Ferramentas para a Construção de Arquivos Digitais de História Oral*

Silvestre Lacerda[1]
lacerda@iantt.pt
Norberto Lopes[2], Nelma Moreira[2], Rogério Reis[2]
{nml,nam,rvr}@ncc.up.pt

[1] Instituto dos Arquivos Nacionais / Torre do Tombo
[2] DCC-FC& LIACC, Universidade do Porto

Resumo Neste trabalho, apresentamos um conjunto de ferramentas, baseadas em tecnologias XML, para a produção, indexação, classificação e pesquisa de documentos multimédia associados a entrevistas. Estas ferramentas foram desenvolvidas no âmbito do projecto de história oral da Universidade Popular do Porto cujo objectivo é a preservação e divulgação da memória social e laboral do Porto no séc. XX.

Palavras chave: especificações de linguagens XML, indexação, editores de XML, *metadados*, *thesaurus*

1 Introdução

A produção de documentos estruturados e anotados é essencial para permitir a disponibilização e pesquisa de informação em formatos digitais e em particular na Web. Para tal é necessário ter definidas especificações de cada tipo de documento e ter disponíveis editores que facilitem a sua criação e modificação.

Neste trabalho, apresentamos um conjunto de ferramentas, baseadas em tecnologias XML, para a construção, indexação, classificação e pesquisa de documentos multimédia associados a entrevistas. Estas ferramentas foram desenvolvidas no âmbito do projecto de história oral da Universidade Popular do Porto (**UPP**) cujo objectivo é a preservação e divulgação da memória social e laboral do Porto no séc. XX.

A cada documento foi associada metainformação, usando especificações do *standard Dublin Core* adaptadas para o tratamento dos vários tipos de documentos relacionados com uma mesma entrevista: áudio, vídeo, transcrição, resumo, etc.

A indexação e classificação dos documentos, com recurso a palavras chave ou a palavras no texto é habitualmente feita com auxílio de vocabulários controlados, e em especial de *thesauri*. Neste contexto, foi também desenvolvida uma

* Trabalho parcialmente financiado pela Fundação para a Ciência e Tecnologia (FCT) e Programa POSI.

especificação XML para *thesauri*, assim como um conjunto de ferramentas para a sua consulta e pesquisa.

Este trabalho está organizado do seguinte modo. Na secção seguinte é apresentado o Centro de Documentação e Informação da Universidade Popular do Porto, no âmbito do qual este trabalho se insere. São aí também descritos alguns dos objectivos a atingir para a construção de uma arquivo digital de acesso livre. Na Secção 3 são apresentadas as especificações das linguagens XML usadas para alguns dos documentos associados às entrevistas. Na Secção 4 são descritos os editores de documentos que foram implementados e, na Secção 5, um interface de pesquisa nos documentos. Na Secção 6 são descritas ferramentas para a consulta e pesquisa a *thesauri*. Finalmente, na Secção 7 é indicado algum trabalho em curso e futuro.

2 Centro de Documentação e Informação sobre o Movimento Operário e Popular do Porto

O Centro de Documentação e Informação sobre o Movimento Operário e Popular do Porto (**CDI**) da **UPP** [19,20] foi criado em 2001 (com o apoio da Sociedade Porto 2001 e em colaboração com a União de Sindicatos do Porto) e tem como principais objectivos:

- contribuir para a preservação da memória e da história oral e social do Porto, valorizando o seu património social e as suas identidades;
- coligir, tratar e difundir informação sobre o movimento popular e de trabalhadores do Porto e apoiar e estimular o estudo sobre ele;
- identificar e conhecer o património arquivístico de sindicatos e outras organizações de trabalhadores do Porto, através do levantamento, diagnóstico e inventário dos seus arquivos, incluindo o levantamento da informação sobre núcleos documentais custodiados por outras instituições públicas ou privadas;
- recolher, em suporte áudio e vídeo, testemunhos e histórias de vida de pessoas que protagonizaram e/ou vivenciaram acontecimentos sociais representativos da vida social e laboral da cidade ao longo do século XX;
- disponibilizar informação relevante sobre condições de trabalho, lutas sociais, associações de trabalhadores e organizações populares, vivências das ilhas e dos bairros sociais, práticas culturais mais relevantes da "cidade do trabalho"a partir de depoimentos dos entrevistados;

