Adapting web content for low-literacy readers by using lexical elaboration and named entities labeling

William M. Watanabe, Arnaldo Candido Jr., Marcelo A. Amâncio, Matheus de Oliveira, Thiago A. S. Pardo, Renata P. M. Fortes and Sandra M. Aluísio
Institute of Mathematical and Computer Sciences, University of São Paulo
400, Trabalhador Sáo-carlense Avenue - Centro
P.O.Box 668. 13560-970 - São Carlos/SP, Brazil
\{watinha@, arnaldoc@, amancio@, matheusuol@grad., taspardo@, renata@, sandra@\}icmc.usp.br,

ABSTRACT
In this paper we describe a web content adaptation tool for assisting low-literacy readers to access online information. The “Educational FACILITA” tool provides innovative features and the design of more intuitive interaction models. Especially, we propose an interaction model and web application that explore the Natural Language Processing tasks of lexical elaboration and named entity labeling for improving web accessibility.

Categories and Subject Descriptors
H.3.1 [Content Analysis and Indexing]: Dictionaries, Linguistic processing, Thesauruses; H.5.3 [Group and Organization]: Web-based interaction; H.5.4 [Hypertext/Hypermedia]: User issues

General Terms
Design, Human Factors

Keywords
Web content adaptation, Web accessibility, Natural language processing

1. INTRODUCTION
The large capacity of Web for providing information leads to multiple possibilities and opportunities for users. The development of high performance networks and ubiquitous devices allow users to retrieve content from any location and in different scenarios or situations they might face in their lives. Further to retrieving information, current technologies also enable users to contribute to the authoring of content on the Web by means of forums, blogs and wikis, in the so called “Web 2.0” [33].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Copyright 2010 ACM 978-1-4503-0045-2 ...$5.00.

Unfortunately the possibilities offered by the Web are not necessarily currently available to all. Individuals who do not have completely compliant software or hardware that are able to deal with the latest technologies, or have some kind of physical or cognitive disability, find it difficult to interact with web pages, depending on the page structure and the ways in which the content is made available. In this context, Web accessibility studies work towards implementing processes and recommendations that lead the design of interfaces that have their access granted for all users regardless of the special needs they might present.

The pre-eminent reference when addressing Web accessibility is the WCAG (Web Content Accessibility Guidelines) [19, 12]. The WCAG establish a set of guidelines that discuss accessibility issues and provide accessibility design solutions [30, 31]. It requires the manual implementation of technological and design solutions by Web developers and content authors [34, 32]. However, there is no guarantee that the guidelines will be followed. Freire et al., for example, conducted a survey about accessibility awareness of people involved in Web development in Brazil [13] and the web accessibility in Brazilian Municipalities websites [11]. Their conclusions were that web accessibility is far from being actually considered in Brazil and that much work remains to be done.

WCAG documents address not only structure and technological aspects. Guideline 14 in WCAG 1.0 (to ensure that documents are clear and simple) and Guideline 3.1 (to make text content readable and understandable) in WCAG 2.0, for example, provide recommendations about how the content should be made available to users. However, Web developers are not always responsible for content preparation and authoring in a Website. Moreover, in the current “Web 2.0” context, users, without any prior knowledge about the guidelines, directly participate in the content authoring process of Web applications. This may make it even more difficult to develop completely WCAG conformance Websites.

Since 2001, the INAF index (National Indicator of Functional Literacy) has been annually computed to measure the levels of functional illiteracy of Brazilian population. The 2009 report presented a still worrying scenario: 7% of the individuals were classified as illiterate; 21% as literate at the rudimentary level; 47% as literate at the basic level; and only 25% as literate at the advanced level [22]. Thus, we argue that an assistive technology for adapting web content is an urgent necessity for digital inclusion of low literacy
people, specially when we consider that a large portion of Brazilian people (about 28%) are functional illiterate.

In this scenario, we present in this paper a Web technology and Natural Language Processing (NLP) combined solution to adapt web content for poor literacy readers. Particularly, mainly following WCAG, we explore the NLP tasks of lexical elaboration and named entity labeling to assist poor readers having access to web content. We developed an interaction model and a web application to provide such functionalities.

Our on-line tool, called “Educational FACILITA”\(^1\), is part of a bigger project – PorSimples \(^1\) – on Text Simplification for Brazilian Portuguese language, which aims at producing tools for aiding web text authors to produce simpler texts and for assisting (poor) readers to have access to relevant written material.

