilidades que nos seriam desconhecidas caso contrário.

No que diz respeito a melhorias, consideramos que devíamos ter iniciado este trabalho prático mais cedo, de forma a que nos permitisse lidar mais atempadamente com os erros e com algumas dificuldades. Consideramos, ainda, que a resolução dos problemas pode ser mais aprofundada e melhorada.

A Ficheiro de Input

```
@inproceedings{graminteractivas1990,
  author={F. Mário Martins and J.J. Almeida and P.R. Henriques},
  title = {Mecanismos para Especificação e Prototipagem de Interfaces
    Utilizador-Sistema},
  note = {(Gramáticas Interactivas guardadas)},
  booktitle={3o Encontro Português de Computação Gráfica},
  address="Coimbra",
  year = 1990,
}

@techreport{tlc89,
  author={J.J. Almeida and J.B. Barros},
  title = {Teoria das Linguagens },
  institution = "umdi",
  type = "Texto didáctico",
  year = 1988,
  keyword = "compiladores",
}

@techreport{estruturasdedados90,
  author={J.B. Barros and J.J. Almeida},
  title = {Estruturas de Dados},
  institution = "umdi",
  type = "Texto didáctico",
  year = 1990,
  keyword = "algoritmos",
}

@techreport{Camila,
  author ={{projecto Camila}},
  editor ={L.S. Barbosa and J.J. Almeida and J.N. Oliveira and Luís Neves},
  title = {\textsc{Camila} - A Platform for Software Mathematical Development},
  url="http://camila.di.uminho.pt",
  type="(Páginas do projecto)",
  institution = "umdi",
  year=1998,
  keyword = "FS",
}

@techreport{Natura,
  author ={J.J. Almeida},
  title = "{Natura} - Natural language processing",
  url="http://natura.di.uminho.pt/",
  type="(Páginas do projecto)",
  note="\url{http://natura.di.uminho.pt/}",
  institution = "umdi",
  keyword = "PLN",
  year = 1997,
}

@techreport{jspell1,
  author = "J.J. Almeida and Ulisses Pinto",
  title = "Manual de Utilizador do {JSpell}",
  year = 1994,
  type = "Manual",
  month = "Jul",
  institution = "umdi",
  keyword = "morphology, lexical analysis,jspell",
  abstract = {},
```

```

    url = "http://natura.di.uminho.pt/~jj/pln/jspellman.ps.gz",
}

@inproceedings{Almeida94b,
    author = "J.J. Almeida",
    title = "{GPC} -- a Tool for higher-order grammar specification",
    booktitle = "Actas del X Congreso de Lenguajes Naturales e Leanguajes Formales, Sevilla",
    year = 1994,
    url = "http://natura.di.uminho.pt/~jj/pln/yalg3.ps.gz",
    editor="Carlos Martin Vide",
    keyword ="DCG,grammar",
}

@inproceedings{Almeida95a,
    author = "J.J. Almeida",
    title = "{YaLG} -- extending {DCG} for natural language processing",
    booktitle = "Actas del XI Congreso de Lenguajes Naturales e Leanguajes Formales, Tortosa",
    year = 1995,
    pages = "621--628",
    editor="Carlos Martin Vide",
    url = "http://natura.di.uminho.pt/~jj/pln/yalg.ps.gz",
    keyword ="jspell,morphology,PLN,DCG,nllex",
}

@inproceedings{Almeida94c,
    author = "J.J. Almeida and Ulisses Pinto",
    title = "Jspell -- um módulo para análise léxica genérica de linguagem natural",
    booktitle = "Actas do X Encontro da Associação Portuguesa de Linguística",
    address={Évora 1994},
    pages = "1--15",
    year = 1995,
    url = "http://natura.di.uminho.pt/~jj/pln/jspell1.ps.gz",
    keyword ="jspell,morphology,PLN,perl",
}

@techreport{jj95,
    author = "J.J. Almeida",
    title = "{NLlex} -- a tool to generate lexical analysers for natural language",
    institution = "umdi",
    year = 1995,
    number = "UM-DI-95.04",
    url = "http://natura.di.uminho.pt/~jj/pln/nllex.ps.gz",
    keyword ="jspell,morphology,lex,PLN,nllex",
}

@techreport{Barbosa95,
    author = "L.S. Barbosa and J.J. Almeida",
    title = "System Prototyping in \textsc{Camila}",
    year = 1995,
    number = "DI-CAM-95:11:1",
    institution="University of Minho",
    note = "Lecture notes for the System Design Course,
           Computer System Engineering, University of Bristol",
    url="http://www.di.uminho.pt/~lsb/pub_camila/LNcam.ps.gz",
    keyword ="Camila, formal specification",
}

@techreport{Barbosa95a,
    author = "L.S. Barbosa and J.J. Almeida",
    title = "\textsc{Camila}: A reference Manual",
    year = 1995,

```



```

number = "DI-CAM-95:11:2",
institution="University of Minho",
url="http://www.di.uminho.pt/~lsb/pub_camila/RMcam.ps.gz",
keyword = "Camila",
}

@techreport{BA97a,
  keyword = "Formal Methods, Prototyping, Camila",
  author = "L.S. Barbosa and J.J. Almeida",
  title = "Systems Prototyping in \textsc{Camila}",
  type = "{Lecture Notes for the Bristol Course (1st ed. 1995)}",
  year = 1998,
  institution = "DI (U. Minho)",
  number = "DI-CAM-95:11:1:v98"
}

@techreport{Barbosa95b,
  author = "L.S. Barbosa and J.J. Almeida",
  title = "Growing Up With \textsc{Camila}",
  institution = "umdi",
  year = 1995,
  number = "DI-CAM-95:7:1",
  url = "http://www.di.uminho.pt/~lsb/pub_camila/romantic.ps.gz",
  keyword = "Camila, formal specification, didactics",
}

@inproceedings{Ramalho95,
  author = "J.C. Ramalho and J.J. Almeida and P.R. Henriques",
  title = "Algebraic Specification of Documents",
  booktitle = "TWLT10 - Algebraic Methods in Language Processing",
  year = 1995,
  month = "6--8 Dec.",
  editor = "A. Nijholt and G. Scollo and R. Steetskamp",
  address = "Twente University, Netherlands",
  note = "AMiLP'95",
  series = "Twente Workshop on Language Technology",
  url="http://natura.di.uminho.pt/~jj/bib/amilp95.ps.gz",
  docpage="http://www.di.uminho.pt/~jcr/projectos/david/ARTIGOS/AMiLP95/amilp95.html",
  pages = "55--64",
  keyword = "PDavid, Camila, SGML",
}

@inproceedings{Almeida96a,
  author = "J.J. Almeida",
  title = "Especificação e tratamento de Dicionários",
  booktitle = "Actas do XI Encontro da Associação Portuguesa de Linguística",
  address={Lisboa 1995},
  year = 1996,
  volume=2,
  url = "http://natura.di.uminho.pt/~jj/pln/etdic.ps.gz",
  keyword = "perl,morphology,lexical analysis, dictionary",
}

@inproceedings{Ulisses96,
  author = "Ulisses Pinto and J.J. Almeida",
  title = "Tratamento automático de termos compostos",
  booktitle = "Actas do XI Encontro da Associação Portuguesa de Linguística",
  address= "Lisboa 1995",
  volume=2,
  year = 1996,
  url = "http://natura.di.uminho.pt/~jj/pln/ptc.ps.gz",

```

```

    keyword = "jspell,morphology,lexical analysis, PLN",
}

@inproceedings{Almeida96b,
    author = "J.J. Almeida and J.B. Barros",
    title = "{YaLG} a tool for higher-order grammar specification",
    booktitle = " II International Conference on Mathematical Linguistics, Tarragona, Spain",
    year = 1996,
    url="http://natura.di.uminho.pt/~jj/pln/yalg2.ps.gz",
    keyword ="yalg,DCG,RS",
}

@article{jj96,
    author = "J.J. Almeida",
    title = "{NLlex} -- a tool to generate lexical analysers for natural language",
    year = 1996,
    month = "Sep",
    volume = "19",
    pages = "81--90",
    journal = {Procesamiento del Lenguaje Natural},
    publisher="Sociedade Española para el Procesamiento del Lenguaje Natural",
    keyword ="jspell,morphology,lex,PLN,nllex",
    url= "http://natura.di.uminho.pt/~jj/pln/nllex2.ps.gz",
}

@techreport{Almeida96c,
    author = "J.J. Almeida and J.C. Ramalho",
    title = "From {BiBTeX} to {HTML} semantic nets",
    institution = "umdi",
    year = 1996,
    number = "DI-DAV-96:1:1",
    docpage = "http://www.di.uminho.pt/~jcr/bib/dbib.html",
    keyword ="PDavid, bibtex, librarian studies,html",
}

@inproceedings{Ramalho96,
    author = "J.C. Ramalho and J.J. Almeida and P.R. Henriques",
    title = "Document Semantics: two approaches",
    booktitle = "Celebrating a Decade of SGML",
    year = 1996,
    month = "Nov.",
    address = "Boston, USA",
    note = "SGML'96 Conference",
    editor = "Graphic Communications Association",
    publisher = "ArborText",
    pages = "473--483",
    docpage="http://www.di.uminho.pt/~jcr/projectos/david/COMS/sgml96/sgml96.html",
    keyword ="PDavid, Camila, SGML,AG,FS",
}

@InProceedings{SGML97,
    author = "J.C. Ramalho and J.G. Rocha and J.J. Almeida and P.R. Henriques",
    title = "SGML Documents: where does quality go?",
    booktitle = "SGML/XML'97 Conference",
    year = 1997,
    address = "Washington D.C. - USA",
    month = "Dec.",
    keyword = "PDavid, SGML, Semantics",
}

@inproceedings{Almeida98,

```

```

author = "J.J. Almeida",
title = "Programação de dicionários",
booktitle= "Actas do XIII Encontro da Associação Portuguesa de Linguística",
year=1998,
pages="21--28",
volume= "1",
address = "Lisboa 1997",
keyword = "perl,morphology,dictionary,parser",
url= "http://natura.di.uminho.pt/~jj/bib/progDic.ps.gz",
}

@inproceedings{Reis98,
author = "Ricardo Reis and J.J. Almeida",
title = "Etiquetador morfo-sintáctico para o Português",
booktitle= "Actas do XIII Encontro da Associação Portuguesa de Linguística",
address = "Lisboa 1997",
year=1998,
keyword ="tagger, PLN",
url= "http://natura.di.uminho.pt/~jj/bib/etiquetador2.ps.gz",
}

@inproceedings{ABN097a,
title = "\textsc{Camila}: Formal Software Engineering Supported by Functional Programming",
author = "J.J. Almeida and Barbosa, L.S. and Neves, F.L. and Oliveira, J.N.",
booktitle = "Proc. II Conf. Latino Americana de Programaci\on Funcional ({CLaPF97})",
editor = "De Giusti, A. and Diaz, J. and Pesado, P.",
year = 1997,
month = "October",
address = "La Plata, Argentina",
pages = "1343--1358",
keyword ="Camila, formal specification",
url = "http://camila.di.uminho.pt/camila-doc/CLaPF97.ps.gz",
}

@inproceedings{ABN097b,
title = "\textsc{Camila}: Prototyping and Refinement of Constructive Specifications",
author = "J.J. Almeida and Barbosa, L.S. and Neves, F.L. and Oliveira, J.N.",
booktitle = "6th International Conference on Algebraic Methods and Software Technology ({AMAST'97})",
editor = "Johnson, M.",
year = 1997,
month = "December",
address = "Sydney, Australia",
publisher = "Springer Lect. Notes Comp. Sci. (1349)",
pages = "554--559",
keyword ="Camila, formal specification",
}

@inproceedings{AH97,
title = "Dynamic Dictionary = cooperative information sources",
author = "J.J. Almeida and P.R. Henriques",
year = 1998,
address = "Australia",
url = "http://natura.di.uminho.pt/~jj/bib/agentes97.ps.gz",
keyword ="dictionary, Agentes",
booktitle = "Proc. II Conference on Knowledge-based Intelligent Electronic Systems ({Kes98})",
month = "April",
}

@inproceedings{museums98,
author={J.G. Rocha and M.R. Henriques and J.C. Ramalho and J.J. Almeida