O projecto "Memórias do trabalho - testemunhos do Porto laboral no século XX"enquadra o trabalho desenvolvido para preservação da memória da história oral. Existem actualmente 80 entrevistas realizadas, com uma duração total superior a 260 horas, de trabalhadores de diferentes profissões e com diferentes experiências de intervenção social, muitos deles dirigentes sindicais, activistas de associações locais, activistas políticos e presos políticos antes do 25 de Abril de 1974.

As narrativas de vida recolhidas constituem um acervo documental ímpar e de grande importância para a investigação do movimento dos trabalhadores, sobre as suas lutas, práticas culturais e condições de trabalho, no Porto no séc. XX. Cada narrativa foi registada em formatos áudio e vídeo, transcrita e elaborado um resumo que pode ser disponibilizado na `Web`. Embora alguns dos documentos multimédia possam ser de acesso restrito é importante a sua catalogação. Para os restantes e, em especial os documentos textuais é essencial a sua estruturação e anotação para o acesso, pesquisa e classificação. Desde a criação do **CDI**, em 2001, foi muito clara a opção de utilizar documentos em formatos abertos e baseados em especificações `XML`. Quer as transcrições, quer os resumos e quer os restantes documentos disponibilizados na *página* `Web` do **CDI** referentes ao cadastro dos arquivos das organizações, são anotados e podem ser pesquisados por texto ou por palavras pertencentes a uma indexação pré-definida.

Para a análise e investigação do acervo, pretende-se agora melhorar o acesso à informação já disponibilizada e criar ferramentas que ajudem a construção e edição dos documentos e dos arquivos digitais associados.

Alguns dos objectivos são:

- Permitir pesquisas mais elaboradas, pelo uso de vocabulários controlados, em especial *thesauri* na área das ciências socais e políticas.
- Permitir resultados de pesquisa mais complexos. Por exemplo: extracção de todas as referências a indústrias têxteis numa dada zona e período temporal.
- Permitir a introdução de comentários associados a segmentos de texto, num sistema aberto a qualquer interessado. Os comentários deverão ser *guardados* numa camada diferente, de modo a não alterar o texto original.
- Organizar os documentos associados a uma dada entrevista pela sua catalogação num sistema semelhante ao usado pelo ISLE (International Standard for Language Engineering) [9], para recursos multimédia para estudos linguísticos.
- Construção de um arquivo digital de acesso livre, de acordo com os princípios da OAI (Open Archives Initiative) [10], pelo uso de *metadados* e pela implementação de protocolos de recolha de informação (OAI-PHM).

De salientar ainda a utilização e o desenvolvimento de ferramentas informáticas em *software* livre. Sendo assumido que a informação contida nos documentos produzidos se pretende mais durável do que a duração habitual dos suportes de *hardware* e *software*, só com a possibilidade de ter os códigos fonte (e os poder alterar) é garantido o acesso a este espólio no futuro.

3 Especificações de Linguagens `XML`

Nesta secção descrevemos um conjunto de especificações de linguagens `XML` para diversos documentos. Todos os documentos, inclusive os multimédia, devem ter um registo de metainformação que permita a sua catalogação. A sua especificação é apresentada na Secção 3.1. Sobre cada documento de texto é possível fazer anotações que serão usadas na pesquisa ou classificação, ou constituírem

comentários que poderão ser úteis na análise. Na Secção 3.2 é descrita uma especificação para as anotações. Nas secções seguintes são apresentadas as especificações para as transcrições e os resumos das entrevistas.

O esquema de `XML` usado para definir cada linguagem foi o `Relax NG` [21]. Para além do seu poder expressivo tem uma sintaxe muito elegante e é baseado numa formalização teórica de autómatos de árvore que permite implementações claras e eficientes. Este esquema manipula elementos, atributos e texto ao mesmo nível, permitindo um uso flexível de padrões.

