

“KAIROS”: A WEB-BASED SYSTEM FOR AUTOMATIC GENERATION OF WEATHER FORECASTS IN TWO LANGUAGES, GREEK-ENGLISH

D. RETALIS &
N. LOURADOS

National Observatory of Athens
Institute for Environmental Research
and Sustainable Development

P.O. Box 20048, 11810 Athens,
Greece

Tel: ++301 3490122

Retalis@env.meteo.noa.gr

S. RETALIS

University of Cyprus
Department of Computer Science
P.O. Box 537, 1678 Nicosia, Cyprus

Tel: ++357 2 892246

retal@cs.ucy.ac.cy

H. PAPAGEORGIOU

Institute for Language and Speech
Processing

Epidavrou & Artemidos 6,

151 25 Athens, Greece

Tel: ++301 6875426

xaris@ilsp.gr

ABSTRACT

Forecast composition is a time consuming task requiring the interaction and the integration of knowledge about meteorology, knowledge about language and knowledge about the conventions used to communicate the right amount of information to meet the needs of diverse user communities. The project “KAIROS” facilitates this work by (semi-)automating the configuration and the generation of forecasts in two languages, English and Greek via a web-based system. Among the major innovative features of “KAIROS” are the following: Generator configuration by means of several resources in which the user specifies different textual plans, the output format of the forecasts, the terminology that will be used as also the concept mappings of the data. Additionally, the modular design and the user intervention functionality of the system enforce two valuable characteristics: reusability and maintainability. In this paper, brief analysis of the architecture of the system under development “KAIROS” is given.

Keywords

Web-based system, Language Generation, XML, metadata, meteorological terminology, weather forecast.

1. INTRODUCTION

With the ever-increasing use of computer technologies in research and development efforts in Meteorology, there is a considerable amount of information stored in digital format. Specialized personnel for whom this information has been stored mainly utilize this information (for statistical analyses, predictions, etc). However, a question arises whether this digital information can be configured in order to meet the needs of diverse user communities. The project “KAIROS”, which is co-funded by the Greek General Secretariat of Research and Technology, is looking for ways to help meteorologists in the automatic composition of weather forecasts from digital meteorological data collections, via a specialized web system. This project adds significant value to the Meteorology Science because weather forecast composition is a time consuming task requiring the interaction and the integration of knowledge about meteorology, knowledge about language and knowledge about the conventions used to communicate the right amount of

information to diverse audience. Because of the “KAIROS” system, the automatic configuration and the generation of weather forecasts is now being made in two languages, English and Greek.

2. THE ARCHITECTURE OF THE “KAIROS” SYSTEM

In Figure 1, the architectural blueprint of the KAIROS system is shown. It is an incremental modularized architecture comprised of:

- The Unit for designing weather forecasts consisting of two major subsystems: the CD (content determination) component responsible for what information should be included in the text and what is minor information that should be eliminated, and the forecast structuring component responsible for designing XML-based forecast structures (based on XML schemata) and the enrichment of these structures with the selected meteorological concepts. These two processes, discourse organisation and text planning are based on a particular user model. For example, in the case of weather summaries, all the information about temperature may be collected into one paragraph and presented before another paragraph which presents all the information about snowfall.
- The Unit for designing structures of sentences which consists of a subsystem for selection of textual units (lexicatisation), and a subsystem for synthesis of textual units into sentences (aggregation).
- The surface linguistic realisation Unit catering for the generation of morphologically and syntactically correct forecasts. The surface realiser deals with all these issues that are effectively non-content aspects of the final sentential form.

In a simplistic way, the weather forecasts are generated as follows: The numerical meteorological data retrieved from the

Data Base are transformed into XML documents of weather forecasts. Then the alphanumerical data become textual data with the aid of the specific sub-language of meteorology. For example, for the wind direction, the numerical data 235° is transformed into “southwest”. The same happens with other values, such as volume, etc.

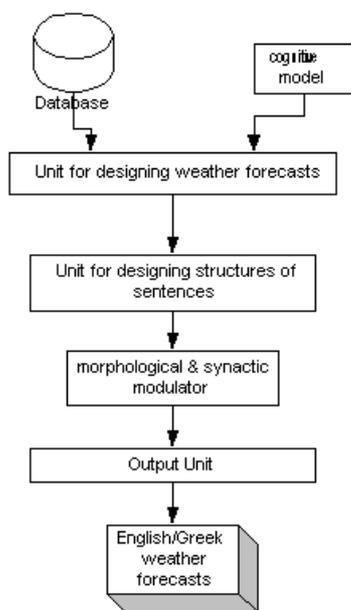


FIGURE 1: KAIROS' architecture

We use a model-driven mapping, which means that a data model is imposed on the structure of the XML document and this is mapped explicitly to the structures in the database and vice versa. What is lost in flexibility (in comparison to Template-Driven Mappings) is gained in simplicity, since the system, which is based on a concrete model, generally does more work for the user. Because the result of transferring data from the database to an XML document follows a single model, XSL is integrated into model-driven products so as to provide the flexibility found in template-driven systems. As a model for data in an XML document, we use a tree of data-specific objects, in which element types generally correspond to objects and content models, attributes, and PCDATA correspond to properties.

The way of generating forecasts into natural language is based on Reiter's pipeline architecture [3], as shown in Figure 2. The outputs are web-based weather forecasts which are easily comprehensible by the laymen.

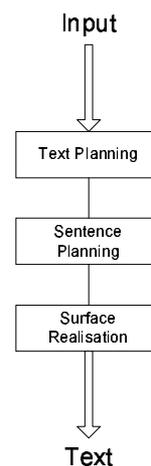


Figure 2: Reiter's pipeline architecture

3. CONCLUDING REMARKS

The project “KAIROS” is still undergoing. The prototype of the system has been formatively evaluated very positively. However, it was found that there was a need for a special (and formal) sub-language on Meteorology for the automatic generation of weather forecasts into natural language as precisely as possible.

For this purpose, a complete catalogue of textual phrases found in a forecast has been created (including their translation into English-Greek). In addition, specific innovation has been made in: a) the development of specific DTDs for weather forecasts, b) the concept mappings of the meteorological data, and c) the output format of the forecasts in a natural language. The KAIROS system can be easily used for automating the configuration and the generation of forecasts in more languages than English and Greek due to its open and modular architecture.

4. ACKNOWLEDGMENTS

The project KAIROS, ref num 99ΕΔ 228, is funded by the EPET program of the General Secretariat of Research and Technology. Co-ordinator is the Institute for Language and Speech Processing. Partners to this project are the National Observatory of Athens, and the Aristotle University of Thessalonica.

5. REFERENCES

- [1] Hodgson, V. 1997: New technologies and learning: Accepting the challenge, *Management Learning*, J. Burgoyne & M. Reynolds (eds), SAGE Publications, London, 1997.
- [2] Reiter, E.; Dale, R. 2000: *Building Natural Language Generation Systems*, Cambridge Univ. Press, 2000.
- [3] Mann, W. C.; Thompson S.A. Rhetorical structure theory: towards a functional theory of text organization, *Text*, 8 (3), pp. 243-281, 1988.