conTXT: Context-Aware Summarization as an Adaptation Factor in Mobile Devices

Luís Fernando Fortes Garcia, Prof. Dr. José Valdeni de Lima

Universidade Federal do Rio Grande do Sul,
Instituto de Informática
Av. Bento Gonçalves, 9500 - Campus do Vale - Bloco IV
Bairro Agronomia - Porto Alegre - RS -Brazil
CEP 91501-970 Caixa Postal: 15064

Keywords. context-aware summarization, adaptation, mobile computing.

Abstract. This article presents the architecture of a context-sensitive automatic text summarizer that is intended to be used in mobile computing systems. The summarization process proposed is based on contextual information, especially spatial localization, temporality and users' profiles. We have extended the TF-ISF algorithm by including additional levels of relevance, which are represented by contextual words. This solution aims at contributing to the availability of adequate and customized information, at the right time and at the right place.

1 Introduction

The increasing amount of information available today, either by electronic means or computer networks, such as the Internet, makes it almost impossible to manage them properly. In the context of mobile computing we highlight the intrinsic necessities and constraints of such platforms, such as low storage and processing capacities, and size and interaction limitations. Adaptation architectures that can promote the improvement of interaction in mobile devices are then required. One alternative is the use of automatic text summarization as an adaptation factor. The generation of summaries from original documents, besides promoting a more universal access to inherently limited devices, contributes to the reduction of information overload, as a selection of the most relevant content is made prior to exhibition.
The current tendency of extraction techniques for text summarization usually available today is to produce non satisfactory results, once they do not associate features and functionalities nor are concerned with the platform limitations. Functionality is meant to be the possibility of using some device in different scenarios as a conse-

Fernando Fortes Garcia L. and Valdeni de Lima P. (2005).
conTXT: Context-Aware Summarization as an Adaptation Factor in Mobile Devices.
In *Proceedings of the 1st International Workshop on Web Personalisation, Recommender Systems and Intelligent User Interfaces*, pages 14-23
DOI: 10.5220/0001413200140023

quence of the variability of spatial localization and temporality. Limitation is concerned to low storage and processing capacities, and interaction constraints.
A possible relation among the concepts of adaptation, automatic summarization and computing arises from this scenario. Information and characteristics provided by context sensitivity, especially in terms of environmental factors (spatial localization and temporality), lead to the refinement of the summarization process in the scope of mobile computing. In the present article, context-aware summarization is defined as the inclusion of additional values, which are associated with the level of relevance and representativeness of the key-words meaning within their contexts, in the calculus of relevance of the extractive summarization algorithms This contextual meaning is the relation between spatial localization, temporality and user's profile.
The solution proposed is that the context-sensitive summarization process shall be used as an adaptation factor in mobile and wireless devices, by generating text extractions with relevant information for the user's context.
We expect that the extractions obtained with context-sensitive summarizers can be more useful to mobile devices' users for being more adequate both to their context and profile, as well as for effectively taking the intrinsic limitations of these devices into account.

1.1 Contextualization

Context-sensitive summarization can be used in different mobile computing scenarios:

- Visualization of information provided in portals in mobile phones through SMS (Short Message Service). The novelty is that the only information displayed to the user would be the most adequate to his profile and context;
- Information from public emergency services, such as the police, fireguard and health care services. For their urgency character, such information should be adapted to each professional's area or to the context of the current emergency;
- Weather forecast information adapted to the user's context, selecting information from the user's geographic region;
- Medical information, such as summaries of patient's history, tests and prescriptions, depending on the physical location of the healthcare provider and patient and time issues, such as deadlines for tests and medications, besides the user's profile;
- Advertising information directed towards the user's profile and context.

It is important to say that users may have different profiles, according to their preferences, abilities and limitations, as well as they can take part of a number of different contexts in the same day. For example, when someone is physically present at the stock exchange, we could consider natural that he was provided with summaries about the financial market such as currency quoting, taxes and shares; however, when he goes to a shopping mall, the new context changes the summarization focus to other issues concerning products, sales and movies showing that day.

2 Related Works

The adaptation of texts for mobile devices have already been implemented, since the most simple ones, as in (Gomes, 2001) and (Corston-Oliver, 2001), passing by those with some degree of sophistication as in (Oh and Wang), and (Anderson et al, 2001) until some that use complex processes of automatic summarization (Buykkokten et al, 2002) or information associated with the user's profile (McKeown, 2001), (McKeown et al, 2003) and (Muñoz, 2003). These works were selected from literature because they have already been implemented and have common features with the object of the present study.