Although the work described in this paper was originally designed for addressing a local (Brazilian Portuguese) issue, our solution was intended to be generic enough to be applied for other languages. In fact, most of our decisions were based on prominent works on Text Adaptation for other languages, some of which will be reported in the next section.

In the next section we briefly introduce the main related works in the area, while our proposal is described in Section 3. Some final remarks are made in Section 5.

2. RELATED WORK

The WCAG checkpoint 14.1 (WCAG 1.0) and success criteria 3.1.5 (WCAG 2.0) discuss the importance of Websites containing textual content with reading levels that match users reading skills.

In order to automatically adapt Websites to the WCAG checkpoint 14.1 “Use the clearest and simplest language appropriate for a site’s content” (WCAG 1.0) and the success criteria 3.1.5 “Reading Level” (WCAG 2.0), Watanabe et al. developed the Facilita application \(^3\). Facilita works as a browser plug-in and uses Automatic Summarization and Syntactic Simplification operations to adapt Website textual content into accessible versions that would be more adequate for low literacy users.

The present work builds up on Watanabe et al. Facilita application \(^3\), however, instead of only simplifying texts available on web pages, we automatically provide textual elaboration (EL) using synonyms for the words which we classify as difficult to be understood by low-literacy readers and also add semantics for specific textual elements available on the Web page. Specifically, the semantics that we add is related to named entities recognition and classification (NERC), which is a task generally associated with the area of information extraction (IE). Following we review several works related to our proposal.

2.1 Content Presentation and Interaction

Many studies report on the development and design of interfaces for low literacy individuals \(^3\). In order to enhance the content comprehension of the content, the studies use different kinds of media and communication modes that partially or totally replace the textual content available on the interfaces.

One of the ways to replace the textual media on the interfaces used in previous studies is the voice and audio interacting with the web page.

The automatic content generation context and deployment over the Web architecture implies some technological restrictions on the applications and interfaces implementation. Particularly, the automatic retrieval of images or graphical representations that fully address the main theme of a Web page or even replace it entirely.

Specifically, when considering the technological restrictions imposed by the Web architecture, the use of audio and voice interfaces is important. Although speech versions of the content can be offered by technologies such as applets for static content, the automatic speech recognition or text-to-speech resources have very restrict uses on the Web platform. Multimodal development technologies, such as the XHTML+Voice \(^3\), already address this issue, but the XHTML+Voice technology currently is only available for the English language. Regarding the Portuguese language, the “Internet Browser with Speech Recognition and Synthesis”\(^3\) aims at the development and improvement of an Internet browser with speech recognition and synthesis functionality for this language, however the project is currently on initial stages.

Web pages often contain pop-up, ads information, and complementary information that might distract or make it difficult for users to access the main content of a Web page. Gupta et al. \(^15\) implemented an automatic content extraction mechanism in order to aid visually impaired and blind users read a page by not having the screen reader software to read all the clutter in the page before reaching the main content. The approach used by Gupta et al. describes the use of a host of heuristics that filters web page’s clutter (adds, animations and extraneous links) to provide users with a more accessible version of the web page. They use the DOM structure of a web page in order to navigate through its content and extract information that might disturb users searching and reading the content. Although the approach used by Gupta et al. removes extraneous content from the web page and might change its layout, it does not adapt the inner parts of the elements that were not filtered. The approach is responsible for deciding on which elements will be displayed, however it does not affect the content that is inside them.

Bigham and Ladner \(^6\) introduce the Accessmonkey framework as a contribution towards the goal of automatically

\(^{1}\) http://vinho.intermidia.icmc.usp.br/watinha/Educational-Facilita/

\(^{3}\) http://www.bv.fapesp.br/projetos-pipe/637/navegador-internet-reconhecimento-sintese-fala
helping users with special needs interact with the Web. Big-
ham and Ladner’s work suggests the use of a Javascript
framework to deliver browser plugins that automatically a-
dapt the webpage in order to make it more accessible to
users. Even though the use of local scripting languages
might provide a great deal of possibilities for the develop-
ment of web accessibility solutions, the study does not focus
on how to adapt the webpages textual content. Our ap-
proach resembles the use of a Javascript framework since we
rely on the use of Jetpack⁷, a Javascript language frame-
work aimed at implementing Firefox browser addons. Our
work directly addresses the textual adaptation scenario on
webpages, describing other resources that are required to
provide the assistance.

