

A NEW eBOOK CONCEPT AND TECHNOLOGY DEDICATED TO GEOGRAPHICAL INFORMATION

Dan CRISTEA^{*}, Ștefan-Gheorghe PENTIUC^{**}

^{*} "Alexandru-Ioan Cuza" University of Iasi, Faculty of Computer Science, Iasi, Romania (dcristea@info.uaic.ro), ^{**} "Stefan cel Mare" University of Suceava, Faculty of Electrical Engineering and Computer Science, Suceava, Romania (pentiuc@usv.ro)

DOI: 10.19062/2247-3173.2016.18.1.57

Abstract: *The paper presents the concept of MappedBook, the heart of preoccupations and the advances of the project MappingBooks - Jump in the book!. A MappedBook presents a real multi-dimensional interpretation of a book content: textual, graphical, geographical and temporal information, dependence from the localization of the reader and augmented reality, in short, an adequate disclosure of the reader. This kind of eBook "understands" part of the text, recognizes entity mentions with a correspondent in the reality, knows where the reader is and what are the real entities nearby. In the designed technology a MappedBook will be able to draw on the map a route which is described in the book, to upload and use geographical data. The technology is based on a client-server architecture.*

Keywords: *e-book, natural language understanding, mobile computing.*

1. INTRODUCTION

A MappedBook is an interactive electronic book (eBook) connected to the reader, and which could be instantly responsive to the location where the user (reader) is, via a mobile phone or a tablet. This eBook could provides information about events in the real or virtual location, about places' names and other named entities contained in the original book. The system will enrich the existing text links with contextual knowledge for completing readers with additional information beyond the text itself so the eBook becomes an interactive and enjoyable reading.

The main issue will be a 2D or 3D hypermap as described in [1]. These hypermaps, which are in fact multimedia systems referencing Romanian geographical information, can help to increase the visibility of Romanian tourism potential to an international public or to potential customers.

The structured information of a hypermap covers some basic levels with continuous coverage superimposed levels as discrete data, raster and vector, multimedia and hyperlinks.

For the implementation of a Mapped Book, the initial processing of the original text with artificial intelligence tools is essential. The basic text analysis includes at least: recognition of the language, text classification, tokenization, tagging of parts of speech, lemmatization, recognizing noun phrase groups and mentioned entities, sentence-level

segmentation and anaphora resolution. All of them are well known techniques, which have reached technological maturity for Romanian¹.

2. MAPPEDBOOK – A SHORT DESCRIPTION

Mainly a MappedBook is an interactive eBook having the following features:

- it is connected to the reader and the real world
- it is sensitive to the location where the reader is
- incorporates elements of mixed reality, by dynamically setting text over the images captured by the camera of the user's mobile device.

The information provided will be differentiated in time and space.

The eBook let users navigate in multimedia data sets, not only ordered by theme but also spatially. The concept discussed here is supplemented by a prototype developed in an authorware environment.

The platform addresses mainly two categories of readers:

- i) middle school, high school students and students of Romanian universities,
- ii) tourists visiting Romania.

For the first category of users the MappedBooks will be a geographical textbook, and for the second category – touristic guides. The end product for both categories of users represents an innovative interactive technology, for now dedicated only to the Romanian language, but extendable in the future to other languages [2].

For example, if the user reads about a museum and its position happens to be close to the user location, the system will try to find on the web its visitation schedule. If there is sufficient time for a visit, it will notify the user that possibility.

Interactive social gaming features can be imagined, children/students engaged in learning activities or traveling, same as tourists visiting Romania, can be put in relation to each other.

Moreover, by using mixed reality technologies, over the image captured by the mobile device (tablet or smartphone) tourist information about the various objectives sought can be overlapped.

An example is shown in Fig. 1 where a user situated nearby Suceava, looking at the image captured by her phone camera, could see instantaneously in what direction are the main localities situated at a maximum distance of 50 km around her position. If he will turn around, the image captured will be updated with the names of another localities. The system make possible the configuration by the user of the main parameters for vizualization (distance, angles, methods for filtering the points of interest).

This function will be available online or offline (the later, in cases when the phones are not connected to Internet). In the online case, if the user touches a locality on the screen, the server will sent a short description of the locality and other supplementary information if required, in order to be displayed over the image.

The online regime is built as a client-server architecture, where the server will be able to provide web services, feeding with textual and geographical information the clients assumed to be mobiles (smartphones, tablets, laptops).