In (Gomes, 2001) a proposal has been made to allow the access and visualization of web documents in mobile computing devices without content changes. The system is based on navigation through different abstraction levels in an interface. Besides, users can customize the system by selecting the parts of the document that will be visualized in detail. The focus was on the user's interface and on heuristics that make possible to display long documents in size-limited devices, with no damage to content understanding.

(Corston-Oliver, 2001) present another simple approach to the adaptation of texts that must be displayed in small devices. Based on text compaction, techniques may vary from simple manipulation of characters up to sophisticated linguistic processing. The process of reducing texts makes a telegraphic representation of each sentence by excluding some elements. Considering only a shallow syntactic analysis, elements that are not theoretically relevant for the understanding of meaning are excluded. The main goal of reducing the text is to fulfill limitation requirements. This process is repeated for each sentence by a syntactic analyzer, which, first of all, excludes punctuation signals. After that, the process of characters removal include since deletion of vowels from, depending on the idiom (English, French, German or Spanish), changes in substantives like companies names, reduction of days of the week to one or three letters, and others.

Automatic summarization techniques based on knowledge were described by (OH and WANG). They aimed at disassembling documents to display them in mobile computing devices. The process has 5 steps: (i) firstly, each document is classified into a previously defined category; (ii) after that, the document structure is analyzed and decomposed into several paragraphs; (iii) based on paragraphs, the relevant sentences are extracted and paragraphs with key-words are marked; (iv) a table of the document's content is made and, (v) eventually, the method converts the table of the document's contents and paragraphs into a WML document for display.

Buykkokten's proposal (2001 and 2002) for text summarization is implemented in five methods, in which each web page is split in semantic textual units that can be partially displayed. Thus, the user is allowed to explore successive portions of text in different levels, according to his particular needs.

The adaptation proposals developed by (McKeown, 2001), (McKeown et al, 2003) and (Muñoz, 2003) take into account contextual information, especially from the user's profile, when adapting contents to mobile devices. However, summarization is not foreseen as an adaptation factor, it is limited to documents in the medical area and considers the user's profile to generate different summaries, with relevant information for patients and physicians.

3 Context-Aware Summarization as an adaptation factor in mobile devices

After outlining the adaptation mechanisms reported on the literature, we noticed that solutions provided do not consider – or consider only partially – information associated with context. We are aware of the importance of managing context-associated information, once they would make it possible to better adequate summaries to their own characteristics. Associating context-related information to the process of automatic summarization shall increase the representativeness of the summary generated.
The architecture of textual summarization proposed herein adds the context-sensitivity concept to the traditional solutions, something that must be considered essential in the scenery of mobile computing. Context awareness happens by the inclusion of contextual information in the summarization process, like user's profile, spatial localization and temporality.
Context, in the present article, is meant to be those relevant and inherent characteristics of spatial localization, which support mobility and temporality, making possible to describe different scenarios during different periods of time. The user's profile is considered to be relevant information inherent to identification and personal preferences.
Inclusion of such contextual factors in the process of summarization is believed to refine it, adapting content to user's preferences and scenarios, as well it takes into account the limitations of the mobile computing platforms. This would make it possible to use the context-aware summarization as a factor of adaptation in the mobile environment.

Table 1. Contextual information

Contextual information	Refers to	Examples
Spatial localization	"Where"	Office, pub, movie-theater
Temporality	"When"	Morning, afternoon, evening
User's profile	"What"	Sports, Religion, Politics, Education

The architecture herein proposed is considered open, as it can be implemented and/or expanded with different summarization techniques and allows the addition of new contextual information. The implementation and adaptation of algorithms for extractive summarization used for the mapping of contextual information in the field of mobile computing through a relation between information and corresponding keyword.

3.1 Architecture

The architecture proposed comprises the automatic summarization of texts by including the factor relevance in words derived from the user's context (spatial localization

and temporality) and profile; namely contextual-words in the present study. This proposal has emerged from the idea that certain words are very good clues of specific scenarios. Their presence in texts being summarized demonstrates that the text will be important for users with a specific profile, at a specific moment and place. The process we are proposing increments the value of original relevance of traditional methods of extractive summarization by adding relevance values that are defined during the user's interaction taking place in a certain context. As a given word tends to be more significant in certain contexts than others, its value may increase or decrease according to context, which will make it more relevant, or not, in the selection of sentences that compose the summary.