3.1 Descrição de *metadados* usando o *standard Dublin Core*

Para a descrição de *metadados* foram usados os 15 elementos base do *standard Dublin Core* [11], nomeadamente:

Título (`title`)
Autor (`creator`)
Editor (`publisher`)
Assunto (`subject`) Cada valor deve ser uma palavra-chave obtida por consulta dum *thesaurus*.
Colaborador (`contributor`) Para cada colaborador deve ser indicada a sua função.
Data (`date`) As datas serão associadas à criação, modificação e disponibilização do documento.
Descrição (`description`) Explicação ou resumo do conteúdo do documento.
Tipo (`type`) No caso dos resumos e transcrições será `texto`.
Formato (`format`) Uma descrição do tipo de média.
Idioma (`language`)
Fonte (`source`)
Relação (`relation`) Para além das relações habituais de pertença (`isPartOf`, `hasPart`, `isRequired`, etc.), dever-se-á relacionar cada documento com a entrevista que lhe deu origem.
Direitos (`rights`)
Âmbito (`coverage`)

Cada um destes elementos pode ser opcional ou ser repetido. Para a especialização dos elementos e para o preenchimento do conteúdo de cada elemento, optou-se pelo uso de atributos e a utilização de vocabulários controlados, em geral com especificação `XML`. Para os atributos, seguiu-se o método proposto pela comunidade OLAC (Open Language Archives Community) [16,3] de usar atributos correspondentes às qualificações `refinamento`, `codificação` e `idioma`, respectivamente `refine`, `code` e `lang`.

3.2 Anotações

Para a indexação e pesquisa nos documentos de texto, é conveniente anotar, de modo manual ou semi-automático, segmentos de texto.

As anotações não são inseridas como *elementos* `XML`, mas sim são referências a segmentos de texto. Deste modo é possível ter vários níveis de anotações e múltiplas anotações num mesmo segmento.

Foi seleccionada uma tipologia de termos que é actualmente usada para essas anotações. A classificação actual incluí: *evento*, *toponímia*, *data*, *acontecimento*, *cargo*, *nome próprio* e *nome de organização*. Cada anotação é caracterizada por um destes tipos, um comentário e marcas de início e fim do segmento de texto. O esquema da linguagem `XML` associada é apresentado na Figura 1.

```
Annotations = element Annotations {
   attribute marks { text },
   Annotation+
}
Annotation = element Annotation {
   attribute startchar { text },
   attribute endchar { text },
   attribute type {"evento" | "toponimia" | "data" |
          "acontecimento" | "cargo" |
           token "nome proprio" | token "nome organizacao" },
   attribute normtext { text },
 text}
```

Figura 1. Esquema `Relax NG` para as anotações.

Pretendemos estender o sistema, de modo a permitir anotações associadas a termos de um *thesaurus* ou outros vocabulários controlados. Note-se que este tipo de anotações poderá também ser feita para documentos de áudio ou vídeo.

3.3 Transcrições

Uma da tarefas mais laboriosas no tratamento de entrevistas (ou outros registos de fala) é a transcrição do áudio para texto, e poucas ferramentas existem em *software* livre. No decurso do projecto utilizou-se inicialmente uma ferramenta desenvolvida na Universidade do Minho, `Escriba` e mais recentemente o `transcriber` [14,1], uma aplicação que é usada pela comunidade linguística internacional mas é especialmente vocacionada para transcrições de emissões radiofónicas. Actualmente, está em curso a construção duma aplicação de transcrição, que se adapte melhor aos requisitos deste projecto.

Numa transcrição, será necessário ter informação sobre cada interveniente e a divisão da entrevista em diversas conversações. Na Figura 2 apresentamos o esquema da especificação para transcrições. De notar a ausência de informação sobre a qualidade ou características da gravação áudio, mas tal informação não nos parece essencial para este tipo de transcrição e poder ser difícil de especificar por um transcritor não especialista em áudio.

```
include "annotations.rnc"
include "metadata.rnc"

start = TranScript
TranScript = element Transcript {FileData, SecTion+, Annotations?}
FileData = element Filedata { MetaData, PerSon+ }
PerSon = element Person { attribute id { text },
  attribute name { text },
  attribute email { text }?,
  attribute type {"interviewer" |
                      "interviewee" | "other" }}
SecTion = element Section { attribute title { text },
    attribute desc { text },
    ConverSation+ }
ConverSation = element ConverSation {attribute starttime { text },
 attribute endtime { text },
 attribute title { text },
 PerSon,
 text}
```

Figura 2. Esquema `Relax NG` para as transcrições.