```

```

    and J.L. Faria and P.R. Henriques},
    title={Adapting Museum Structures for the Web: No Changes Needed!},
    booktitle = "Museums and the Web 1998",
    note = "Toronto - Canadá",
    year= 1998,
}

```

```

@inproceedings{ABBN98,
    author = "Almeida, J.J. and Barbosa, L.S. and Barros, J.B. and
        Neves, L.F.",
    title = "On The Development of \textsc{Camila}",
    booktitle = "Workshop on Research Themes on Functional Programming",
    year = 1998,
    month = "18 Sep.",
    editor = "L.S. Barbosa and J.A. Saraiva",
    publisher = "Proc. 3rd Summer School on Advan. Funct. Prog., Braga"
}

```

```

@article{Ramalho98,
    author = "J.C. Ramalho and J.J. Almeida and P.R. Henriques",
    title = "Algebraic specification of documents",
    year = 1998,
    number = "199",
    pages = "231--247",
    journal = "Theoretical Computer Science",
    url="http://natura.di.uminho.pt/~jj/bib/amilp95.ps.gz",
    docpage="http://www.di.uminho.pt/~jcr/projectos/david/ARTIGOS/AMiLP95/amilp95.html",
    keyword ="PDavid, Camila, SGML",
}

```

```

@inproceedings{Gis99,
    author = "Jorge Rocha and Ana Silva and Ricardo Henriques and J.J. Almeida and Pedro Henriques",
    title={Formal Methods for {GI} Systems Development},
    booktitle = "Conferência da Association of Geographic Information
Laboratories for Europe (AGILE)",
    address="Roma",
    year=1999,
    keyword="GIS",
}

```

```

@inproceedings{RPA99,
    author={Jorge Rocha and Tiago Pedroso and J.J. Almeida},
    title ={{MAPit} - A tool set for automatically generation of {HTML} Maps},
    booktitle = "Conferência da Association of Geographic Information
Laboratories for Europe (AGILE)",
    address="Roma",
    year=1999,
    keyword="GIS, XML, mapit",
}

```

```

@inproceedings{RSea99,
    author = {Jorge Gustavo Rocha and Ana Silva and J.J. Almeida and Mario Ricardo Henriques and Pedro Range
    title= {Sobre a Utilização de Metodologias Formais no Desenvolvimento de
{SIG}},
    booktitle = {GISBRASIL'99, Salvador},
    year=1999,
    keyword="GIS",
}

```

```

@inproceedings{xmldt99,
    author = { J.J. Almeida and José Carlos Ramalho},

```

```

    title = {{XML::DT} a Perl Down-Translation module},
    booktitle = "XML-Europe'99, Granada - Espanha",
    month = "May",
    year= 1999,
    keyword="XML, perl",
}

@article{RRAH99,
  author={J.C. Ramalho and J.G. Rocha and J.J. Almeida and P.R. Henriques},
  title = {SGML documents: Where does quality go?},
  journal={Markup Languages: theory and practice},
  Volume = "1",
  pages = "75--90",
  publisher="MIT Press",
  year= 1999,
  keyword="SGML",
}

@inproceedings{Barbosa2000,
  author = "L.S. Barbosa and J.B. Barros and J.J. Almeida",
  title = "Polytypic Recursion Patterns",
  year = 2000,
  month = "May",
  booktitle = "{SBLP'2000} (to appear as a ENTCS volume)",
  address = "{UFP}, Recife, Brasil",
  keyword="FS, Camila",
}

@inproceedings{jj2001x,
  author = "J.J. Almeida",
  title = "Smallbook -- comando para produção de livros em pequena escala",
  year = 2000,
  pages = "445--450",
  booktitle = "Actas da II Conferência Internacional de Tecnologias de
Informação e Comunicação na Educação",
  address="Braga",
  keyword="publishing, latex, smallbook",
}

@article{speaker:sepln2001,
  author={J.J. Almeida and A. M. Simões},
  title = {Text to speech -- a rewriting system approach},
  journal = {Procesamiento del Lenguaje Natural},
  address="Sevilha",
  volume = "27",
  pages = "247--255",
  publisher="Sociedade Española para el Procesamiento del Lenguaje Natural",
  month="Sep.",
  year= 2001,
  keyword="TTS,RS,fonética",
}

@inproceedings{mp2001,
  author= {J.J. Almeida and J. Gustavo Rocha and P. Rangel Henriques and
Sónia Moreira and Alberto Simões},
  title = {{Museu da Pessoa} -- Arquitectura} ,
  booktitle = {Congresso Nacional de Bibliotecários, Arquivistas e
Documentalistas},
  address = {Porto},
  url = "http://natura.di.uminho.pt/~jj/bib/museuDaPessoa2001.ps.gz",
  month = "Maio",

```

```
year= 2001,  
}
```

```
@inproceedings{alfarrabio2001,  
author= {J.J. Almeida and P. Rangel Henriques and J. Gustavo Rocha and  
        Alberto Simões},  
title = {Alfarrábio: Adding value to an Heterogeneous Site Collection} ,  
booktitle = {Congresso Nacional de Bibliotecários, Arquivistas e  
            Documentalistas},  
address = {Porto},  
url = "http://natura.di.uminho.pt/~jj/bib/alfarrabio2001.ps.gz",  
month = "Maio",  
year= 2001,  
}
```

```
@inproceedings{freq2002,  
author= {Paulo A. Rocha and Alberto M. Simões and J.J. Almeida},  
title={ Cálculo de frequências de  
palavras para entradas de dicionários através do uso conjunto de analisadores  
morfológicos, taggers e corpora},  
booktitle={Actas do XVII Encontro da Associação Portuguesa de Linguística},  
address={Lisboa 2001},  
pages="407--418",  
year= 2002,  
url = "http://natura.di.uminho.pt/~jj/bib/apl:freqnormpt.ps.gz",  
abstract = {Apresentamos neste documento uma possível abordagem à  
            extracção de frequências de palavras a partir de  
            corpora, baseada numa utilização cooperativa de várias  
            ferramentas de Processamento de Linguagem Natural.},  
}
```

```
@inproceedings{jspell2002,  
author= {Alberto M. Simões and J.J. Almeida},  
title={ Jspell.pm -- um módulo de análise morfológica  
para uso em processamento de linguagem natural},  
booktitle={Actas do XVII Encontro da Associação Portuguesa de Linguística},  
address={Lisboa 2001},  
pages="485--495",  
abstract = {Neste documento é nosso propósito apresentar as  
            características presentes no analisador morfológico  
            jspell e quais as suas consequências ao nível de  
            aplicações de processamento de linguagem natural. Como  
            ferramenta que é frequentemente integrada em software  
            mais específico, apresentamos um módulo Perl  
            desenvolvido com o objectivo de facilitar a interligação  
            do analisador morfológico com pequenas aplicações  
            desenvolvidas em linguagens de scripting. Devido à  
            constante necessidade de melhoramento de dicionários, e  
            em particular dos analisadores morfológicos, discutimos  
            as propriedades que estes devem conter para facilitar o  
            seu processamento e enriquecimento automático.},  
year= 2002,  
}
```

```
@inproceedings{dag2002,  
author= {Alberto M. Simões and J.J. Almeida and Pedro R. Henriques},  
title={Directory Attribute Grammars},  
booktitle={VI Simpósio Brasileiro de Linguagens de Programação},  
pages = {297--308},  
address = {Rio de Janeiro, Brasil},  
year= 2002,
```

```

}

@inproceedings{elpub2002,
  author= {Alberto M. Simões and J.J. Almeida},
  title={Library::* -- a toolkit for digital libraries},
  booktitle={Elpub 2002 -- International Conference on Electronic Publishing},
  address = {Karlovy Vary, República Checa},
  month="Nov.",
  pages = {203--211},
  year= 2002,
  isbn = "3-89700-357-0",
  abstract = {
    In last years the amount of digital documents has increased
    dramatically. Unfortunately the same did not occur with the
    structure and organization of the information. Traditionally we
    built a digital library using a catalog with documents'
    meta-information including a conceptual classification and an
    ontology of concepts.

    In this document we present a set of modules to help in the task of
    building and maintaining a digital library. It includes a module to
    work with ontologies, a set of modules to handle specific catalog
    formats (like Bib\TeX), a module to define new catalog formats
    and a tool to integrate ontologies and multi-format
    catalogs in a web browse-able knowledge-base.
  }
}

@article{parguess2002,
  author= {J.J. Almeida and Alberto M. Simões and J. Alves de Castro},
  title={Grabbing parallel corpora from the web},
  publisher="Sociedade Española para el Procesamiento del Lenguaje Natural",
  journal = {Procesamiento del Lenguaje Natural},
  volume ="29",
  pages = "13--20",
  month="Sep.",
  year= 2002,
  abstract = {
    Multilingual resources are useful for linguistic studies, translation,
    and many other tasks. Unfortunately, these resources are difficult to obtain
    and organize.

    In this document we describe a set of tools designed to help in the
    task of mining bilingual resources from the web, from a specific site,
    from a file system, from a list of URLs, or from a translation memory.

    As a design goal we intend to build tools that can be used both
    cooperatively (in pipeline) and also in a independent way. }
}

@InProceedings{APL2k2.Parguess,
  author = "J.J. Almeida and Alberto Manuel Simões and José Alves Castro",
  title = "Extracção de corpora paralelo a partir da web: construção e
  disponibilização",
  booktitle={Actas do XVIII Encontro da Associação Portuguesa de Linguística},
  address={Porto 2002},
  year = "2003",
  lang = "PT",
  url = {http://alfarrabio.di.uminho.pt/~albie/publications/APL2k2.Parguess.pdf},
  abstract = {
    Ao longo deste documento descrever-se-á um conjunto de ferramentas
    construídas para extracção automática de recursos bilingues a partir
    da Web, a partir de um \emph{site} específico, a partir de um

```

sistema de ficheiros contendo alguns textos que sejam traduções de outros, ou ainda a partir de memórias de tradução.

Neste trabalho apresenta-se todo o processo de construção de corpora paralelo desde os algoritmos de minagem dos dados (data mining) até à construção de vários tipos de recursos bilingues incluindo a construção automática de corpus paralelos pesquisáveis na Internet.

```
}
}

@InProceedings{APL2k2.Synthesis,
  author = "J.J. Almeida and Alberto Manuel Simões",
  title   = "Geração de voz com sotaque",
  booktitle={Actas do XVIII Encontro da Associação Portuguesa de Linguística},
  address={Porto 2002},
  year    = "2003",
  url     = {http://alfarrabio.di.uminho.pt/~albie/publications/APL2k2.Synthesis.pdf},
  lang    = "PT",
  abstract = {
    Como é sabido os sotaques podem estar ligados a uma zona geográfica,
    a um grupo social, podem até ser uma característica pessoal. O seu
    estudo e descrição tem interessado muitos investigadores embora
    normalmente esse estudo tem sido feito de modo pouco formal.
```

No trabalho que aqui se relata, tentou-se descrever formalmente sotaques e disfunções através de criação de regras a integrar como variantes num gerador de voz. Deste modo, pretendeu-se criar um ambiente de experimentação dos modelos construídos para descrever algumas características de certos sotaques ou certas disfunções, de modo a permitir a sua validação.