The implementation of contextual-words may be considered an adaptation of the "cue-phrases" concept by (PAICE, 1981). Differences consist of the way where and how they are obtained as well as of the extension of meaning proposed. The implementation of the architecture we are proposing is made through an adaptation proxy responsible for performing the process of context-sensitive summarization, as well as the maintenance of contextual information through contextual-words associated with the user's context and profile. Figure 1 shows the architecture of the adaptation process proposed. It consists of three main modules: (i) User's Profile Management Module; (ii) Context Management Module and (iii) Summarizing Module, which are detailed as it follows:

- User/Mobile Device: represents the mobile device which will display the summarized document. It is responsible for obtaining and supplying information related to spatial localization, temporality and user's identification;
- Documents Repository: represents the set of textual documents that are able to be summarized with the context-awareness approach;
- Data base – User's profile: It stores information associated with the user's profile, including its associated context-words and respective indexes of relevance;
- Data base – Context: It stores information associated to different contexts, their contextual-words associated and respective indexes of relevance multiplication;

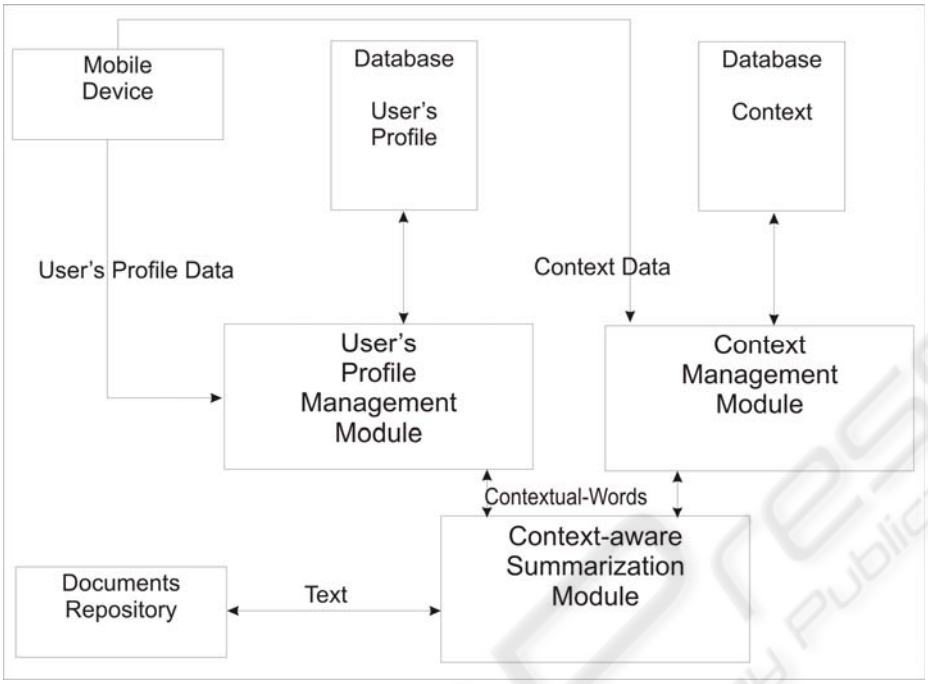

Fig. 1. Architecture of the adaptation Proxy.

- User's profile management module: It is responsible for managing the user's profile, by monitoring the human-computer interaction and/or by receiving explicit data from forms, for example. Profile is meant to be the user's identification and preferences and relevant topics and words. The user's profile refinement can be considered dynamic and is constantly updated by this module. Information on the user's profile are supplied both by the mobile device and its user. They are made available in the user's profiles database, where a user's model is composed of contextual-words related to user-defined topics. An expansion of this model could include, besides favorite subjects, other information as mood, preferences and cognitive data.
- Context Manager Module: Responsible for acquiring, converting and managing information from the user's context, as spatial localization and temporality. Based on this contextual information, this module is responsible for providing the most representative words for the current context. The words are directly associated to context, and consequently to the user's profile.
- Context-aware summarization module: It implements the automatic text summarization improved by the inclusion of the multiplication of indexes of contextual-words supplied by the User's profile and Context Management modules.