Na literatura não é do nosso conhecimento nenhum *standard* para especificações de documentos deste tipo, havendo caracterizações no âmbito do projecto TEI [5] e de alguns projectos de história oral [6].

3.4 Resumos

Os resumos das entrevistas são pequenos textos de acesso livre que sumarizam as narrativas. Presentemente são a maior fonte de informação para quem pretende consultar o arquivo. Os resumos podem ser anotados e estão divididos em secções, que designamos por `histórias`. O esquema apresentado na Figura 3 é baseado no já utilizado pelos resumos que estão disponíveis na *página* `Web` do **CDI**. As principais alterações foram a separação da metainformação e o facto das anotações não corresponderem a elementos `XML` inseridos no texto.

4 Editores

Embora existam editores genéricos para documentos `XML`, estes normalmente obrigam a conhecimentos mínimos da sua sintaxe, assim como do esquema específico associado a cada documento. Ambientes gráficos que escondam pormenores técnicos para o utilizador não especializado facilitam e aceleram a produção e modificação de documentos estruturados. Exemplos de editores de anotações `XML` são o UXO [15] para anotações linguísticas e que permite acesso a base de dados, e o Cadixe [4], desenvolvido no âmbito de um projecto de bioinformática.

```
include "metadata.rnc"
include "annotations.rnc"

start = Resume
Resume = element Resume { MetaData, Stories, Annotations? }
Stories = element Stories {
   attribute title { text },
   attribute desc { text },
   Story+
}
Story = element Story {
   attribute startchar { text },
   attribute endchar { text },
   attribute title { text },
   attribute desc { text },
   text
}
```

Figura 3. Esquema `Relax NG` para os resumos.

Por outro lado, há actualmente a tendência para que os editores sejam integrados em navegadores `Web`. No entanto não optamos por esta alternativa, pois não nos parece que os benefícios desse tipo de ferramentas – não ser necessária instalação e acesso directo a arquivos remotos, por exemplo – compensem as suas desvantagens, nomeadamente de maior custo de desenvolvimento e limitações dado envolverem diferentes tecnologias `Web`, para além da obrigatoriedade do utilizador ter de estar *on-line* para a produção dos documentos.

Assim, optou-se pelo desenvolvimento de aplicações gráficas que contudo devem ser de instalação fácil em qualquer plataforma. O desenvolvimento do sistema informático do **CDI** é essencialmente baseado na linguagem de programação `Python` [22]. Os interfaces gráficos utilizamos a `API` gráfica `wxWidgets` [18], devido à sua portabilidade e vasto `toolkit` disponível. Em particular, a construção dos editores de texto baseia-se na classe `wxStyledTextCrl` que implementa a poderosa componente de edição de texto `Scintill` [8].

A manipulação de documentos `XML` é feita essencialmente através da `API` `elementtree` [12], mas usando a implementação `lxml` [13] que é baseada na biblioteca `libxml2` [17] e tem um maior suporte de `XSLT` e `XPath`. As principais vantagens desta `API` são a sua proximidade com a estrutura do `Python`, a sua maior velocidade e pouco gasto de memória: cerca de 15 vezes mais rápida e 5 vezes menos consumidora de memória do que a mais recente biblioteca de `DOM` do `Python`, e cerca de 2 vezes mais rápida que bibliotecas de `SAX`.