Constatou-se que se consegue obter certas disfunções e certos sotaques com facilidade por simples acrescento de regras opcionais em certas fases da geração da voz. Outros, aparentam ser de maior dificuldade, ou por não conhecermos suficiente bem os fenómenos neles envolvidos ou envolverem maior complexidade prosódica.

```
},
}
```

```
@InProceedings{xata:xmltdt,
  author = "J.J. Almeida and Alberto Manuel Simões",
  title   = "Engenharia reversa de {HTML} usando tecnologia {XML}",
  booktitle = "{XATA --- XML}, Aplicações e Tecnologias Associadas",
  year      = "2003",
  url      =
{http://alfarrabio.di.uminho.pt/~albie/publications/xata2003xml.pdf},
  lang     = "PT",
  abstract = { O proliferar de ferramentas criadores de HTML e o uso
    de HTML guiado pelo aspecto, tem vindo a arruinar o
    seu lado conceptual. Este problema foi reconhecido e
    deu origem a vários formatos ou tecnologias com o
    objectivo de separar o aspecto do conceito. No
    entanto a realidade actual mostra uma enorme
    quantidade de páginas HTML com péssima leitura
    conceptual e estrutural, invalidando uma série de
    usos possíveis da informação nelas contida. Nesta
    comunicação apresenta-se um trabalho (em fase
    inicial) que pretende fazer engenharia reversa de
    HTML para permitir aumentar a sua acessibilidade, a
    fim de ser usada num \emph{browser} para invisuais.
  },
```



```

    editor ="José Carlos Ramalho",
}

@InProceedings{xata:museudapessoa,
    author = "Alberto Manuel Simões and J.J. Almeida",
    title   = {{H}istórias de {V}ida + {P}rocessamento {E}strutural = {M}useu
da {P}essoa},
    booktitle = "{XATA --- XML}, Aplicações e Tecnologias Associadas",
    year      = "2003",
    lang      = "PT",
    abstract  = {
        Este artigo apresenta a arquitectura actual do Museu da Pessoa,
        contemplando a forma como os documentos estão a ser editorados,
        catalogados, arquivados, e processados para a criação das estruturas
        necessárias ao Museu.
    },
    url = {http://alfarrabio.di.uminho.pt/~albie/publications/xata2003mp.pdf},
    editor ="José Carlos Ramalho",
}

@InProceedings{elpub2003,
    author = "Alberto Manuel Simões and J.J. Almeida",
    year= 2003,
    title= "Music publishing",
    isbn = "972-98921-2-1",
    note = {Guimarães, Jun. 2003},
    publisher = "Universidade do Minho",
    url = {http://alfarrabio.di.uminho.pt/~albie/publications/elpub2003.pdf},
    editor ="Sely Costa et al.",
    booktitle= "ElPub 2003 -- International conference on electronic publishing",
    month = "June",
    pages = "288--298",
    lang="EN",
    abstract = {
        Current music publishing in the Internet is mainly concerned with
        sound publishing. We claim that music publishing is not only to make
        sound available but also to define relations between a set of music
        objects like music scores, guitar chords, lyrics and their
        meta-data. We want an easy way to publish music in the Internet, to
        make high quality paper booklets and even to create Audio CD's.
        In this document we present a workbench for music publishing based
        on open formats, using open-source tools and script programming over
        them. The workbench is based on an archive specification written in
        a text-based format which includes sound references, music scores,
        chords and lyrics and their meta-information.
    },

    keyword ="música, bibliotecas digitais",
}

@InProceedings{cp3a:terminum2003,
    author = "J.J. Almeida and Alberto Simões and José Castro and Bruno
Martins and Paulo Silva",
    year= 2003,
    title= "Projecto {TerminUM}",
    publisher = "Universidade do Minho",
    booktitle={CP3A 2003 -- Workshop em Corpora Paralelos: aplicações e
    algoritmos associados},
    note = {Braga},
    month="Jun.",
    pages = "7--14",

```

```

keyword = "terminum, parallel corpora",
url = {http://alfarrabio.di.uminho.pt/~albie/publications/cp3a2003-terminum.pdf},
abstract = { 0 projecto TerminUM tem como objectivos principais o
    estudo, experimentação e a criação de recursos na
    área dos corpora paralelos, terminologia
    (descritiva) e recursos multilingues ligados a
    corpora: fazer extracção tão automática quanto
    possível de corpora a partir da web; fazer extracção
    de dicionários, de terminologia e de outros recursos
    ligados à tradução; criar e interligar as
    ferramentas desenvolvidas; criar e disponibilizar:
    (1) listas de Bitextos, corpora e corpora paralelos,
    (2) ferramentas de criação e transformação de
    corpora, (3) recursos multilingues derivados/ligados
    a corpora. Nesta apresentação serão abordadas
    algumas tarefas presentemente a decorrer no âmbito
    do projecto, nomeadamente: ciclo de vida da
    construção e transformação de corpora; resumo das
    ferramentas desenvolvidas (e em desenvolvimento);
    construção de corpora paralelos tomando como base
    legendas de filmes (subtitles), ficheiro de
    internacionalização (mensagens de software .po) e
    ficheiros de memórias de tradução (TMX); animação de
    corpora paralelos via web (criação de motores de
    consulta usando diversas ferramentas). }
}

@InProceedings{cp3a:kvec2003,
    author = "Bruno Martins",
    year= 2003,
    title= "{Lingua-Biterm}: um módulo Perl para extracção de terminologia bilingue",
    publisher = "Universidade do Minho",
    booktitle={CP3A 2003 -- Workshop em Corpora Paralelos: aplicações e
        algoritmos associados},
    note = {Braga},
    month="Jun.",
    pages = "65--70",
    keyword = "kvec, terminum, parallel corpora, word alignment",
}

@InProceedings{cp3a:natools2003,
    author = "Alberto Simões",
    year= 2003,
    title= "Alinhamento de corpora paralelos",
    publisher = "Universidade do Minho",
    booktitle={CP3A 2003 -- Workshop em Corpora Paralelos: aplicações e
        algoritmos associados},
    note = {Braga},
    month="Jun.",
    pages = "71--77",
    keyword = "natools, terminum, parallel corpora, word alignment",
}

@article{sepln2003,
    author= {Alberto M. Simões and J.J. Almeida},
    title="{NATools} -- A Statistical Word Aligner Workbench",
    publisher="Sociedade Española para el Procesamiento del Lenguaje Natural",
    journal = {Procesamiento del Lenguaje Natural},
    volume = "31",
    pages = "217--224",
    month="Sep.",

```

```

year= 2003,
abstract = {This document presents the TerminUM project and the
work done in its statistical word aligner workbench (NATools). It
shows a variety of alignment methods for parallel corpora and
discusses the resulting terminological dictionaries and their use:
evaluation of sentence translations; construction of a multi-level
navigation system for linguistic studies or statistical
translations. },
keyword ="natools, terminum, parallel corpora, word alignment",
}

@phdthesis{tesejj,
author = {José João Dias de Almeida},
title = {Dicionários dinâmicos multi-fonte},
school = {Universidade do Minho},
type = "Tese de Doutoramento",
supervisor = "Pedro Rangel Henriques",
url="http://natura.di.uminho.pt/~jj/bib/tesejj.pdf",
month="Dez.",
year = 2003,
lang = "PT",
}

@MastersThesis{teseamsb,
author = {Alberto Manuel Brandão Simões},
title = {Parallel Corpora word alignment and applications},
school = {Escola de Engenharia - Universidade do Minho},
url="http://alfarrabio.di.uminho.pt/~albie/publications/msc.pdf",
type = "Tese de Mestrado",
supervisor = "José João Almeida and Pedro Rangel Henriques",
year = {2004},
lang = "EN",
}

@InProceedings{xata04:tx,
author = {José João Almeida and Alberto Simões},
title = {{TX} --- {V}alidação de {XML} baseada em tipos dinâmicos},
booktitle = {{XATA 2004} - XML, Aplicações e Tecnologias Associadas},
pages = {217--224},
year = {2004},
url =
{http://alfarrabio.di.uminho.pt/~albie/publications/xata04-tx.pdf},
lang = "PT",
isbn = {972-99166-0-8},
editor = {José Carlos Ramalho and Alberto Simões},
month = {February},
abstract = {
Desde o advento do SGML e posteriormente do XML, que a validação de
documentos tem sido focada.

```

Esta validação surgiu para analisar a estrutura dos documentos SGML e XML usando DTDs. Além dessa, e devido às restrições do XML em relação ao SGML, a validação de XML bem formado também tem sido usada. Mais recentemente, os Schema e Schematron vieram permitir a validação a um nível superior: não só a estrutura do documento mas também alguma validação de conteúdo.

Neste artigo apresentamos a ferramenta TX que visa outro nível de validação, em que os tipos possam ser mais ricos e/ou calculados dinamicamente, e onde se possa definir funções de anotação e/ou correcção das porções do documento que não sigam as especificações.

```

    }
}

@InProceedings{xata04:mtd,
  author =      {Alberto Simões and José João Almeida and Xavier Gomez
Guinovart},
  title =       {Memórias de Tradução Distribuídas},
  booktitle =   {{XATA 2004} --- XML, Aplicações e Tecnologias Associadas},
  pages =       {59--68},
  year =        {2004},
  lang          = "PT",
  url = {http://alfarrabio.di.uminho.pt/~albie/publications/xata04-mtd.pdf},
  isbn =        {972-99166-0-8},
  editor =      {José Carlos Ramalho and Alberto Simões},
  month =       {February},
  abstract = {
    Neste documento apresenta-se o conceito de memórias de tradução
    distribuídas, discutindo-se o seu interesse na área da tradução, bem
    como as vantagens que uma ferramenta de tradução pode tirar do seu
    uso.

    É apresentada uma possível implementação de memórias de tradução
    distribuídas usando WebServices numa arquitectura de cooperativismo.
    São definidos as mensagens (API) que um serviço deste género deve
    implementar para que uma ferramenta de tradução possa tirar partido
    da colaboração entre tradutores.
  }
}

@article{xmldt2,
  author = {Alberto Simões},
  title = {{XML::DT} - Down-Translating XML},
  journal = {The Perl Review},
  number = 1,
  volume = 1,
  year = 2004,
}

@article{sepln2004,
  author= {Alberto Simões and Xavier Gómez Guinovart and J.J. Almeida},
  title={Distributed Translation Memories implementation using WebServices},
  publisher="Sociedade Española para el Procesamiento del Lenguaje Natural",
  journal =      {Procesamiento del Lenguaje Natural},
  pages =        {89--94},
  volume =       {33},
  month =        {July},
  year =         {2004},
  lang =         {EN},
  url = {http://alfarrabio.di.uminho.pt/~albie/publications/dtm-sepln.pdf},
  keyword = {TMs, MT, distributed translation memories, WebServices, CAT},
  abstract= { Translation Memories are very useful for translators
    but are difficult to share and reuse in a community of translators.
    This article presents the concept of Distributed Translation
    Memories, where all users can contribute and sharing translations.
    Implementation details using WebServices are shown, as well as an
    example of a distributed system between Portugal and Spain.}
}

@inproceedings{linguateca,