The following stages compose the summarization process:

- Document retrieval: the document that will be summarized is provided by the documents repository. The document should be in an only text format (ASCII) with no specific marks;

- Context retrieval: the modules of context and user's profile provide context data and the user's profile and context database make them available;
- Preparing for summarization process – Exclusion of stop words: It consists of removing stop words from the original document. Another process that could be implemented is stemming, however the prototype we have built does not have this functionality;
- Selection of Contextual-words: this process consists of the selection of contextual-words associated both with the user's profile and the user's context;
- Summary generation: it takes place by selecting the most relevant sentences from a text, which are indicated by the context-aware summarization process. Later, it is sent to the mobile device;
- Context-aware summarization: the context-aware summarization process takes place through the application of the TF-ISF algorithm (LAROCCA NETO,2001). This algorithm was adapted for the present study, that is, besides the calculus for relevance of key-words by the original method of TF-ISF, the words have their relevance value multiplied by the indexes of relevance associated, but only if they are in the list of representative contextual-words. The increase in the word relevance value increases in a direct proportion to the index defined for the probability of the sentence selected to compose the final summary. If the word is not in the list of contextual-words, the original relevance value is kept and its original contribution in the sentence selection is preserved.

The TF-ISF algorithm (LAROCCA NETO,2001) calculates the importance of a word w in a sentence s (1):

$$\boxed{TF\_ISF(w,s) = TF(w,s) \times ISF(w)} \quad (1)$$

Let TF(w,s) be the number of times the word w occurs in the sentence s; the inverse frequency ISF(w) is calculated with the formula (2):

$$\boxed{ISF(w) = \log(\frac{tam(s)}{SF(w)})} \quad (2)$$

Let the sentence frequency SF(w) be the number of sentences in which the word w occurs and tam(s) the size of sentence s.
The calculus of key-words relevance through the inclusion of addition relevance indexes of contextual-words is shown in the formula (3):

$$\boxed{TF\_ISFcc(w,s) = TF(w,s) \times ISF(w) \times IP(w) \times IC(w)} \quad (3)$$

TF_ISFca = TF-ISF context aware;
IP(w) = Multiplication Index of the word w given by the profile;
IC(w) = Multiplication Index of the word w given by the context.

3.2 Evaluation

Results obtained with the process of context-aware summarization are assessed (i) in terms of values of recall and precision as compared to reference text excerpts, and (ii) by judging the relevance in terms of adequacy and utility criteria of mobile device's users, in the context and user's profile formally defined.
It is important to note that not all original documents had an associated text excerpt to be used as reference, and even in cases where available, they were conceived in a generic form, that is, they only took into account characteristics of the traditional process of extractive summarization to the detriment of aspects associated both with the user's profile and context. This justifies the need for additional human judging.
The evaluation environment proposed for the validation of the context-aware summarization architecture includes the prototype implemented and the TeMario corpus of texts (PARDO, 2003). For evaluation purposes, the following was also considered:

- High related context: It is a text generated when profile and context are strongly related with the topic approached in the original document, according to a human analysis;
- Average related context: It is a text generated when profile and context are quite related with the topic approached in the original document, according to a human analysis, considering, for example, a suitable profile and a non-representative context;
- Low related context: It is a text generated when profile and context are not adequate to the original text, according to human evaluatio.

The graphic of contextual information recall (Figure 2) shows that coverage values of summaries created without contextual information are usually equal to those created with low related contexts and, in other cases, values are significantly smaller.

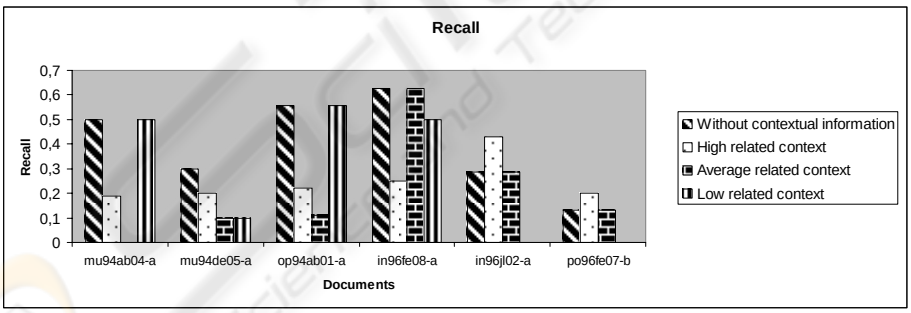

Fig. 2. Recall measures

The precision graphic (Figure 3) shows that the accuracy values of summaries created after highly related context, when available, are higher than those related to other types of contextual summaries.