4.1 Editor para metainformação

O editor de *metadados* pode ser usado integrado no editor de documentos de texto ou ser usado isoladamente para a catalogação de outros documentos. Em

relação à especificação referida na Secção 3.1, tem algumas limitações, por exemplo apenas permite que os documentos tenham só um título. Utiliza diversos vocabulários controlados. Para além dos habituais *standard*s para idiomas, tipos e formatos de documentos, etc., tem vocabulários específicos, por exemplo para funções dos colaboradores (entrevistadores, transcritores, etc.). De futuro, pretende-se que estes vocabulários possam ser escolhidos/configurados para cada elemento. A possibilidade de inserção de qualquer número de elementos de cada tipo será também revista, mas de modo a que o editor continue simples e de uso intuitivo.

4.2 Editor para a anotação e classificação de documentos

O editor de resumos permite a criação e modificação de documentos que seguem a especificação dada na Secção 3.4. Tem contudo algumas restrições conceptuais. Foi desenhado não como um editor genérico de texto, mas sim como um classificador e anotador de textos. Assim, após a anotação de um segmento de texto não é possível a edição de texto da secção correspondente (neste caso, duma *história*). Deste modo é possível, garantir que quem anota não altera o conteúdo do que está escrito. Poderá contudo, ser desejável um sistema mais flexível mas em que a opção de não edição continue disponível. A inserção de histórias está actualmente a ser generalizada para a classificação de textos segundo uma hierarquia de categorias. Deste modo, pretende-se desenvolver um editor de classificação (e anotação) de documentos mais geral e que será usado para a análise de conteúdo das narrativas. Na Figura 4 apresentamos uma imagem tanto do editor de *metadados* como do de resumos.

O editor de transcrições deve ser integrado numa ferramenta que permita o acesso e manipulação de ficheiros áudio e, embora se possa aproveitar o esquema geral do que foi desenvolvido para os resumos não foi ainda implementado na sua totalidade.

5 Pesquisa nos documentos

Foi implementado um protótipo de uma *página* `Web` para o acesso e pesquisa nos documentos produzidos com as novas ferramentas, em particular, nos resumos de entrevistas. A pesquisa é feita com recurso ao `XPath`. Ainda não está implementado nenhum sistema de indexação que optimize a pesquisa e permita pesquisas e resultados mais avançados como os descritos na Secção 2.

6 Manipulação de um *thesaurus*

Os *thesauri* são hierarquias semânticas de termos, que se situam entre a mera classificação e ontologias mais complexas. Cada termo pode ter associados outros termos por diversas relações, em particular hierárquica e equivalência. Os termos *descritores* são definidores de um conceito e devem ser usados em detrimento de outros que sejam considerados no *thesaurus* como equivalentes.

Figura 4. Editores de resumos e de *metadados*.

Existem disponíveis *thesauri* em diversas áreas temáticas e daí ser importante o acesso à sua manipulação digital. O `Zthes` [23] é um *standard* para a representação, acesso e navegação em *thesauri*. Quando este trabalho começou a ser desenvolvido, no início de 2006, havia apenas disponível um esquema em `DTD` para a versão 0.5 do `Zthes` e que foi adaptado para o esquema em `Relax NG` que se apresenta na Figura 6. Mais recentemente, a *página* `Web Zthes` foi actualizada e disponibiliza diversos esquemas para uma nova versão do `Zthes`, a 1.0.

Tendo a **UPP** acesso a um *thesauri* pluridisciplinar, foi feita a sua conversão para a linguagem `Zthes`. Este *thesaurus* é constituído por cerca de 10000 termos e organizado em 25 áreas temáticas.

A navegação no *thesaurus* pode ser por:

- listagem alfabética de todos os termos
- listagem hierárquica organizada por *microthesauri* dos termos descritores

Para optimizar a listagem hierárquica de termos, foi implementada uma indexação dos termos num dicionário `Python`.

A pesquisa de termos utilizando o `XPath`, pode ser exacta ou aproximada, e permitir múltiplas palavras. O *thesaurus* pode ser consultado através de um interface `Web`, como o exemplificado na Figura 7. Dada a relativa dimensão do *thesaurus* que utilizamos, o desempenho dos procedimentos de pesquisa e navegação, baseados quase exclusivamente em manipulação de `XML`, é bastante satisfatório.