```

```

author = {Diana Santos and Alberto Simões and Ana Frankenberg-Garcia and
  Ana Pinto and Anabela Barreiro and Belinda Maia and Cristina Mota and Débora
  Oliveira and Eckhard Bick and Elisabete Ranchhod and J.J. Almeida
  and Luís Cabral and Luís Costa and Luís Sarmento and Marcirio Chaves and Nuno
  Cardoso and Paulo Rocha and Rachel Aires and Rosário Silva and Rui Vilela and
  Susana Afonso},
title = {Linguateca: um centro de recursos distribuído para o processamento
  computacional da língua portuguesa},
year = 2004,
booktitle = {Workshop on Linguistic Tools and Resources for Spanish and
  Portuguese},
editor = {IBERAMIA 2004},
lang = "EN",
pages="147--154",
address="Puebla, México",
url = {http://alfarrabio.di.uminho.pt/~albie/publications/linguateca.pdf},
abstract = {
  Neste artigo apresentamos uma panorâmica da actividade da Linguateca na criação
  e disponibilização de recursos e ferramentas para a língua portuguesa. Começamos
  por uma descrição dos objectivos e pressupostos da Linguateca e uma breve história
  da sua intervenção, e finalizamos com algumas considerações sobre a melhor forma
  de prosseguir na organização da área.
}
}

```

```

@inproceedings{xata05:fs,
  author="Rui Vilela and Alberto Simões and Eckhard Bick and J.J. Almeida",
  title="Representação em {XML} da {F}loresta {S}intáctica",
  month="Fev.",
  year=2005,
  editor="José Carlos Ramalho and Alberto Simões and João Correia Lopes",
  booktitle="XATA 2005, Aplicações e Tecnologias Associadas",
  publisher="Departamento de Informática, Universidade do Minho",
  location="Braga",
  keyword = {XML, Floresta Sintáctica, tigerXML, Lingua::PT::Dirty},
}

```

```

@inproceedings{xata05:tdt,
  author="J.J. Almeida and Alberto Simões",
  title="Inferência de tipos em documentos {XML}",
  month="Fev.",
  year=2005,
  editor="José Carlos Ramalho and Alberto Simões and João Correia Lopes",
  booktitle="XATA 2005, Aplicações e Tecnologias Associadas",
  publisher="Departamento de Informática, Universidade do Minho",
  location="Braga",
  keyword = {XML, XML::DT},
}

```

```

@inproceedings{xata06:navegante,
  author="J.J. Almeida and Alberto Simões",
  title="Navegante: um proxy de ordem superior para navegação intusiva",
  month="Fev.",
  year=2006,
  editor="José Carlos Ramalho and Alberto Simões and João Correia Lopes",
  booktitle="XATA 2006, Aplicações e Tecnologias Associadas",
  publisher="ESTGP",
  address = {Portalegre},
  Note="poster",
  pages="376--377",
}

```

```

keyword = {XML, XML::DT, HTML},
}

@inproceedings{xata06:xmlauto,
  author="J.J. Almeida and Alberto Simões",
  title =      {Geração dinâmica de {API}s {P}erl para criação de {XML}},
  month="Fev.",
  year=2006,
  editor="José Carlos Ramalho and Alberto Simões and João Correia Lopes",
  booktitle =  {{XATA 2006} --- 4ª Conferência Nacional em XML, Aplicações e Tecnologias Aplicadas},
  publisher="ESTGP",
  url = {http://alfarrabio.di.uminho.pt/~albie/publications/xata2006-xmlwritersimple.pdf},
  address =    {Portalegre},
  pages="307--314",
  keyword = {XML, XML::DT, HTML},
  isbn =      {972-99166-2-4},
  lang =      "PT",
  abstract =   {
    É consensual que o XML como linguagem para a estruturação de documentos
    tem vindo a tomar um lugar relevante. É também evidente a vantagem
    obtida no uso de XML como linguagem de intercâmbio.
    No entanto, a sua sintaxe é
    demasiado descritiva pelo que a geração de documentos de forma
    manual é dolorosa sendo útil dispor de módulos
    que simplifiquem essa tarefa.

    Neste artigo propomos um módulo Perl (XML::Writer::Simple) configurável via
    DTD que simplifica a tarefa de gerar XML.
  },
}

@article{sepln06,
  author =      {Alberto Simões and J. João Almeida},
  title =      {{NatServer:} A Client-Server Architecture for building Parallel Corpora applications},
  year =      {2006},
  journal =    {Procesamiento del Lenguaje Natural},
  address =    {Zaragoza, Spain},
  url = {http://alfarrabio.di.uminho.pt/~albie/publications/sepln06.pdf},
  month =      {September},
  lang =      {EN},
  volume =     {37},
  pages =      {91--97},
  abstract =   {Parallel corpora are important resources for most
    Natural Language processing tasks. From the common
    applications, like machine translation, to the
    usually mono-lingual tasks as paraphrase detection
    and word sense disambiguation, most researchers are
    using massive parallel corpora. Thus, the
    availability of an efficient way to manage them is
    very important. This paper presents a Client-Server
    architecture to query efficiently parallel corpora
    and probabilistic translation dictionaries.},
}

@inProceedings{eamt06,
  author =      {Alberto Simões and J. João Almeida},
  title =      {Combinatory Examples Extraction for Machine Translation},
  shortin =    {{EAMT}},
  year =      {2006},
  url = {http://alfarrabio.di.uminho.pt/~albie/publications/eamt06.pdf},
  booktitle =  {11th Annual Conference of the European Association for Machine Translation},

```

```

editor =      {Jan Tore Lonning and Stephan Oepen},
address =     {Oslo, Norway},
pages =       {27--32},
month =       {19--20, June},
isbn =        {82-7368-294-3},
lang = {EN},
abstract = { One of the bottlenecks of example-based machine
            translation (EBMT) is to be able to amass
            automatically quantities of good examples. In our
            work in EBMT, we are investigating how far one can
            go by performing example extraction from parallel
            corpora using Probabilistic Translation Dictionaries
            to obtain example segmentation points. In fact, the
            success of EBMT highly depends on examples quality
            and quantity, but also in their length. Thus, we
            give special importance on methods to extract
            different size examples from the same translation
            unit. With this article we show that it is possible
            to extract quantities for examples from parallel
            corpora just using probabilistic translation
            dictionaries extracted from the same corpora.},
}

@InProceedings{lrec06,
  author =      {José João Almeida and Alberto Simões},
  title =       {{{T}_20$} --- Recycling Thesauri into a Multilingual Ontology},
  shortin =     {{LREC}},
  booktitle =   {Fifth international conference on Language Resources and Evaluation, LREC 2006},
  year =        {2006},
  url = {http://alfarrabio.di.uminho.pt/~albie/publications/lrec06.pdf},
  address =     {Genova, Italy},
  month =       {May},
  lang =        {EN},
  abstract =    {In this article we present {T}_20$ --- a workbench to
                assist the process of translating heterogeneous
                resources into ontologies, to enrich and add
                multilingual information, to help programming with
                them, and to support ontology publishing. {T}_20$ is
                an ontology algebra.},
}

@inProceedings{elpub06-t2o,
  author =      {J. João Almeida and Alberto Simões },
  title =       {Publishing multilingual ontologies: a quick way of obtaining feedback},
  year =        {2006},
  booktitle =   {{{ElPub 2006} --- Digital Spectrum: Integrating Technology and Culture},
  address =     {Bansko, Bulgaria},
  note="poster",
  url = {http://alfarrabio.di.uminho.pt/~albie/publications/elpub06-t2o.pdf},
  month =       {June},
  lang =        {EN},
  abstract =    {Dictionary and Thesaurus are valuable resources for
                Natural Language Processing but do not exist as
                freely available as expected, especially for
                languages other than English and, when they exist,
                they are just available for querying online. Our
                main goal with T20 --- Thesaurus to Ontology
                framework --- is to create a multilingual ontology:
                freely available online and to download; with a
                computer readable format; with a good API; with a

```

```

        structure as rich as possible; reusing all the
        structured information we can get; },
    }

@InProceedings{elpub06-blind,
    author = {António R. Fernandes and Alexandre Carvalho and J. João Almeida and Alberto Simões },
    title = {Transcoding for Web Accessibility for the Blind: Semantics
from Structure},
    shortin = {{ElPub}},
    booktitle = {ElPub 2006 --- Digital Spectrum: Integrating Technology and Culture},
    year = {2006},
    url = {http://alfarrabio.di.uminho.pt/~albie/publications/elpub06-blind.pdf},
    address = {Bansko, Bulgaria},
    month = {June},
    abstract = {True accessibility requires minimizing the scanning time
        to find a particular piece of
        information. Sequentially reading web pages do not
        provide this type of accessibility, for instance
        before the user gets to the actual text content of
        the page it has to go through a lot of menus and
        headers. However if the user could navigate a web
        page based through semantically classified blocks
        then the user could jump faster to the actual
        content of the page, skipping all the menus and
        other parts of the page. We propose a transcoding
        engine that tackles accessibility at two distinct,
        yet complementary, levels: for specific known sites
        and general unknown sites. We present a tool for
        building customized scripts for known sites that
        turns this process in an extremely simple task,
        which can be performed by anyone, without any
        expertise. For general unknown sites, our approach
        relies on statistical analysis of the structural
        blocks that define a web page to infer a semantic
        for the block.},

    lang = "EN",
}

% ===== 2007 =====

@InCollection{avalon:jspell,
    author = {José João Almeida and Alberto Simões},
    title = {Jspellando nas morfolimpiadas: Sobre a participação do {Jspell} nas Morfolimpiadas},
    booktitle = {Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa},
    shortin = {Avaliação conjunta, cap. 8},
    year = {2007},
    editor = {Diana Santos},
    pages = {83--90},
    publisher = {{IST Press}},
}

@InCollection{avalon:avalinha,
    author = {Alberto Simões and José João Almeida},
    title = {Avaliação de alinhadores},
    shortin = {Avaliação conjunta, cap. 18},
    booktitle = {Avaliação conjunta: um novo paradigma no processamento computacional da língua portuguesa},
    publisher = {{IST Press}},
    year = {2007},
}

```



```

pages =      {219--230},
editor =     {Diana Santos},
}

@InProceedings{xata07:xmltmx,
  author =    {José João Almeida and Alberto Simões},
  title =     {{XML::TMX} --- Processamento de Memórias de Tradução de Grandes Dimensões},
  shortin =   {{XATA}},
  booktitle = {{XATA 2007} --- 5ª Conferência Nacional em XML, Aplicações e Tecnologias Aplicadas},
  year =      {2007},
  month =     {February},
  url =       {http://alfarrabio.di.uminho.pt/~albie/publications/xmltmx07.pdf},
  pages =     {83--93},
  isbn =      {978-972-99166-4-9},
  editor =    {José Carlos Ramalho and João Correia Lopes and Luís Carríço},
  abstract =   { As ferramentas de tradução assistida por computador
                tentam reutilizar as traduções realizadas pelo
                tradutor sempre que uma frase semelhante tenha sido
                já traduzida. Para o intercâmbio destes documentos
                foi definido um formato denominado TMX (Translation
                Memory Exchange) baseado em XML. Este tipo de
                documento ganha facilmente tamanhos incontroláveis
                para o seu processamento com métodos tradicionais.
                Neste artigo propomos uma metodologia de ordem
                superior para o processamento de documentos de
                estrutura repetitiva (em que se inserem as memórias
                de tradução) com uma abordagem baseada na conjunção
                de SAX e DOM. São apresentados vários exemplos de
                filtros sobre memórias de tradução bem como um
                conjunto de medidas da sua eficiência. }
}

@InProceedings{MP07,
  author =    {Alberto Simões and Rúben Fonseca and José João Almeida},
  title =     {{Makefile::Parallel} Dependency Specification Language},
  booktitle = {Euro-Par 2007},
  year =      {2007},
  address =   {Rennes, France},
  month =     {August},
  pages =     {33--41},
  editor =    {Anne-Marie Kermarrec and Luc Bougé and Thierry Priol},
  volume =    {4641},
  series =    {LNCS},
  publisher = {Springer-Verlag},
  abstract =   { Some processes are not easy to be programmed from scratch
for
parallel machines (clusters), but can be easily split on simple
steps. Makefile::Parallel is a tool which lets users to specify how processes
depend on each other.

The language syntax resembles the well known Makefile
makefiles format, but instead of specifying files or targets
dependencies, Makefile::Parallel specifies processes (or jobs) dependencies.

The scheduler submits jobs to the cluster scheduler (in our case,
Rocks PBS) waiting them to end. When each process finishes,
dependencies are calculated and direct dependent jobs are submitted.

Makefile::Parallel language includes features to specify parametric rules,
used
to split and join processes dependencies. Some tasks can be split