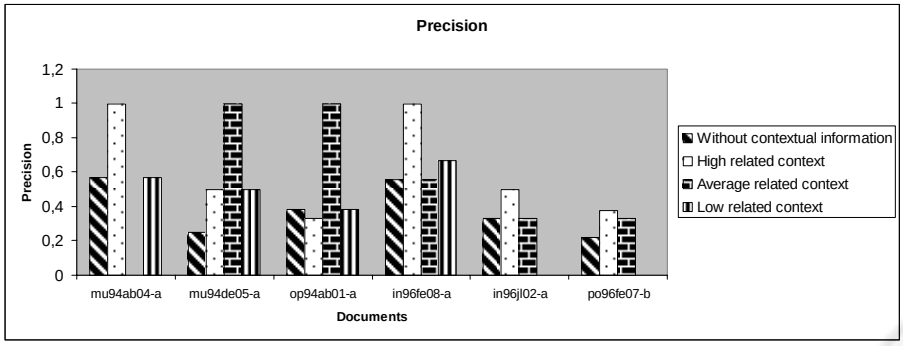

Fig. 3. Precision measures

These graphics show how summaries created with highly related contextual information tend to contain sentences with the contextual-words represented in related profiles and contexts; this tends to demonstrate that they are more representative of original texts. Summaries created with low related context tend to keep the original relevance values calculated by the traditional TF-ISF formula, producing summaries very similar to those created without contextual information, that is, at minimum, the summary produced is equivalent to the one that does not take contextual information into account, this happens when information are not relevant.

4 Conclusions

The large amount of information available in computer networks and the consequent difficulties of managing them have been increasing the necessity of adaptation.
This work proposed a new alternative for adaptation through the use of context-aware summarization. It aims at adapting textual objects by including contextual information available in the scenario of mobile computing into the summarization process.
The first outcomes show that the summarization process supported by contextual information tend to give more accurate and representative – adapted – results in the scenario of mobile computing, confirming our initial hypothesis.
The most important contributions of the solution proposed can be its application as an adaptation factor within the context of mobile computing, the development of an architecture for automatic summarization of texts that can benefit from contextual information for the refinement of the summarization process, producing summaries adequate to the user's profile and context, and the improvement of summarized texts quality. Summaries could be more adequate to the user's profile and context, as well as textual genres and specific domains would have better results. Yet, we may advance to an open architecture that enables the adaptation to other languages, the adoption of new functionalities and methods of summarization; expanding it to other media using sound and image, by implementing corresponding methods of summarization based on contextual information.

References

1. ANDERSON, C.; DOMINGOS, P.; WELD, D. Web Site Personalizers for Mobile Devices. IN: 17th International Joint Conference on Artificial Intelligence. Seattle, Washington, USA, 2001.
2. BUYKKOKTEN, O. et al. Efficient Web Browsing on Handheld Devices Using Page and Form Summarization. ACM Transactions on Information Systems. V.20, N.1, January 2002.
3. CORSTON-OLIVER, S. Text Compaction for Display on Very Small Screens. IN: The Second Meeting of the North American Chapter of the Association for Computational Linguistics. Pittsburgh, USA, 2001.
4. GOMES, P. et al. Web Clipping: Compression Heuristics for Displaying Text on a PDA. IN: Mobile HCI 2001: Third International Workshop on Human Computer Interaction with Mobile Devices. Lille, France, September 2001.
5. LAROCCA NETO, J. Contribuição ao estudo de técnicas para sumarização automática de textos. 2001. 109p. Dissertação (Mestrado em Informática) – Pontifícia Universidade Católica, PUC-PR, Curitiba.
6. LAROCCA NETO, J. et al. Generating Text Summaries Through the Relative Importance of Topics. IN: Simpósio Brasileiro de Inteligência Artificial. Atibaia, SP, 2000.
7. McKeown, K. PERSIVAL, a System for Personalized Search and Summarization over Multimedia Healthcare Information. IN: JCDL´01. Virginia, 2001.
8. McKeown, K.; ELHADAD, N.; HATZIVASSILOGLOU, V. Leveraging a Common Representation for Personalized Search and Summarization in a Medical Digital Library. IN: 2003 Joint Conference on Digital Libraries. Houston, Texas, 2003.
9. MUÑOZ, M. et al. Context-Aware Mobile Communication in Hospitals. IEEE Computer, September 2003.
10. OH, S.; WANG, X. Automatic Adaptation of News Articles to Smaller Screen Device. http://courses.cs.tamu.edu/cpsc689/hohin/fall02/projects/SmartClick/document/CasStudy2.pdf.
11. PAICE, C. The automatic generation of literature abstracts: an approach based on the identification of self-indicating phrases. IN: Information Retrieval Research. 1981.
12. PARDO, T.; RINO, L. TeMário: Um Corpus para Sumarização Automática de Textos. 2003. Relatório Técnico. Universidade de São Carlos. São Carlos/SP.