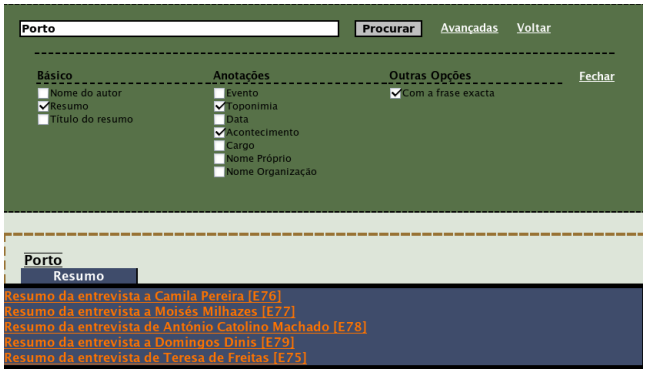

Figura 5. Pesquisa nos resumos.

7 Conclusões e Trabalho futuro

As ferramentas descritas apesar de estarem ainda em fase de desenvolvimento, já podem ser utilizadas. É necessário melhorar os editores de modo a serem mais gerais e mais flexíveis; permitirem a classificação e anotação de textos genéricos, pela definição duma hierarquia de categorias; e integrarem o acesso a *thesauri*. Em especial, pretendemos analisar a possível utilização do modelo de grafos de anotação [2,7] que permite uma maior independência entre o interface com utilizador e os formatos dos documentos. Num grafo de anotação a informação é representada por um conjunto de nós e arcos que podem ser anotados com os mais variados atributos. Note-se que a estrutura usada para implementação de anotações segue já este tipo de formalismo.

Durante 2006 decorreu o projecto *Memórias, Vivências e Identidade dos Trabalhadores Têxteis do Porto - Percursos de Vida e de Trabalho na Indústria Têxtil*, no âmbito do Concurso *Investigação Científica na Pré-graduação*, da Universidade do Porto. Dos materiais produzidos, incluí-se o tratamento de seis entrevistas e algumas das ferramentas informáticas aqui descritas, e que foram já utilizadas no decurso desse projecto.

Os documentos existentes no **CDI** irão ser convertidos para os novos formatos e será desenvolvida uma nova versão da *página* `Web` do **CDI**, segundo os objectivos definidos na Secção 2.

8 Agradecimentos

O trabalho aqui descrito teve contribuições dos vários membros da equipa do projecto *Memórias do Trabalho- Testemunhos do Porto Laboral no Século XX*, nomeadamente Manuel Loff, Teresa Medina, Cristina Nogueira e Eloy Rodrigues. A selecção, recolha e tratamento das narrativas envolveu e envolve um grande número de pessoas em especial alunos de licenciaturas da Universidade do Porto. Entre 2000 e 2004, Luís Pessoa colaborou no desenvolvimento e manutenção da actual *página* `Web` e ferramentas informáticas associadas.

```
start = Zthes
Zthes = element Zthes {Term+}
Term = element term {terment,
   termnote?,
   admin?,
   relation*}
terment  = attribute termId {text},
        element termName {text},
        element termQualifier {text}?,
        attribute termType {text}?,
        attribute termLanguage {text}?
admin = attribute termCreatedDate {text}?,
        attribute termCreatedBy {text}?,
        attribute termModifiedDate {text}?,
        attribute termModifiedBy {text}?
termnote = element termNote {text,
   attribute type { "SN" | "NE"}}
relation = element relation {
   attribute relationType {text},
   element sourceDb {text}?,
   attribute termId {text},
   attribute termName {text}
}
```

Figura 6. Esquema `Relax NG` para *thesauri*.