```

```

    into n smaller jobs working on different portions of files. At the
    end, another process can be used to join the results.
},
}

@InProceedings{epia-bio-2007,
  booktitle = {New Trends in Artificial Intelligence},
  editor = {José Neves and Manuel Filipe Santos and José Manuel Machado},
  pages = {541--552},
  isbn13 = {978-989-95618-0-9},
  author = {Anália Lourenço and Alberto Simões and José Almeida and
Miguel Rocha and Isabel Rocha and Eugénio Ferreira},
  title = {An Ontology-Based Approach To Systems Biology Literature
Retrieval and Processing},
  shortin = {Epia, CMBSB},
  year = {2007},
  month = {December},
  abstract = {This paper details the \emph{SysBio Explorer}, a
Systems Biology Literature Retrieval and Processing
Framework, whose aim relies on the automatic
inference of regulatory and metabolic networks based
on biomedical literature. The \emph{SysBio Explorer}
does not focus on any organism or problem in
particular and encompasses a number of processing
and analysis techniques. It works over full-text
documents, applying Natural Language Processing
techniques and using biomedical dictionaries and
ontologies together with hand-made rules. Besides
biological entity recognition and relation
extraction, document classification, relevance
assessment and authoring networks are also within
its present scope. The framework is described in
terms of its design requirements and implementation
decisions, exposing current achievements, but also
highlighting present obstacles and future
work. Experiments over real-world problems
concerning the organisms \emph{E. coli},
\emph{S. cerevisiae} and \emph{H. pylori} are used
in its validation.}
}

@InProceedings{epia-music-2007,
  booktitle = {New Trends in Artificial Intelligence},
  editor = {José Neves and Manuel Filipe Santos and José Manuel Machado},
  shortin = {Epia, TEMA},
  pages = {791--799},
  author = {Alberto Simões and Anália Lourenço and José João Almeida},
  title = {Using Text Mining Techniques for Classical Music Scores
Analysis},
  year = {2007},
  month = {December},
  abstract = {Abstract.Music Classification is a particular area
of Computational Musicology that provides valuable
insights about the evolving of composition patterns
and assists in catalogue generation. The proposed work
detaches from former works by classifying music based
on music score information. Text Mining techniques
support music score processing while Classification
techniques are used in the construction of decision
models. Although research is still at its earliest

```

beginnings, the work already provides valuable contributes to symbolic music representation processing and subsequent analysis. Score processing involved the counting of ascending and descending chromatic intervals, note duration and meta-information tagging. Analysis involved feature selection and the evaluation of several data mining algorithms, ensuring extensibility towards larger repositories or more complex problems. Experiments report the analysis of composition epochs on a subset of the Mutopia project open archive of classical LilyPond-annotated music scores.

```
    },
}
```

```
@InCollection{harem:rena,
  author =      {J.João Almeida},
  title =       {{RENA} - Reconhecedor de Entidades},
  booktitle =   {Reconhecimento de entidades mencionadas em português},
  year =        {2007},
  note = {Documentação e actas do HAREM, a primeira avaliação conjunta na
  área},
  pages =       {157-172},
  url = {http://acdc.linguatca.pt/aval_conjunta/LivroHAREM/Cap13-SantosCardoso2007-Almeida.pdf},
  ISBN = {978-989-20-0731-1},
  editor =      {Diana Santos and Nuno Cardoso},
  shortin ={{HAREM} cap. XIII},
}
```

```
@Book{RH02,
  author=   {José Carlos Ramalho and Pedro Rangel Henriques},
  title =   {XML \& XSL: da teoria à prática},
  publisher =   {Editora FCA},
  series =     {Série Tecnologias de Informação ISBN-972-722-347-8},
  edition=     {1st Ed.},
  month =      {Oct},
  year =       {2002}
}
```

```
@Incollection{RRH02,
  author = {F\'atima Rodrigues and Carlos Ramos and Pedro Rangel Henriques},
  title = {Data Reduction to Improve Knowledge Extraction},
  booktitle = {Enterprise Information Systems III},
  editor = {J.Filipe and B.Sharp },
  publisher = {Kluwer Academic Publishers},
  pages = {137-145},
  year = 2002
}
```

```
@Incollection{ORH06,
  author = {Paulo Oliveira and F\'atima Rodrigues and Pedro Rangel Henriques},
  title = {Data Cleaning by Reusing Domain Knowledge},
  booktitle = {Knowledge and Decision Technologies},
  editor = {},
  publisher = {ISBN: 972-8688-39-3},
  pages = {67-74},
  year = 2006
}
```

%-----

```
@Article{KMHVZ04,
  author = {Tomaz Kosar and Marjan Mernik and Pedro Henriques and Maria João Varanda and Viljem Zumer},
  title = {Grammatical Approach to Problem Solving},
  journal = {Informatica Journal},
  volume = {28},
  pages = {393-404},
  year = {2004}
}

@Article{HVMLGW05,
  author = {Pedro Henriques and Maria João Varanda and Marjan Mernik and Mitja Lenic and Jeff Gray and Hui},
  title = {Automatic Generation of Language-based Tools using LISA System},
  journal = {IEEE Software Journal},
  year = {2005},
  month = {April},
  volume = {152},
  number = {2},
  pages = {54-70}
}

@Article{RMHV06,
  author = {Damijan Rebernak and Marjan Mernik and Pedro Rangel Henriques and Maria João Varanda},
  title = {AspectLISA: an aspect-oriented compiler construction system based on attribute grammars},
  journal = {Electronic Notes in Theoretical Computer Science (ENTCS)},
  editor = {J. Boyland and A. Sloane},
  publisher = {Elsevier},
  year = {2006},
  month = {Oct},
  volume = {164},
  number = {2},
  pages = {37-53}
}

@Article{RMHCV06,
  author = {Damijan Rebernak and Marjan Mernik and Pedro Rangel Henriques and Daniela da Cruz and Maria Jo},
  title = {Specifying Languages using aspect-oriented approach: AspectLISA},
  journal = {CIT -- Journal of Computing and Information Technology},
  editor = {},
  publisher = {},
  year = {2006},
  month = {Dec},
  volume = {14},
  number = {4},
  pages = {343-350}
}

@Article{GDH06,
  author = "Gancarski, Alda and Doucet, Anne and Henriques, Pedro",
  title = "{AG-based interactive system to retrieve information from XML documents}",
  journal = "{IEEE Software Journal}",
  year = 2006,
  month = {Apr},
  volume = {153},
  pages = "51-60"
}

@Article{BH98,
  author = {Leonor Barroca and Pedro Rangel Henriques},
  title = {A Framework and Patterns for the Specification of Reactive Systems},
```

```

journal = {IST: Information and Software Technology},
editor = {},
publisher = {Elsevier},
year = {1998},
month = {},
volume = {},
number = {40},
pages = {135-142}
}

```

```

@Article{RAH98,
author = {José Carlos Ramalho and José João Almeida and Pedro Rangel Henriques},
title = {Algebraic Specification of Documents},
journal = {TSC: Theoretical Computer Science},
editor = {},
publisher = {Elsevier},
year = {1998},
month = {Jun},
volume = {199},
number = {},
pages = {}
}

```

```

@Article{RARH98,
author = {José Carlos Ramalho and José João Almeida and Jorge Gustavo Rocha and Pedro Rangel Henriques},
title = {SGML Documents: Where does quality go?},
journal = {Mark-up Languages: theory and practice},
editor = {},
publisher = {MIT Press},
year = {1998},
month = {Nov},
volume = {1},
number = {1},
pages = {}
}

```

```

@Article{GRH06,
author = {Giovani Librelotto and José Carlos Ramalho and Pedro Rangel Henriques},
title = {Metamorphosis - A Topic Maps Based Environment to Handle Heterogeneous Information Resources},
journal = {Lecture Notes in Computer Science (LNCS)},
editor = {},
publisher = {Springer-Verlag GmbH},
year = {2006},
month = {},
volume = {3873},
number = {},
pages = {14-25}
}

```

```

@Article{JGRH04,
author = {Marta Jacinto and Giovani Librelotto and José Carlos Ramalho and Pedro Rangel Henriques},
title = {XCSL Tutorial},
journal = {Acm Crossroads},
editor = {},
publisher = {ACM Press},
year = {2004},
month = {},
volume = {8},
number = {5},
pages = {5}
}

```

```

@Article{JGRH03,
  author = {Marta Jacinto and Giovani Librelotto and José Carlos Ramalho and Pedro Rangel Henriques},
  title = {XCSL: XML Constraint Specification Language},
  journal = {CLEI Eletronic Journal},
  editor = {},
  publisher = {ACM Press},
  year = {2003},
  month = {},
  volume = {6},
  number = {},
  pages = {3}
}

@Article{GRH04,
  author = {Giovani Librelotto and José Carlos Ramalho and Pedro Rangel Henriques},
  title = {TM-Builder: An Ontology Builder based on XML Topic Maps},
  journal = {CLEI Eletronic Journal},
  editor = {},
  publisher = {ACM Press},
  year = {2004},
  month = {},
  volume = {7},
  number = {2},
  pages = {4}
}

@Article{GRH05a,
  author = {Giovani Librelotto and José Carlos Ramalho and Pedro Rangel Henriques},
  title = {Geração automática de interfaces Web para Sistemas de Informação: Metamorphosis},
  journal = {Sistemas de Informação},
  editor = {},
  publisher = {APSI: Associação Portuguesa de Sistemas de Informação},
  year = {2005},
  month = {},
  volume = {17},
  number = {},
  pages = {53-72}
}

@Article{RH98a,
  author = {José Carlos Ramalho and Pedro Rangel Henriques},
  title = {Qualidade na Publicação Electrónica: como controlá-la?},
  journal = {Sistemas de Informação},
  editor = {},
  publisher = {APSI: Associação Portuguesa de Sistemas de Informação},
  year = {1998},
  month = {},
  volume = {9},
  number = {},
  pages = {69-83}
}

@Article{MSH05,
  author = {Giovana Mendes and Nuno Alberto Silva and Pedro Rangel Henriques},
  title = {Utilizando uma Base de Dados XML Nativa aplicada ao tratamento de erros num sistema de logs},
  journal = {Sistemas de Informação},
  editor = {},
  publisher = {APSI: Associação Portuguesa de Sistemas de Informação},
  year = {2005},
  month = {},