2.2 Text adaptation and NLP

Text adaptation is a very well known practice used in ed-
ucational settings. Young [36] mentions two different tech-
niques for text adaptation: Text Simplification and Text
Elaboration. The first can be defined as any task that re-
duces the lexical or syntactic complexity of a text while try-
ing to preserve meaning and information and it is subdivided
into Syntactic Simplification, Lexical Simplification, Auto-
matic Summarization, and other techniques. Text Elab-
oration aims at clarifying and explaining information and
making connections explicit in a text, using definitions, syn-
onyms or hypernyms of the text words. Text simplification
can be used to maximize the comprehension of low literacy
users since mismatching the reading level of the application
textual media with the reading skills of users can impact
severally on the users access to the application [4, 26].

Specifically for English language learners (ELLs), educa-
tional research suggests that text adaptation can facilitate
both reading comprehension of content and English lan-
guage skills development [7]. According to Yano et al. [35],
text modifications seem to have positive effects on second
language comprehension. Another study by Urano [28] also
present some support for the assumption that lexical modifi-
cation, or at least lexical elaboration, facilitates second lan-
guage reading comprehension. The results from his study
confirmed that elaboration is more favorable than simplifi-
cation in terms of vocabulary acquisition. Moreover, he also
states that one occurrence of each target word with lexical
elaboration may not be sufficient to acquire its meaning, al-
though more advanced learners may be able to learn some of
the words with only one occurrences. Although our en-
visioned user is different from the ones of Urano, we follow
this study by using two Portuguese thesauruses for present-
ing simpler synonyms for complex words.

Petersen and Ostendorf [23] show two important opera-
tions for lexical adaption for bilingual education: (a) to
remove unnecessary information from the text, and (b) to
add information that better explains difficult terms. Our
work follows the second option since we use short definitions
from Wikipedia to define some text entities and propose the
use of other Web applications and Social Web resources to
improve the identification and to better explain the textual
entities found on a Web page.

Elhadad [10] investigates how to improve access to medical
literature for health consumers, focusing on medical termi-
ology. They present a method to predict automatically in

⁷https://jetpack.mozillalabs.com/
naturally extract the textual content to be passed on to the textual processing operations, to have users activating the assistance directly from the Web page, and to present the results directly on the Web application in which the user was reading the original content.

3. EDUCATIONAL FACILITA

In Brazil, a large part of the population faces difficulties with reading and comprehending texts depending on text size and linguistic complexity; therefore, access to media that use texts as their primary way to convey information is limited. In this scenario, we developed Educational FACILITA, which is a Web application aimed at assisting users in understanding textual content available on the Web.

The WAI (Web Accessibility Initiative) through the WCAG 2.0 in the Success Criteria 3.1.3 acknowledges the importance of providing mechanisms to help users to understand unusual words that might be presented on Web sites. Educational FACILITA, in this context, automatically adapts the content of Web sites in order to provide mechanisms that present users with synonymous or short definitions for words, which are classified as unusual, or difficult ones to be understood by the users. Educational FACILITA also displays additional and complementary information about named entities (NEs) that are contained on the Web sites text. Educational FACILITA’s assistance is illustrated on Figure 1, representing the reading assistance mechanisms automatically included on the web page to be presented to the end users.

It is expected that this additional information presented in the text by the proposed approach would help users better understand Websites’s textual content and allow users to learn the meaning of new or unusual words/expressions.

In order to present the reading assistance to users, Educational FACILITA makes use of the following NLP (Natural Language Processing) modules: Lexical Elaboration (Section 3.1.1) and Named Entities Recognition and PoS-Classification (Section 3.1.2).

We were also very careful with the interface design of Educational FACILITA. Given our envisioned final user, it is necessary to develop an interface that takes low-literacy individuals limitations into account, implementing interactions that are simple to use, and more adequate to the users computer skills. It is worth observing that Educational FACILITA requires two resources: (a) the browser Firefox 3.5 previously installed, and (b) the Jetpack installed.

In this context, we also report the development and integration of different Web technologies that were used with the intention of keeping the application’s interaction as simple and as natural as possible. As a result, Educational FACILITA became able to automatically identify the textual content on the Web page, and insert the reading assistance mechanisms in the original Web page itself, without having to change its design or functionality. Therefore Educational FACILITA requires less effort of users.