Referências

1. Claude Barras, Edouard Geoffrois, Zhibiao Wu, and Mark Liberman. Transcriber: development and use of a tool for assisting speech corpora production. *Speech Communication*, 33(1–2):5–2, 2001.
2. Steven Bird and Mark Liberman. A formal framework for linguistic annotation. *Speech Communication*, 33(1–2):23–60, 2001.
3. Steven Bird and Gary Simons. The OLAC metadata set and controlled vocabularies. In *Proceedings of the ACL 2001 Workshop on Human Language Technology and Knowledge Management*, pages 7–18, Morristown, NJ, USA, 2001. Association for Computational Linguistics.
4. Gilles Bisson. *Cadixe user manual: why and what.* CADERIGE Project, 2005.
5. The TEI Consortium. Tei: The Text Encoding Initiative. `http://www.tei-c.org/`, 2006.
6. Janet Crum. Oral history markup language. `http://www.ohsu.edu/library/staff/crumj/oml`, 1999.
7. Edouard Geoffrois, Claude Barras, Steve Bird, and Zhibiao Wu. Transcribing with annotation graphs. In *Second International Conference on Language Resources and Evaluation (LREC)*, pages 1517–1521, 2000.
8. Scintilla Project Group. Scintilla. `http://www.scintilla.org/`, 2006.
9. IMDI Team. IMDI Metadata Elements for Session Descriptions. Technical report, MPI Nijmegen, October 2003.

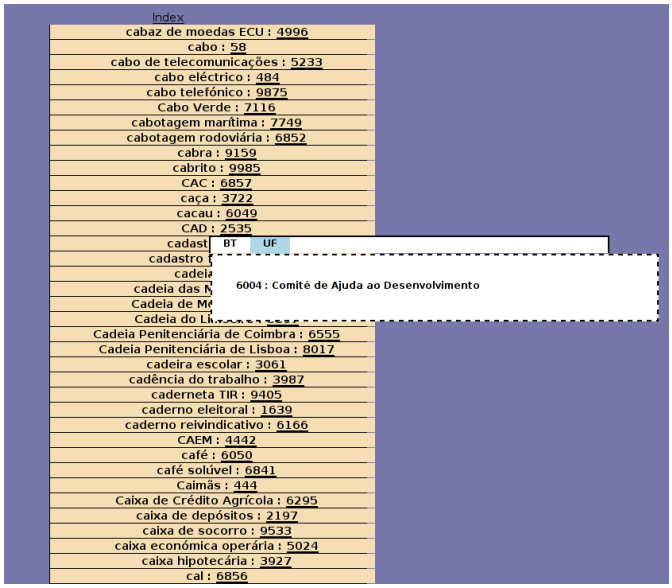

Figura 7. Pesquisa alfabética num *theasaurus*

10. Open Archives Initiative. OAI-PMH Core Resources. `http://www.openarchives.org/`, 2006.
11. The Dublin Core Metadata Initiative. Dublin Core Metadata Terms. `http://dublincore.org/`, 2006.
12. Fredrik Lundh. ElementTree API. `http://effbot.org/zone/element-index.html`, 2006.
13. lxml development team. lxml. `http://codespeak.net/lxml/`, 2006.
14. Sylvain Galliano Mathieu Manta, Fabien Antoine and Claude Barras. Transcriber. `http://trans.sourceforge.net/`, 2006.
15. Jan-Torsten Milde. Uxo: An xml-based extensible annotation editor. In *Proccedings of the GLDV-Spring Meeting*, pages 151–159, Giessen University, 2001.
16. Open Language Archives Community. Olac metadata. `http://www.language-archives.org/OLAC`, 2006.
17. The GNOME Project. The XML library for Gnome. `http://xmlsoft.org/`, 2006.
18. Julian Smart, Robert Roebling, Vadim Zeitlin, and Robin Dunn. *wxWidgets 2.6.3: A portable C++ and Python GUI toolkit.*
19. Universidade Popular do Porto. Centro de Documentação e Informação sobre o Movimento Operário e Popular do Porto. `http://cdi.upp.pt/`.
20. Universidade Popular do Porto. O Centro de Documentação e Informação sobre o Movimento Operário e Popular do Porto da Universidade Popular do Porto. In Manuel Loff e Maria da Conceição Meireles Pereira, editor, *PORTUGAL: 30 ANOS DE DEMOCRACIA (1974-2004)*, pages 287–294. Universidade do Porto, 2006.
21. Eric van der Vlist. *RELAX NG*. O'Reilly, 2003.
22. Guido van Rossum. *Python Library Reference*, 2.4.2 edition, 2005.
23. The Zthes working group. The Zthes specifications for thesaurus representation, access and navigation. `http://zthes.z3950.org/`, 2006.