```

```

    volume = {17},
    number = {},
    pages = {91-100}
}

```

```

@Article{ALHF02,
  author = {Gustavo Arnold and Giovani Librelotto and Pedro Rangel Henriques and Jaime Fonseca},
  title = {O Uso da Linguagem RS em Robótica},
  journal = {Revista Electrónica e Telecomunicações},
  editor = {},
  publisher = {Departamento de Electrónica e Telecomunicações da Universidade de Aveiro},
  year = {2002},
  month = {Apr},
  volume = {},
  number = {},
  pages = {501-508}
}

```

%-----

```

@InProceedings{HKMVZ03,
  author = {Pedro Henriques and Tomaz Kosar and Marjan Mernik and Maria João Varanda and Viljem Zumer},
  title = {Grammatical Approach to Problem Solving},
  booktitle = {ITI'03 - 25th International Conference on Information Technology Interfaces, Croacia},
  publisher = {IEEE},
  year = {2003},
  month = {June}
}

```

```

@InProceedings{VH01,
  author = {M. J. Varanda and Pedro Rangel Henriques},
  title = {Visualization / Animation of Programs based on Abstract Representations and Formal Mappings},
  booktitle = {HCC'01 - 2001 IEEE Symposia on Human-Centric Computing Languages and Environments},
  publisher = {IEEE},
  year = {2001},
  month = {September}
}

```

```

@InProceedings{VH02,
  author = {P. Henriques and M. J. Varanda and M. Mernik and M. Lenic},
  title = {Automatic Generation of Language-based Tools},
  booktitle = {LDTA - Workshop on Language, Descriptions, Tools and Applications (ETAPS'02)},
  year = {2002},
  month = {April}
}

```

```

@InProceedings{VH03,
  author = {M. J. Varanda and P. Henriques},
  title = {Visualization / Animation of Programs in Alma: obtaining different results},
  booktitle = {VMSE2003 - Symposium on Visual and Multimedia Software Engineering (HCC'03), New Zealand},
  publisher = {IEEE},
  year = {2003},
  editor = {},
  month = {October}
}

```

```

@InProceedings{RMHVC06,
  author = {Damijan Rebernak and Marjan Mernik and Pedro Rangel Henriques and Maria João Varanda Pereira},
  title = {Specifying Languages using Aspect-oriented Approach: AspectLISA},
  booktitle = {ITI'06 - 28th International Conference on Information Technology Interfaces, Croacia},

```

```

publisher = {IEEE},
pages      = {},
year       = {2006},
month      = {June}
}

```

```

@InProceedings{BHVU07d,
  author = {Mario Ber\`on and Pedro Rangel Henriques and M. Jo\~ao Varanda and Roberto Uzal},
  title = {PICS una Herramienta para la Comprensi3n e Inspecci3n de Programas},
  booktitle = {CACIC'07 - XIII Argentine Congress on Computer Science},
  publisher = {Corrientes, Argentina},
  year = {2007},
  editor = {},
  month = {Aug}
}

```

```

@InProceedings{BHVU07c,
  author = {Mario Ber\`on and Pedro Rangel Henriques and M. Jo\~ao Varanda and Roberto Uzal},
  title = {Program Inspection to Incerconnect Behavioral and Operational View for Program Comprehension},
  booktitle = {York Doctoral Symposium, 2007},
  publisher = {University of York, UK},
  year = {2007},
  editor = {},
  month = {Oct}
}

```

```

@InProceedings{BHVU07b,
  author = {Mario Ber\`on and Pedro Rangel Henriques and M. Jo\~ao Varanda and Roberto Uzal},
  title = {Comprensi3n de Programas por Inspecci3n Visual y Animaci3n},
  booktitle = {IX Workshop de Investigadores en Ciencias de la Computaci3n (WICC 2007), Argentina},
  year = {2007},
  editor = {},
  month = {}
}

```

```

@InProceedings{BHVU07a,
  author = {Mario Ber\`on and Pedro Rangel Henriques and M. Jo\~ao Varanda and Roberto Uzal},
  title = {Static and Dynamic Strategies to Understand C Programs by Code Annotation},
  booktitle = {OpenCert'07, 1st Int. Workshop on Foundations and Techniques for Open Source Software Certification},
  year = {2007},
  editor = {},
  month = {Apr}
}

```

```

@InProceedings{CHLB07a,
  author = {Elisabete Cunha and Pedro Rangel Henriques and Sandra Lopes and Mario B3ron},
  title = {O Sitio de Pico, Software Educativo para Crian7as com Paralisia Cerebral},
  booktitle = {IX Workshop de Investigadores en Ciencias de la Computaci3n (WICC 2007), Argentina},
  year = {2007},
  editor = {},
  month = {}
}

```

```

@InProceedings{BHVU06a,
  author = {Mario B3ron and Pedro Rangel Henriques and Maria Jo3o Varanda and Roberto Uzal},
  title = {Herramientas para la compresi3n de programas},
  booktitle = {VIII Workshop de Investigadores en Ciencias de la Computaci3n (WICC 2006), Universidade de S3o Carlos},
  year = {2006},
  editor = {},
  month = {}
}

```



```

@InProceedings{BHVU06b,
  author = {Mario Béron and Pedro Rangel Henriques and Maria João Varanda and Roberto Uzal},
  title = {Comprensión de Algoritmos de Ruteo},
  booktitle = {XXXII Conferencia Latinoamericana de Informática (CLEI 2006), Santiago de Chile},
  year = {2006},
  editor = {},
  month = {}
}

@InProceedings{BHVUM06,
  author = {Mario Béron and Pedro Rangel Henriques and Maria João Varanda and Roberto Uzal and G. Mon},
  title = {A Language Processing Tool for Program Comprehension},
  booktitle = {CACIC'06 - XII Argentine Congress on Computer Science},
  publisher = {Universidad Nacional de San Luis, Argentina},
  year = {2006},
  month = {Oct}
}

@InProceedings{BHVU08,
  author = {Mario Béron and Pedro Rangel Henriques and Maria João Varanda and Roberto Uzal},
  title = {Simplificando la Comprensión de Programas a través de la Interconexión de Dominios},
  booktitle = {CACIC'08 - XIV Argentine Congress on Computer Science, Universidad Nacional de San Luis, A},
  publisher = {Universidad Nacional de Chilecito, La Rioja, Argentina},
  year = {2008},
  month = {Oct}
}

@InProceedings{BHV06,
  author = {Mario Béron and Pedro Rangel Henriques and Maria João Varanda},
  title = {A System for Evaluate and Understand Routing Algorithms},
  booktitle = {Interacção'06, Universidade do Minho},
  year = {2006},
  editor = {},
  month = {}
}

@InProceedings{BCVHU08,
  author = {Mario Béron and Daniela da Cruz and Maria João Varanda and Pedro Rangel Henriques and Robe},
  title = {Evaluation Criteria of Software Visualization Systems used for Program Comprehension},
  booktitle = {Interacção'08 -- 3ª Conferência Interacção Pessoa-Máquina},
  editor = {Universidade de Évora},
  year = {2008},
  month = {Oct}
}

@InProceedings{BUHV08,
  author = {Mario Berón and Roberto Uzal and Pedro Rangel Henriques and Maria João Varanda Pereira},
  title = {Inspección de Código para relacionar los Dominios del Problema y Programa para la Comprensi},
  booktitle = {X Workshop de Investigadores en Ciencias de la Computación (WICC 2008), La Pampa, Argentina},
  publisher = {Facultad de Ingeniería de la Universidad Nacional de la Pampa},
  year = {2008},
  editor = {UNLPam, RedUNCI},
  month = {May}
}

@InProceedings{OVH05,
  author = {Eva Oliveira and Maria João Varanda and Pedro Rangel Henriques},
  title = {Compreensão de Aplicações Web: O Processo e as Ferramentas},
  booktitle = {CAPSI05 - Conferência da Associação Portuguesa de Sistemas de Informação, ESTiG Bragança},
  year = {2005},

```

```

    month =      {October}
}

%> Proposta de um sistema para compreensão de aplicações web
@InProceedings{OHV06,
  author =      {Eva Oliveira and Pedro Rangel Henriques and Maria João Varanda },
  title =       {Proposta de um Sistema para Compreensão de Aplicações Web},
  booktitle =   {CAPSI06 - Conferência da Associação Portuguesa de Sistemas de Informação, Univ. Aveiro},
  year =        {2007},
  month =       {January}
}

%>Analyzing the structure of scientific articles to improve information retrieval
@InProceedings{GH07b,
  author =      {Alda Gancarski and Pedro Rangel Henriques},
  title =       {Analyzing the structure of scientific articles to improve information retrieval},
  booktitle =   {Proceedings of the Int. IADIS Conference on Applied Computing 2007 (IADIS AC07 - Salamanca, Spain)},
  year =        {2007},
  editor =      {Pedro Isaias and Nuno Mamede},
  month =       {Feb}
}

%>Using data together with metadata to improve xml information access
@InProceedings{GH07a,
  author =      {Alda Gancarski and Pedro Rangel Henriques},
  title =       {Using data together with metadata to improve XML information access},
  booktitle =   {Proceedings of the Int. Conference on Web Information Systems and Technologies (WEBIST07 - Porto)},
  year =        {2007},
  editor =      {},
  month =       {Mar}
}

@InProceedings{GFH08,
  author =      {Alda Gancarski and Flávio Xavier Ferreira and Pedro Rangel Henriques},
  title =       {Using data together with metadata to improve XML information access},
  booktitle =   {Proceedings of the 10th Int. Conference on Enterprise Information Systems (ICEIS'08)},
  address =     {Barcelona, Spain},
  year =        {2008},
  month =       {Jun}
}

@inproceedings{FGH08,
  author =      {Flávio Xavier Ferreira and Alda Lopes Gancarski and Pedro Rangel Henriques},
  title =       {Information access from XML using semantics and context: application to the Portuguese Emigration Museum resources},
  booktitle =   {CAPSI'08 -- 8ª Conferência da Associação Portuguesa de Sistemas de Informação},
  address =     {Setubal, Portugal},
  year =        {2008},
  month =       {Oct},
}

@InProceedings{FH08,
  author =      {Flávio Xavier Ferreira and Pedro Rangel Henriques},
  title =       {Using OWL to specify and build different views over the Emigration Museum resources},
  booktitle =   {XATA08 --- {XML} : Aplicações e Tecnologias Associadas, Évora, Portugal},
  year =        {2008},
  month =       {February}
}

%> Navegando na rede semântica dos topic maps com o ulisses
@inproceedings{LPRH07,
  author={Giovani Rubert Librelotto