¹ For more references, see the pages of the NLP-Group@UAIC-FII at <http://nlptools.info.uaic.ro/>.



FIG.1. Information on the point of interest superposed on the image captured with the camera of the mobile device

On the other hand, meaningful linguistic data achieved in the project will be stored in public archives to be useful for future research, that will handle the implementation / development of processing technologies of the Romanian language, as well as the development of electronic dictionaries. An example of this is the extensive list of Romanian geographical names acquired during the project, which will become instances of the Romanian WordNet [3] concepts. Moreover, the structure, content and design of hypermaps will be contextualized. The project has created an inventory of public information and educational tourism in a catalogue metadata space by using geocoded public data available.

3. MAIN MODULES

As an eBook will be based firstly on text (and hypertext) information, the textual data providers will be, generally, publishing houses, which, based on partnership agreements, will offer the primary texts (electronic forms of manuals and travel guides) to the partners in this project. On the server, the texts are subject to a linguistic processing chain followed by a geographical processing and crawling the web for relevant connected information. The overall structure of the system was proposed in [4] and [5] and later developed in [6].

The main modules of the system are the following (see Fig. 2):

1. Text Analytics (TA) – by accessing the NLP services of the NLP-Group&UAIC-FII;
2. Name Entity Recognition (NER) – by using a gazetteer (a large collection of entity names) and a collection of patterns identifying classes;
3. Entity Crawling (EC) – for now, only on a list of predetermined sites;
4. Relations Detection (RD) – namely space relations, based on a list of patterns;
5. Geography (GEO) – exploiting GIS data;
6. Maps & Trajectories (M&T) – for positioning geo-names and journey paths on maps;

7. Augmented Reality – a module running on the client, dealing with orientation of the camera and placing of information in real time;
8. Interfaces – implementing interfaces adequate for different types of devices;
9. Client-Server – responsible for the communication between server and client's device;

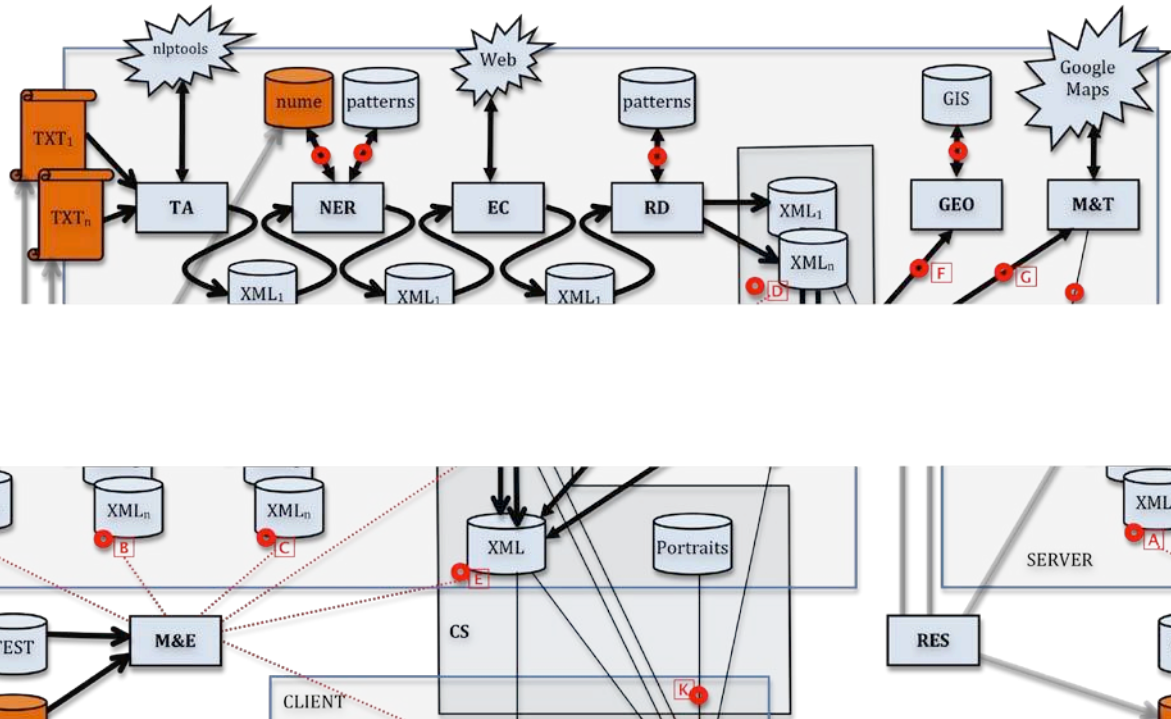


FIG.2. Relations between the main components of the MappingBooks system

MappingBooks will be able to recognize and annotate different types of entities, linking those statements in the book with data from the real and virtual world, and synchronize by applying artificial intelligence techniques [7], [8] for linking references found in the book with actual images captured with the user's mobile device.