In the following sections, we first describe our NLP modules (Section 3.1), the key features designed for Educational FACILITA (Section 3.2), and the implementation details (Section 3.3), including the proposed architecture and the technological requirements.

3.1 NLP Modules

3.1.1 Lexical Elaboration

The first part of the lexical simplification consists of tokenizing the original text and marking the words that are considered complex. In order to judge a word as complex or not, we use 3 dictionaries created for the PorSimples Project, as cited in Section 2: one for common words to youngsters, one composed of frequent words, and one of concrete words. If a word is not in the simple words dictionary and is not a proper noun, we assume it to be a complex word.

The lexical elaboration module also uses the Unitex-PB dictionary5 for finding the lemma of the words in the text, so that it is possible to look for it in the simple words dictionary.

The problem of looking for a lemma directly in a dictionary is that there are ambiguous words and we are not able to deal with word senses. For instance, in Portuguese, the word “canto” may mean the noun “corner”, as well as the verb “sing”.

For dealing with the Part of Speech (PoS) ambiguity, we use the MXPOST PoS tagger25 trained over NILC tagset, whose purpose is to automatically tag the text and identify each word PoS category.

After the text is tagged, the words from the text that are not proper nouns, prepositions and numerals are selected, and their PoS tags are used to look for their lemmas in the dictionaries. The tagger does not acheive not a 100% precision and some words may not be in the dictionary. Hence, we look for the lemma only (without the tag) when we are not able to find the word-lemma combination in the dictionary. Still, if we are not able to find the word, the lexical elaboration module assumes that the word is complex and marks it.

The last step of the process consists in providing simpler synonyms for the marked words. For this task, we use the thesauruses TeP 2.06 and PAPEL6 for Portuguese language.2 This operation is carried out when the user clicks

\[\text{http://www.nilc.icmc.usp.br/nlinc/projects/unitexpb/web/dicionarios.html}\]

\[\text{http://www.nilc.icmc.usp.br/nlinc/TagSet/ManualEtiquetagem.htm}\]

\[\text{http://www.nilc.icmc.usp.br/top2/}\]

\[\text{http://www.linguateca.pt/PAPEL/}\]
on a marked word. It triggers a search in the thesauruses for synonymous words that are also in the common words dictionary. If simpler words are found, they are listed in order, from the simpler to the more complex ones. To determine this order, we used Google API to search each word in the web: we assume that the more frequently a word happens, the simpler it is. Although this may lead to some imprecision, it is a good approach to the problem.

As far as we know, TeP has about 45,000 words grouped into 20,000 synonym clusters. PAPEL is a larger repository, with about 100,000 words.

The general process of annotating and dealing with complex words is shown in Figure 2.

3.1.2 Named entities recognition and pos-classification

Following results from educational research that suggests that text adaptation can facilitate both reading comprehension and vocabulary acquisition, we added encyclopedia knowledge as additional information for some terms from the text. Particularly, we added information from Wikipedia. Figure 3 shows the general architecture.

The chosen terms are the named entities (NEs), which were already defined in Section 2, to which we add small extracts from Wikipedia. We look for extracts that define the NE under focus, which generally appear in the beginning of the Wikipedia documents.

Names of people, places and organizations are examples of NEs, e.g., “João Almeida”, “Rio de Janeiro” and “Petrobrás”. NE recognition (NER) corresponds to the task of automatically identifying such terms. The sentence below shows an example of a sentence with a NE identified:

\[ <\text{NE class}='\text{PersonName}'>\text{Alberto Santos Dumont}</\text{NE}> \text{ was a Brazilian inventor.} \]

It is important to notice that the NER task may be also manually carried out. Usually, if the annotation task is well defined, humans perform better than NER systems.

In previous evaluations of NER systems for Portuguese conducted by Linguateca7, one of the best systems was Rembrandt [8], achieving a general accuracy of about 57%. It classifies NEs into 9 general classes, which in turn may be subcategorized into other 47 classes, and accesses Wikipedia for improving its categorization and indexing the NEs. Therefore, besides the class of a NE, the system also retrieves its basic definition from Wikipedia. The system is freely available and open source.