```

```

    and Renato Preigschadt de Azevedo
    and José Carlos Ramalho
    and Pedro Rangel Henriques},
title={Navegando na Rede Semântica dos Topic Maps com o Ulisses},
booktitle={XATA 2007 - XML: Aplicações e Tecnologias Associadas (5ª Conferência Nacional)},
year={2007},
address={Lisboa - Portugal},
month={Feb},
}

@inproceedings{LPRH07-TM,
author={Giovani Rubert Librelotto
    and Renato Preigschadt de Azevedo
    and José Carlos Ramalho
    and Pedro Rangel Henriques},
title={Topic Maps Constraint Specification Languages: comparing AsTMA!, OSL, and XTche},
booktitle={Topic Maps 2007 - The First International Topic Maps User Conference},
year={2007},
address={Oslo - Norway},
month={Mar},
}

@inproceedings{LRHGT08,
author={Giovani Rubert Librelotto
    and José Carlos Ramalho
    and Pedro Rangel Henriques
    and Jonas Bulegon Gassen
    and Rogério Corrêa Turchetti},
title={A Framework to specify, extract and manage Topic Maps driven by ontologie},
booktitle={SIGDPC'08 -- 26th ACM International Conference on Design of Communication},
address={Lisboa, Portugal},
year={2008},
month={Sep},
}

@inproceedings{LGFSSAH08,
author={Giovani Rubert Librelotto
    and Jonas Bulegon Gassen
    and Leandro O. Freitas
    and Matheus C. Silveira
    and Fábio L. Silva
    and Iara Augustin
    and Pedro Rangel Henriques},
title={Uma Ontologia aplicada a um Ambiente Pervasivo Hospitalar},
booktitle={CAPSI'08 -- 8ª Conferência da Associação Portuguesa de Sistemas de Informação},
address={Setubal, Portugal},
year={2008},
month={Oct},
}

@InProceedings{LMMVRH08,
author={Giovani Rubert Librelotto
    and Mirkos Martins
    and Henrique Machado
    and Juliana Vizzotto
    and José Carlos Ramalho
    and Pedro Rangel Henriques},
title = {Generating a Semantic Network for PubMed},
booktitle = {XATA08 --- {XML} : Aplicações e Tecnologias Associadas, Évora, Portugal},
year = {2008},
month = {February}

```

```

}
%.-----2008.9.18
> LIBRELOTTO, Giovanni R. ; AZEVEDO, R. P. ; GASSEN, Jonas ; RAMALHO, José
> Carlos ; HENRIQUES, Pedro Rangel . Visualização de Topic Maps com o UliSSES.
> Revista do CCEI, v. 12, p. 61-67, 2008.
>
> LIBRELOTTO, Giovanni R. ; AZEVEDO, R. P. ; RAMALHO, José Carlos ; HENRIQUES,
> Pedro Rangel . Comparing Topic Maps Constraint Specification Languages.
> Lecture Notes in Computer Science, v. 4999, p. 86-97, 2008.
>
> LIBRELOTTO, Giovanni R. ; RAMALHO, José Carlos ; HENRIQUES, Pedro Rangel .
> Representação de Conhecimento na Semantic Web. In: Antônio A. F. Loureiro;
> Marinho P. Barcellos. (Org.). XXV Congresso da Sociedade Brasileira de
> Computação: XXIV Jornadas de Atualização em Informática. : , 2005, v. , p.
> 1210-1261.
>
> LIBRELOTTO, Giovanni R. ; GASSEN, Jonas ; RAMALHO, José Carlos ; HENRIQUES,
> Pedro Rangel . Using Web Services to put Metamorphosis on the Web. In:
> XXXIII Conferencia Latinoamericana de Informática, 2007, San José, Costa
> Rica. Anais da XXXIII Conferencia Latinoamericana de Informática, 2007.
>
> LIBRELOTTO, Giovanni R. ; AZEVEDO, R. P. ; RAMALHO, José Carlos ; HENRIQUES,
> Pedro Rangel . Topic Maps Constraint Specification Languages: comparing
> AsTma!, OSL, and XTche. In: The First International Topic Maps User
> Conference, 2007, Oslo, Norway. The First International Topic Maps User
> Conference, 2007.
>
> LIBRELOTTO, Giovanni R. ; AZEVEDO, R. P. ; TURCHETTI, Rogério Correa ;
> RAMALHO, José Carlos ; HENRIQUES, Pedro Rangel . Implementing TMCL: XTche -
> a topic map schema and constraint specification language. In: XXXIII
> Conferencia Latinoamericana de Informática, 2007, San José, Costa Rica.
> Anais da XXXIII Conferencia Latinoamericana de Informática, 2007.
>
> LIBRELOTTO, Giovanni R. ; FERREIRA, P. G. D. ; MARTINS, M. O. ; MACHADO, H.
> T. ; RAMALHO, José Carlos ; HENRIQUES, Pedro Rangel . Topic Maps applied to
> PubMed. In: Extreme Markup Languages 2007, 2007, Montreal, Canadá. Extreme
> Proceedings 2007, 2007.
>
> LIBRELOTTO, Giovanni R. ; AZEVEDO, R. P. ; RAMALHO, José Carlos ; HENRIQUES,
> Pedro Rangel . Topic Maps Constraint Languages: understanding and comparing.
> In: Portuguese Conference on Artificial Intelligence, 2007, Guimarães,
> Portugal. New Trends in Artificial Intelligence. José Manuel Neves, Manuel
> Filipe Santos e José Manuel Machado (Eds.), 2007. v. 13.
>
%-----2008.09.18

%-----
@Article{CHV08ja,
  author = {Daniela da Cruz and Pedro Rangel Henriques and Maria João Varanda},
  title = {Alma versus DDD},
  journal = {ComSIS -- Computer Science and Information Systems Journal, Special Issue on Compilers, Related Languages and Tools},
  editor = {Ivan Lukovic and Simao Melo de Sousa },
  publisher = {Faculty of Technical Sciences, Trg Dositeja Obradovica 6, 21000 Novi Sad, Serbia},
  year = {2008},
  month = {Dec},
  volume = {5},
  number = {2},
  pages = {119-136},
  note = {ISSN: 1820-0214}
}

```

```

@Article{FPCH08jb,
  author = {Elisabete Ferreira and Rogério Paulo and Daniela da Cruz and Pedro Rangel Henriques},
  title = {Integration of the {ST} Language in a Model-Based Engineering Environment for Control Systems},
  journal = {ComSIS -- Computer Science and Information Systems Journal, Special Issue on Compilers, Related Languages and Tools},
  editor = {Ivan Lukovic and Simao Melo de Sousa },
  publisher = {Faculty of Technical Sciences, Trg Dositeja Obradovica 6, 21000 Novi Sad, Serbia},
  year = {2008},
  month = {Dec},
  volume = {5},
  number = {2},
  pages = {87-101},
  note = {ISSN: 1820-0214}
}

@Article{PMCH08j,
  author = {Maria João Varanda Pereira and Marjan Mernik and Daniela da Cruz and Pedro Rangel Henriques},
  title = {Program Comprehension for Domain-Specific Languages (invited paper)},
  journal = {ComSIS -- Computer Science and Information Systems Journal, Special Issue on Compilers, Related Languages and Tools},
  editor = {Ivan Lukovic and Simao Melo de Sousa },
  publisher = {Faculty of Technical Sciences, Trg Dositeja Obradovica 6, 21000 Novi Sad, Serbia},
  year = {2008},
  month = {Dec},
  volume = {5},
  number = {2},
  pages = {1-17},
  note = {ISSN: 1820-0214}
}

@Article{CHV07,
  author = {Daniela da Cruz and Pedro Rangel Henriques and Maria João Varanda},
  title = {Constructing program animations using a pattern-based approach},
  journal = {ComSIS -- Computer Science and Information Systems Journal, Special Issue on Advances in Programming Languages},
  editor = {Ivan Lukovic and Marjan Mernik },
  publisher = {Faculty of Technical Sciences, Trg Dositeja Obradovica 6, 21000 Novi Sad, Serbia},
  year = {2007},
  month = {Dec},
  volume = {4},
  number = {2},
  pages = {97-114},
  note = {ISSN: 1820-0214}
}

@InProceedings{CPH07f,
  author = {Daniela da Cruz and Maria João Varanda Pereira and Pedro Rangel Henriques},
  title = {Pattern-based Program Visualization},
  booktitle = {1st Workshop on Advances in Programming Languages, WAPL'07 --- evento satélite da
    \textsf{International Multiconference on Computer Science and Information Technology, IMCSIT'07}},
  year = {2007},
  editor = {},
  month = {Oct},
}

@InProceedings{CH07g,
  author = {Daniela da Cruz and Pedro Rangel Henriques},
  title = {Slicing wxHaskell modules to derive the User Interface Abstract Model (short paper and poster)},
  booktitle = {WAPL'07 ---1st Workshop on Applications of Programming Languages,
    a satellite event of \textsf{International Multiconference on Computer Science and Information Technology, IMCSIT'07}},
  year = {2007},
  month = {October}
}

@InProceedings{CH07h,

```



```

author = {Daniela da Cruz and Pedro Rangel Henriques and Maria João Varanda Pereira},
title = {Exploring and Visualizing the "alma" of {XML} Documents},
booktitle = {XATA08 --- {XML} : Aplicações e Tecnologias Associadas, Évora, Portugal},
year = {2008},
month = {February}
}

@InProceedings{CHP08b,
author = {Daniela da Cruz and Pedro Rangel Henriques and Maria João Varanda Pereira},
title = {Alma vs {DDD}},
booktitle = {CoRTA08 --- Compilers, Related Technologies and Applications, Bragança, Portugal},
year = {2008},
month = {July}
}

@InProceedings{CPH08c,
author = {Daniela da Cruz and Jorge Sousa Pinto and Pedro Rangel Henriques},
title = {Properties Preservation during Transformation (short paper)},
booktitle = {CoRTA08 --- Compilers, Related Technologies and Applications, Bragança, Portugal},
year = {2008},
month = {July},
annote = {Repository, ID: http://hdl.handle.net/1822/8106}
}

@InProceedings{FPCH08d,
author = {Elisabete Ferreira and Rogério Paulo and Daniela da Cruz and Pedro Rangel Henriques},
title = {Integration of the {ST} Language in a Model-Based Engineering Environment for Control Systems:
An Approach for Compiler Implementation},
booktitle = {CoRTA08 --- Compilers, Related Technologies and Applications, Bragança, Portugal},
year = {2008},
month = {July}
}

@InProceedings{PMCH08e,
author = {Maria João Varanda Pereira and Marjan Mernik and Daniela da Cruz and Pedro Rangel Henriques},
title = {Visual{LISA}: a Visual Interface for an Attribute Grammar based Compiler-Compiler (short paper)},
booktitle = {CoRTA08 --- Compilers, Related Technologies and Applications, Bragança, Portugal},
year = {2008},
month = {July}
}

@InProceedings{PMCH08f,
author = {Maria João Varanda Pereira and Marjan Mernik and Daniela da Cruz and Pedro Rangel Henriques},
title = {Program Comprehension for Domain-Specific Languages},
booktitle = {CoRTA08 --- Compilers, Related Technologies and Applications, Bragança, Portugal},
year = {2008},
month = {July}
}
%%%.....
@InProceedings{CHP09a,
author = {Daniela da Cruz and Pedro Henriques and Jorge Sousa Pinto},
title = {Code Analysis: Past and Present},
booktitle = {Proceedings of the Third Int. Workshop on Foundations and Techniques for Open Source Software},
editor = {Luis Barbosa and Antonio Cerone and Siraj Shaikh },
publisher = {Electronic Communications of the EASST},
Volume = {X (2009)},
year = {2009},
month = {March},
note = {ISSN: 1863-2122}
}