CONCLUSIONS

The MappingBooks technology addresses a new concept of an eBook dedicated to geographical subjects in an integrated manner [9], [10]. When ready, the project will be of help to different types of users, from school children to students, and from travellers to people eager to discover touristic places in Romania, in an interactive, attractive manner. The technology processes texts of manuals of geography and travelling guides.

The feasibility of the technology for automatic annotation of texts described above is very much dependent on the maturity of language technologies modules. The more accurate are the annotations obtained automatically, the less expensive the whole production cycle will be for a new book, because manual correction of massive annotations obtained automatically will be avoided.

ACKNOWLEDGMENTS

The paper benefits by the financial support of the project MAPPINGBOOKS - JUMP IN THE BOOK!, in the frame of PARTNERSHIPS - Collaborative Projects Applied Research - Competition 2013 (PCCA 2013). The consortium is composed of "Alexandru-Ioan Cuza" University of Iasi – as coordinator, SC SIVECO Romania SA., and the "Stefan cel Mare" University of Suceava.

REFERENCES

- [1] Kraak, M.-J., Rico, V.D. (1997): *Principles of hypermaps, Computers & Geosciences*, 23(4), pp. 457-464.
- [2] Anechitei, Daniel, Cristea, Dan, Dimosthenis, Ioannidis, Ignat, Eugen, Karagiozov, Diman, Koeva, Svetla, Kopeć, Mateusz, Vertan Cristina (2013): *Summarizing Short Texts Through a Discourse-Centered Approach in a Multilingual Context.*, in Neustein, A., Markowitz, J.A. (eds.), *Where Humans Meet Machines: Innovative Solutions to Knotty Natural Language Problems*. Springer Verlag, Heidelberg/New York.
- [3] Dan Tufiş, Dan Cristea, Sofia Stamou (2004). BalkaNet: Aims, Methods, Results and Perspectives. A General Overview. In *Romanian Journal of Information Science and Technology*, Romanian Academy, Bucharest, Romania, Dan Tufiş (ed.) Special Issue on BalkaNet, July, 7(1-2), ISSN 1453-8245, pages 9–43.
- [4] Dan Cristea, *MappingBooks- Let me jump in the book! Structure proposal*, Communication at The 7th International Conference on Speech Technology and Human-Computer Dialogue "SpeD 2013", Cluj-Napoca, October 16-19, 2013.
- [5] Dan Cristea, Ț Ionu Cristian Pistol (2014). *MappingBooks Jump in the book!* Communication at ConsILR – Craiova, 18-19 September 2014
- [6] Dan Cristea, Eugen Ignat, *Linking Book Characters. Toward A Corpus Encoding Relations Between Entities*, Invited paper: The 7th International Conference on Speech Technology and Human-Computer Dialogue, "SpeD 2013", Cluj-Napoca, October 16-19, 2013.
- [7] Florea A., Pentiu, Gh. Șt., Kayser, D. (2004): *Intelligence artificielle et agents intelligents*, Ed. Printech, Bucharest, ISBN 973-652-976-2, 420 p.
- [8] Dan Cristea, Ionu Cristian Pistol (2014). *MappingBooks: Linguistic Support For Geographical Navigation Systems*. In Mihaela Colhon, Adrian Iftene, Verginica Barbu Mititelu, Dan Cristea, Dan Tufiş (eds.) (2014). *Proceedings of the 10th International Conference "Linguistic Resources And Tools For Processing The Romanian Language, Craiova, 18-19 September 2014"*, „Alexandru Ioan Cuza” University Publishing House, pag. 189-198
- [9] Longley, P. A., Goodchild, M. F., Maguire, D. J. & Rhind, D. W. (2005): *Geographical information systems and science*, p. 511.
- [10] Kraak, M.-J. & Ormeling, F. (2010): *Cartography. Visualization of Spatial Data*, Prentice Hall.

APPLIED
MATHEMATICS,
COMPUTER
SCIENCE, IT&C