For this project, we had to adapt Rembrandt to our final users. The first adaptation was the redefinition of the classes, in order to remove complex subcategories and to change difficult class names to more common ones. These changes make the NE labels more easily understandable by poor readers. For instance, the subcategory “ephemeris” of the class “happening” was renamed to “historical event”. Such decisions were based on consults to several information sources, as TeP 2.0, Wikipedia, and dictionaries.

We also verified whether the first sentences of Wikipedia pages were really the definition of NEs. We performed a statistical test with a group of 10 randomly selected pages from DBPedia, which is a sample of Wikipedia with 250,000 articles. In 9 out of this 10 groups we got a positive answer, what results in a 95% confidence level for our hypothesis. Our estimate is that 73.5% of Wikipedia articles are definition sentences.

3.2 Key features

One of the biggest concerns when dealing with low-literacy users is related to their lack of computer experience. Taking that into account, all of our design decisions were made with focus on simplicity of interaction, in order to require less effort from users.

The first key feature of Educational FACILITA is to automatically extract textual content from the Web page currently being accessed (viewed) by the user. The second one is to integrate the reading assistance mechanism on the same Web page that activated Educational FACILITA.

In order to implement the functionalities, Educational FACILITA makes use of the following textual processing modules:

- **Readability module**: some of the text processing techniques that we use require the establishment of concrete relations among the words that belong to a phrase in the text. From this stage we can identify phrases inside any Web page that is being presented to users we adapted the Readability application1 to only identify the main textual content present into Web pages. The Readability Module selects a specific element from the DOM tree of the Web page as the main element of the document, in which it becomes easier to indentify phrases.

- **Lexical Elaboration**: this module evaluates the complexity of each word separately. Accordingly to the result of this evaluation Educational FACILITA will present the user with synonyms for the complex words. The synonyms provided by Educational FACILITA are simpler, therefore more easily understood by users.

- **Named Entities Recognition**: this module is responsible for identifying and presenting users with information about named entities in a text. In order to

---

7http://www.linguateca.pt/harem/
make this information available to users, we use Rembrandt to both label the named entities and also direct them to their correspondent link in the Wikipedia\textsuperscript{8}. It is important to notice that the Named Entities Recognition methods identify the class of the named entities present in the text. And by doing that we can extend Educational FACILITA prototype to search for references about the named entity identified in other resources available on the Web, in the future. For example, if we find the named entity “New York”, the named entity recognition system would identify it as being of the class “Location”. With that information we could look for a reference of the named entity found in the Google Maps Web application\textsuperscript{9} and present the query’s result to the user. It is important to notice also that even though there are possibilities of integrating Educational FACILITA with others Web applications like the one mentioned previously, currently the prototype only implements the Wikipedia short definition and image extraction routines. This kind of operation is illustrated on Figure 3.

The three modules work as presented on Figure 4. First we execute the readability module on the Web page to identify the main text contained in it. Next we simultaneously run the Named Entity Recognition and the Lexical Elaboration modules having as input the main text identified by the Readability module. Next, we integrate the outputs from both the Named Entity Recognition and the Lexical Elaboration modules into a single HTML code. Finally, we insert the HTML code into the Web page which activated the Educational FACILITA assistance.

It is worth noticing the Educational FACILITA was implemented, initially, just for Brazilian Portuguese. Given that the applications and modules already implemented are devoted to this language. However, porting Educational FACILITA to other languages just require the replacement of the NLP modules (Section 3.1) by other modules with the same functionality for the respective languages of interest, preserving the same architecture.

To avoid interfering the accessibility mechanisms or design characteristics already implemented on the Web site that activated the reading assistance, while integrating the generated mechanisms on the Web page, Educational FACILITA attempts to change the page’s functionality and design at minimal rates.

The integration with minimal changes of original web design exploited in Educational FACILITA was achieved just by adapting the Web page parts pointed out by the textual processing methods. Our approach respects the existing Web mechanisms. For example, when our approach points out accessibility issues in parts of the original Web page that we identify some kind of mechanism or functionality already implemented (such as a link or anchor) Educational FACILITA chooses to keep the previously implemented functionality unchanged.

3.3 Web implementation

Educational FACILITA’S development required several resources that are not directly available for common Web technologies. A few clear examples of development limitations of currently available resources imposed by the Web platform include:

- **Same-origin policy**: the same-origin policy is a security restriction imposed on the client scripts (like JavaScript) that operate in a Web application. The policy restricts the use of HTTP (HyperText Transport Protocol) remote requests that target domains that are different from the domain that originated the Web page’s interaction. It means that Educational FACILITA would not be able to directly request information from other sources, like Wikipedia or Google Maps.

