INTELLIGENT HANDLING OF WEATHER FORECASTS

Stephan Kerpedjiev
Institute of Mathematics
Acad. G. Bonchev St., bl.8
1113 Sofia, BULGARIA

Veska Noncheva
Laboratory of Applied Mathematics
N. Vapzarov St., 15
4000 Plovdiv, BULGARIA

ABSTRACT

Some typical cases of intelligent handling of weather forecasts such as translation, visualization, etc. are decomposed into two subprocesses -- analysis and synthesis. Specific techniques are presented for analysis and synthesis of weather forecast texts as well as for generation of weather maps. These techniques deal with the weather forecasts at different levels -- syntactic, discourse and semantic. They are based on a conceptual model underlying weather forecasts as well as on formal descriptions of the means of expression used in particular natural and cartographic sublanguages.

1. INTRODUCTION

Weather forecasts (WF) are the subject of various manipulations. Evolving as sets of numerical data and qualitative estimations they grow into complete images of expected meteorological situations represented in the form of texts in NL, weather maps, data tables or combined information objects. Then the WF could be translated from one language to another, transformed from verbal to pictorial form or vice versa, etc. Such kind of manipulations are often performed by forecasters, translators, editors in order to obtain better representation of the WF from the point of view of its perception and understanding.

There is some experience with such kind of automatic handling of different classes of specialized texts. The METEO system [7] translates WFs from English to French. It is developed on the basis of the TAUM system. A technique for creating computer animation scenarios of storie is described in [6]. The SOCCER system [1] interprets video recordings of soccer games and comments on them in NL.

The purpose of the present project is to develop techniques for various manipulations of WFs including translation from one natural sublanguage into another, transformation from textual to cartographic form, retelling WFs in the same NL but according to other discourse structure. Each of these manipulations can be decomposed into analysis of the source WF and synthesis of the target WF. These two processes are mediated by a uniform internal representation which is language independent. The example in Fig.1 illustrates the basic processes and objects involved in the project.

2.CONCEPTUAL MODEL

Our conceptual model was created on the basis of both conceptual analysis of WFs from mass media and the recommendations for formulating WFs given in [2].

In our conceptual model each WF is a set of assertions having the form
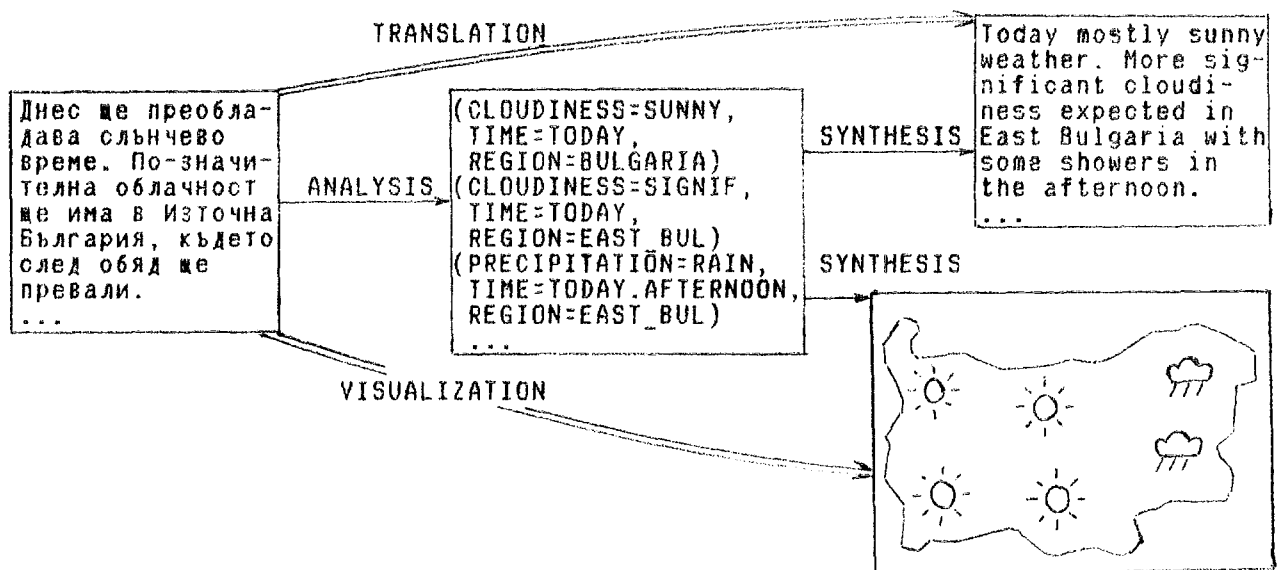
$$(W , T , R).$$