```

```

@InProceedings{OPCH09a,
  author = {Nuno Oliveira and Maria João Varanda Pereira and Daniela da Cruz and Pedro Rangel Henriques},
  title = {Visualization of Domain-Specific Programs' Behavior},
  booktitle = {Proceedings of VISSOFT'09 --- 5th IEEE International Workshop on Visualizing Software for U
    Edmonton, Canada},
  pages = {37 --- 40},
  publisher = {IEEE Computer Society},
  year = {2009},
  month = {Sept}
}

@inproceedings{oliveira09b,
  title = {Applying Program Comprehension Techniques to Karel Robot Programs},
  author = {Oliveira, Nuno and and Henriques, Pedro Rangel and
    da Cruz, Daniela and Pereira, Maria João Varanda and
    Mernik, Marjan and Kosar, Tomaz and Crepinsek, Matej},
  booktitle = {Proceedings of the International Multiconference on Computer Science and Information Techn
  pages = {697 --- 704},
  publisher = {IEEE Computer Society Press},
  address = {Mragowo, Poland},
  year = {2009},
  month = {October},
}

@inproceedings{oliveira09c,
  title = {VisualLISA: Visual Programming Environment for Attribute Grammars Specification},
  author = {Oliveira, Nuno and and Henriques, Pedro Rangel and da Cruz, Daniela and Pereira, Maria João
  booktitle = {Proceedings of the International Multiconference on Computer Science and Information Techn
  pages = {689 -- 696},
  publisher = {IEEE Computer Society Press},
  address = {Mragowo, Poland},
  year = {2009},
  month = {October},
}

@inproceedings{kosar09,
  title = {Influence of domain-specific notation to program understanding},
  author = {Kosar, Tomaz and Mernik, Marjan
    and Henriques, Pedro Rangel and da Cruz, Daniela and
    Pereira, Maria João Varanda and Oliveira, Nuno},
  booktitle = {Proceedings of the International Multiconference on Computer Science and Information Techn
  pages = {673 --- 680},
  publisher = {IEEE Computer Society Press},
  address = {Mragowo, Poland},
  year = {2009},
  month = {October},
}

@inproceedings{FCHGD09a,
  author = {Flávio Xavier Ferreira and Daniela da Cruz and Pedro Rangel Henriques and Alda Lopes Gancarsk
  title= {A Query-by-Example Approach for XML Querying},
  booktitle={CISTI'09 --- 4ªConferência Ibérica em Sistemas e Tecnologias da Informação
    (WSIAI'09 -- Workshop em Sistemas Inteligentes e Aplicações)},
  editor = {Álvaro Rocha and Francisco Restivo and Luís Paulo Reis and Sofia Torrão},
  pages = {611-614},
  address= {Póvoa do Varzim, Portugal},
  year= {2009},
  month= {Jun},
}

@inproceedings{ORH09a,
  author = {Paulo Jorge Oliveria and Maria de Fátima Rodrigues and Pedro Rangel Henriques},

```



```

title=      {SMARTCLEAN: uma ferramenta para a limpeza incremental de dados},
booktitle={CISTI'09 --- 4ª Conferência Ibérica em Sistemas e Tecnologias da Informação
(WSIAT'09 -- Workshop em Sistemas Inteligentes e Aplicações)},
editor =      {Álvaro Rocha and Francisco Restivo and Luís Paulo Reis and Sofia Torrão},
pages = {621-626},
address=      {Póvoa do Varzim, Portugal},
year=         {2009},
month=        {Jun},
}
@inproceedings{LPH09a,
author =      {Giovani Rubert Librelotto and Rafael Teodósio Pereira and Pedro Rangel Henriques},
title=        {Uma metodologia para Consultas aos Bancos de Dados do NCBI},
booktitle={CISTI'09 --- 4ª Conferência Ibérica em Sistemas e Tecnologias da Informação},
editor =      {Álvaro Rocha and Francisco Restivo and Luís Paulo Reis and Sofia Torrão},
pages = {103-108},
address=      {Póvoa do Varzim, Portugal},
year=         {2009},
month=        {Jun},
}

@Article{LARH09,
author={Giovani Rubert Librelotto
and Renato Preigschadt de Azevedo
and José Carlos Ramalho
and Pedro Rangel Henriques},
title = {Topic Maps Constraint Languages: understanding and comparing},
journal = {Int. Journal Reasoning-based Intelligent Systems},
publisher = {Inderscience Enterprises, Ltd.},
year = {2009},
month = {Jun},
volume = {1},
number = {3/4},
pages = {173-181},
note = {ISSN: }
}

@inproceedings{OPHC09a,
author      = {Nuno Oliveira and Maria João Varanda Pereira and
Pedro Rangel Henriques and Daniela da Cruz},
title       = {Domain Specific Languages: A Theoretical Survey},
booktitle   = {INForum'09 --- Simpósio de Informática},
publisher   = {Faculdade de Ciências da Universidade de Lisboa},
address     = {Lisboa, Portugal},
pages       = {35 --- 46},
year        = {2009},
month       = {September}
}

@inproceedings{OPHCC09,
author      = {Nuno Oliveira and Maria João Varanda Pereira and
Pedro Rangel Henriques and Daniela da Cruz and Bastian Cramer},
title       = {{VisualLISA}: A Domain Specific Visual Language for Attribute Grammars},
booktitle   = {INForum'09 --- Simpósio de Informática},
publisher   = {Faculdade de Ciências da Universidade de Lisboa},
address     = {Lisboa, Portugal},
pages       = {155 --- 167},
year        = {2009},
month       = {September},
}

@inproceedings{MKCHCP009,

```

```

author      = {Marjan Mernik and Tomaz Kosar
               and Pedro Rangel Henriques and Daniela da Cruz and
               Maria João Varanda Pereira and Nuno Oliveira },
title       = {Comparison of XAML and C\# Forms using Cognitive Dimensions Framework},
booktitle   = {INForum'09 --- Simpósio de Informática},
publisher   = {Faculdade de Ciências da Universidade de Lisboa},
address     = {Lisboa, Portugal},
pages       = {180 --- 191},
year        = {2009},
month       = {September}
}

@inproceedings{OHCP09,
  author      = {Nuno Oliveira and Pedro Rangel Henriques and Daniela da Cruz and Maria João Varanda Pereira},
  title       = {XAGra - An XML Dialect for Attribute Grammars},
  publisher   = {Faculdade de Ciências da Universidade de Lisboa},
  booktitle   = {INForum'09 --- Simpósio de Informática},
  address     = {Lisboa, Portugal},
  pages       = {409 --- 420},
  year        = {2009},
  month       = {September}
}

@inproceedings{FCHGD09b,
  author      = {Daniela da Cruz and Flávio Xavier Ferreira and Pedro Rangel Henriques and Alda Lopes Gancars},
  title       = {GuessXQ, an inference Web-engine for querying XML Documents},
  publisher   = {Faculdade de Ciências da Universidade de Lisboa},
  booktitle   = {INForum'09 --- Simpósio de Informática},
  address     = {Lisboa, Portugal},
  pages       = {322 -- 325},
  year        = {2009},
  month       = {September}
}
%
@InProceedings{BHVU09,
  author      = {Mario Béron and Pedro Rangel Henriques and Maria João Varanda Pereira and Roberto Uzal},
  title       = {Instrumentaciones de Programas Escritos en C para Interrelacionar las Vistas Comportamental y},
  booktitle   = {CACIC'09 - XV Argentine Congress on Computer Science},
  publisher   = {Universidad Nacional de Jujuy, Argentina},
  year        = {2009},
  month       = {Oct}
}

@article{CBHP09,
  author      = {Daniela da Cruz and Mario Béron and Pedro Rangel Henriques and Maria João Varanda Pereira},
  title       = {Code Inspection Approaches for Program Visualization},
  journal     = {Acta Electrotechnica et Informatica},
  editor      = {Linus Michaeli},
  publisher   = {Faculty of Electrical Engineering and Informatics, Technical University of Kosice, Slovakia},
  year        = {2009},
  month       = {Jul-Sep},
  volume      = {9},
  number      = {3},
  pages       = {32-42},
  note        = {ISSN: 1335-8243}
}

@misc{cruz09,
  author      = {Daniela da Cruz and Nuno Oliveira and Pedro Rangel Henriques},
  title       = {GraAL - A Grammar Analyzer},
  publisher   = {Faculdade de Ciências da Universidade de Lisboa},

```

```

    year      = {2009},
    month     = {September},
    address   = {Lisbon, Portugal},
    note = {Available at: url{http://inforum.org.pt/INForum2009/programa}},
}

@inproceedings{CH09d,
    address   = {Milan, Italy},
    author    = {Daniela da Cruz and Pedro Henriques },
    booktitle = {ICEIS --- 11th International Conference on Enterprise Information Systems},
    title     = {Assessing Databases in .Net: comparing approaches},
    year      = {2009},
    month     = {May}
}

@INPROCEEDINGS{CH2010a,
    AUTHOR =      {Daniela da Cruz and Pedro Rangel Henriques},
    TITLE =      {Exploring, Visualizing and Slicing the Soul of XML Documents},
    BOOKTITLE =  {Proceedings of 25th Symposium On Applied Computing - Document Engineering},
    YEAR =      {2010},
    note =      {to be published}
}

@Article{OPHCC2010,
    author =      {Nuno Oliveira and Maria João Varanda Pereira and Pedro Rangel Henriques and Daniela da Cruz},
    title =      {VisualLISA: A Visual Environment to Develop Attribute Grammars},
    journal = {ComSIS -- Computer Science an Information Systems Journal, Special Issue on Compilers, Related Languages and Tools},
    editor = {Ivan Lukovic and António Leitão },
    publisher = {Faculty of Technical Sciences, Trg Dositeja Obradovica 6, 21000 Novi Sad, Serbia},
    year = {2010},
    month = {January},
    note = {to be published}
}

@Article{KOMPCCH2010,
    author =      {Tomaz Kosar and Nuno Oliveira and Marjan Mernik and Maria João Varanda Pereira and Matej Crnković},
    title =      {Comparing General-Purpose and Domain-Specific Languages: An Empirical Study},
    journal = {ComSIS -- Computer Science an Information Systems Journal, Special Issue on Compilers, Related Languages and Tools},
    editor = {Ivan Lukovic and António Leitão },
    publisher = {Faculty of Technical Sciences, Trg Dositeja Obradovica 6, 21000 Novi Sad, Serbia},
    year = {2010},
    month = {January},
    note = {to be published}
}

```


B.2 Alínea c

```
import re

def processa_nome(m):
    primeiro = m[1]
    ultimo = m[2].rstrip()
    return f"{ultimo},{primeiro[0]}."

text = open('exemplo-utf8.bib', 'r', encoding='UTF-8')
file = open('indice.html', 'w', encoding='UTF-8')
autores=[]
file.write('<!DOCTYPE HTML PUBLIC>\n<HTML>\n      <HEAD>\n          <meta charset="UTF-8"/>\n          <TITLE>Índice de Autores\n      \n      ')
for line in text:
    authors = re.findall(r'^(\\s)*(?i:author)(\\s)*=(\\s*)(({\\|\\"}[^}|"]*(\\}\\|\\")))', line)
    for a in authors:
        if a !=None and (a[3] not in autores):
            autores.append(a[3][1:len(a[3])-1])

resultado=[]
for element in autores:
    for i in element.split(' and '):
        resultado.append(i)
resultado.sort()
#print(resultado)

#retirar elementos repetidos da lista de autores -> working now
repetidos = []
[repetidos.append(x) for x in resultado if x not in repetidos]
repetidos.sort()
#print(repetidos)

#colocar os abreviados numa lista -> working i think
abreviados = []
for element in resultado:
    e = re.match('(\\.)(\\.)(\\.)(\\.)', element, flags=0)
    if e!= None and e not in abreviados:
        abreviados.append(e)
#print(abreviados)

#por os nomes que nao sao abreviaturas de uma lista
nome = []
for r in repetidos:
    if r not in abreviados:
        nome.append(r)

repetidos = []
[repetidos.append(x) for x in nome if x not in repetidos]
repetidos.sort()

file.write('\n          <h1>Author Index</h1>')
for r in repetidos:
    file.write(f'<P>{r} \\n</P>')
```