- **HTTP Proxy approach**: to adapt existing Web applications structure and content, a development alternative, from the Web platform point of view, would be to implement an HTTP Proxy. HTTP Proxy would be able to act in the middle of common Web interaction, between the user and the Web page that activated Educational FACILITA. However, in order to activate the reading assistance, it would be necessary...
to redirect the user to the Web application proxy, literally changing the entire context of the previous Web page, which activated Educational FACILITA. This approach is implemented by Google Translate application\(^{10}\), and might break the Web page’s local scripting functionality, due to Javascript incompatibilities.

For those reasons, Educational FACILITA was implemented on the browser application level, instead of being developed on the common Web applications level. We used the Mozilla Jetpack platform\(^{11}\), and developed Educational FACILITA as an extension for the Firefox Browser\(^{12}\). By choosing the browser level platform of development we also had advantage of the browser features to enhance Educational FACILITA’s design and functionality. Examples of the browser features that are not available on the Web application level include:

- No restriction over HTTP remote requisitions
- Better control over browser persistent data (allowing the implementation of cache strategies)
- Browser interface components management functions (tabs, notifications, menus).

The use of browser level routines also facilitates the integration of the different text processing modules that are required to Educational FACILITA. Each of the three modules (Section 3.2) is implemented in a different environment, in a way that they cannot necessarily be implemented in Javascript or in the Jetpack Platform. For example, the Lexical Elaboration module was developed in PHP, while the Named Entities recognition runs over Groovy\(^{13}\) in Java language. These requirements lead us to use webservice to deliver the modules functionality together in Educational FACILITA.

From the technological point of view, the designed architecture is illustrated in Figure 5. In Figure 5, we also illustrate how other services and data resources such as Facebook and Google Maps could be integrated in Educational FACILITA.

4. EDUCATIONAL FACILITA PROTOTYPE

The Educational FACILITA prototype that we implemented runs as a Jetpack Extension for the Firefox Browser.

The assistance is made available to end users by means of the following interactions:

1. Activating the Educational FACILITA link, which inserts the Reading Assistance mechanisms in the web page.

2. Browsing the highlighted textual entities, presented by the JetPack feature (Figure 6); the highlighted textual entities could be every difficult or complex word and named entities identified in the text.

3. Clicking on the highlighted words or expressions, that user could be interested in seeing the synonymous (Figure 7) or additional information, as illustrated on Figure 8.

\(^{10}\)http://translate.google.com.br
\(^{11}\)https://jetpack.mozillalabs.com/
\(^{12}\)http://br.mozdev.org/
\(^{13}\)groovy.codehaus.org/
5. FINAL REMARKS

This paper presented the proposal of an online tool with educational purposes. It has not been evaluated so far, although we have indications from literature that this path is worth following. The evaluation is in preparation to be performed in the near future and is expected to be in large scale, with children and adults under alphabetization process.

To the best of our knowledge, the system we propose is unique, as it is evidenced by our section on related work. Its main innovation comes from the final user that we envision, which demands new solutions and the design of more intuitive interaction models. Our evaluation must also consider this at some moment.

The usual lack of computer experience of poor literacy readers may still motivate other future solutions. For instance, we can mention the adaptation of our already adapted content to screen readers, as well as the portability of our text adaptation model to other computational platforms, as tablets, with the drawbacks of possibly having high software reengineering and hardware costs.

Finally, as Educational FACILITA is part of PortSimples – a Text Simplification project – it may also benefit from other modules under development, e.g., a summarization module, which would probably enrich even more the user experience.

In future steps, we envision to have the tools and respective modules, that have been under our concerns, with an improved performance of PLN processing and, simultaneously, to evolve the technical solutions, regarding the technological evolutions in Web according to W3C standards, as well the accessibility guidelines.

We also intend to exploit the semantic of named entities in such way that if “São Paulo” is shown, a map to get a view of this city could be presented to end users, by integrating the Google Maps resource available.

6. ACKNOWLEDGMENTS

Our thanks to FAPESP and Microsoft Research for supporting this work.

7. REFERENCES

metrics based on their attributes. In SIGDOC '08: Proceedings of the 26th annual ACM international conference on Design of communication, pages 73–80, New York, NY, USA, 2008. ACM.