Fig.1. Different types of transformation can be represented as a composition of analysis and synthesis

It is read as "Weather characteristics W are expected in the region R during the period T". For each of the items W, T, R a taxonomy is created that defines the quantities used together with their scales. Each taxonomy represents a particular submodel. The present model does not reflect the degree of uncertainty in WFs.

The internal representation of any WF is a consistent set S of assertions formulated in terms of the conceptual model.

3. TEXT ANALYSIS

Analysis is the process of transforming the text of a given WF into its internal representation. It consists of the following steps:
- data extraction -- identifying the strings in the text that represent the values of the assertion items (such strings are called fillers of the corresponding quantity);
- data translation -- representing the fillers in terms of the conceptual model;
- ellipsis resolution -- completing elliptical phrases with information extracted from their contexts;
- inconsistency elimination -- replacing contradictory assertions by non-contradictory ones.

3.1. Data extraction

The technique for data extraction is based on a formal description of the typical phrases of WFs by means of transition networks called analyzers [3]. The conditions along the arcs of the network are checks for certain semantic primitives defined in a dictionary or comparisons with certain strings. Some of the nodes are marked by two types of symbols $mb(x)$ and $me(x)$, where x is the identifier of a quantity in the conceptual model, mb means that a filler of x begins, me marks the end of the filler of x.

3.2. Data translation

Those entries of the dictionary that may occur as fillers have sections where the meaning of the filler is defined by the corresponding term of the conceptual model. Thus, the data translation process consists of looking up in the dictionary for the current filler, picking out the contents of the corresponding section and replacing the filler by the term extracted. Numerical values such as temperature degrees are calculated automatically from the text.

3.3. Ellipsis resolution

A good deal of sentences in WFs are elliptical. For example, consider the WF in Fig.1, where the region is missing from the first assertion and the time period is not specified in the second assertion. In order to complement them a context analysis is performed. It is based on certain heuristics represented as rules. For the particular case of medium-range forecasts in Bulgarian such rules are given in [4] where four variables are used, namely a short-term and a long-term backgrounds of both the time-period and the region. The rules specify the conditions under which these variables change their values as well as the variable whose value should fill in the current elliptical assertion.

3.4. Inconsistency elimination

Two assertions (W_1,T_1,R_1) and (W_2,T_2,R_2) are said to be contradictory if they predict inconsistent weather characteristics (e.g. sunny weather and overcast) for overlapping regions and time periods. The following heuristic is applied to eliminating contradictory assertions.

The more specifically the information is given, the higher priority it has.

In terms of the conceptual model it means that if W_1 and W_2 are inconsistent, $R_1 \subset R_2$ and $T_1 \subset T_2$ then the second assertion is replaced by the assertion (W_2,T_2-T_1,R_2-R_1), where T_2-T_1 and R_2-R_1 are calculated on the basis of the corresponding submodels.

4. TEXT SYNTHESIS

Synthesis is the process of transforming the internal representation of a given WF into a text in a certain NL and according to a certain scheme (discourse structure). It consists of the following stages:
- generation of words and phrases representing the items of the assertions;
- generation of sentences describing the assertions;
- generation of text fragments containing all the weather information related to a certain region, time period or weather phenomenon depending on the scheme adopted.

The three generators are called phrase generator, sentence generator and discourse generator, respectively. They are presented in detail in [5] for the case of short-range forecasts in Bulgarian.

4.1. Discourse generator

The process of generation is a top-down one. First, the assertions are arranged according to the discourse structure, defined preliminarily in terms of the conceptual model. For example, one feasible discourse structure is the requirement to present the WF in four sections containing information for the regions North, East, South and West Bulgaria, respectively. In each section the information should be arranged according to weather phenomena in the following order: cloudiness, precipitation, wind and

temperature. The discourse generator scans the internal representation and sorts out the assertions into the predefined sections, whereby some assertions (e.g. those referring to the whole country) may fall into more than one section. Inside each section the assertions are reordered by their weather characteristics. Other discourse structures can be processed as well.

4.2. Sentence generator

The generation of simple sentences representing single assertions and compound sentences representing groups of assertions is performed according to certain rules. The simple sentence structure is defined by means of patterns. The choice of the actual pattern is made on the basis of the values of the current assertion. Compound sentences are used when certain patterns of assertions occur in the final form of the internal representation. Six specific rules for constructing compound sentences are defined, one of them is shown below:

$$(W_1, T_1, R_1 = BUL) \implies W_1 T_1 \text{ "when" } R_2 W_2$$
$$(W_2, T_1, R_2 \subset R_1)$$

4.3. Phrase generator

After the sentence structure has been settled, the phrase generator is invoked to form the phrases describing the assertion items. It replaces the terms of the internal representation by corresponding phrases in NL. For this purpose a reversed dictionary is used.

5. MAP GENERATION

Map generation is the process of transforming the internal representation of a given WF into a weather map. WFs are expressed on the map by means of stylized depictions as pictograms and symbols. The form and possibly the color of the depictions reflect the characteristics of certain weather phenomena while the positions of the depictions on the map correspond to the regions where these phenomena are expected to happen. There are no established rules for representing time on the map. The most simple solution from such a point of view consists of creating different maps for each of the periods of a given scale. More sophisticated solutions provide with specific means of time expression such as animation, symbolic representation of processes, explicit time notation of the phenomena being represented, etc.

The technique for map generation employed in this project consists of two steps:
- time analysis of the internal representation resulting in a proper division of the assertions into groups according to the model adopted for time representation;

- scanning the assertions of each group and translating them into depictions corresponding to the weather items and arranged in some positions on the map depending on the regions.

The arrangement is based on a preliminarily designed template of the map. The template contains the spots where depictions could potentially appear. Together with the template a mapping M is defined. For any pair (W,R) M determines the set of spots where the corresponding depiction is to be placed. The depictions used are prepared in advance through a specialized graphical editor.

6. CONCLUSION

The three techniques presented here have been implemented on an IBM PC/AT microcomputer with EGA monitor for the case of medium-range WFs for Bulgaria issued by the *Office of Hydrology and Meteorology* in Sofia.

The technique for text analysis was estimated by analyzing 100 WFs from Bulgarian newspapers. About 85-90% of the data items were extracted and translated correctly by the system. The subsystems for text synthesis and map generation produce WFs of acceptable quality.

R E F E R E N C E S

1. E. Andre, G. Herzog and Th. Rist, On the Simultaneous Interpretation of Real World Image Sequences and their Natural Language Description: the System SOCCER, Proc. ECAI 88, p.449.
2. E. M. Dobryshman, Review of Forecast Verification Techniques, World Meteorological Organization, Technical Report No 120, Geneva, 1972.
3. S. Kerpedjiev, A Technique of Automatic Data Extraction from Specialized Texts, SERDICA Bulg. math. public. 13 (1987), p. 239.
4. S. Kerpedjiev. Transformation of Weather Forecasts from Textual to Cartographic Form, Computing Techniques in Physics, Proc. 8th European Summer Schools,1989, p. 212.
5. V. Noncheva, Synthesis of a weather forecast text. Mathematics and Education in Mathematics, Proc. 19th Conf. of Union of Bulg. Math.,1990, p.269.
6. H.Shimazu, Y.Takashima and M. Tomono, Understanding of Stories for Animation, Proc. Coling 88, p. 620.
7. B.Thouin, The METEO System, in: Practical Experience of Machine Translation, ed. V. Lawson,p.39.