TAMIC-P: A SYSTEM FOR NL ACCESS TO SOCIAL INSURANCE DATABASES

Johannes Matiasek, Alexandra Klein, Harald Trost[1]

The TAMIC-P project, which is funded by the European Union, provides access to heterogeneous database systems for social insurance information. The system uses a card-index metaphor to structure the interface and the presentation of information. Cards allow for presenting personal data, legal texts, help texts and domain knowledge simultaneously. The cards itself and the entities displayed on them can be accessed via natural-language queries.

1. Introduction

Accessing heterogeneous databases (as often to be found in Public Administration, [1]) poses considerable problems to non-expert users for various reasons:

- both the location and the field name of the required information may not be known to the user
- the contents of a retrieved field may be coded in a way the user is unable to interpret
- the presentation of the information on the screen may be in such a way that the user finds it difficult to locate it.

The TAMIC-P system which will be described here is an attempt to overcome all these problems. The development of the system is supported by the European Community. It is intended for use with various languages. Currently, an Austrian (German) and an Italian implementation exist which provide an interface to information stored in social insurance databases. For both applications, the main functionality and the structure of the interface are identical, while of course the language-

---

[1] Austrian Research Institute for Artificial Intelligence, A-1010 Wien, Schottengasse 3
`{john|alexandra|harald}@ai.univie.ac.at`

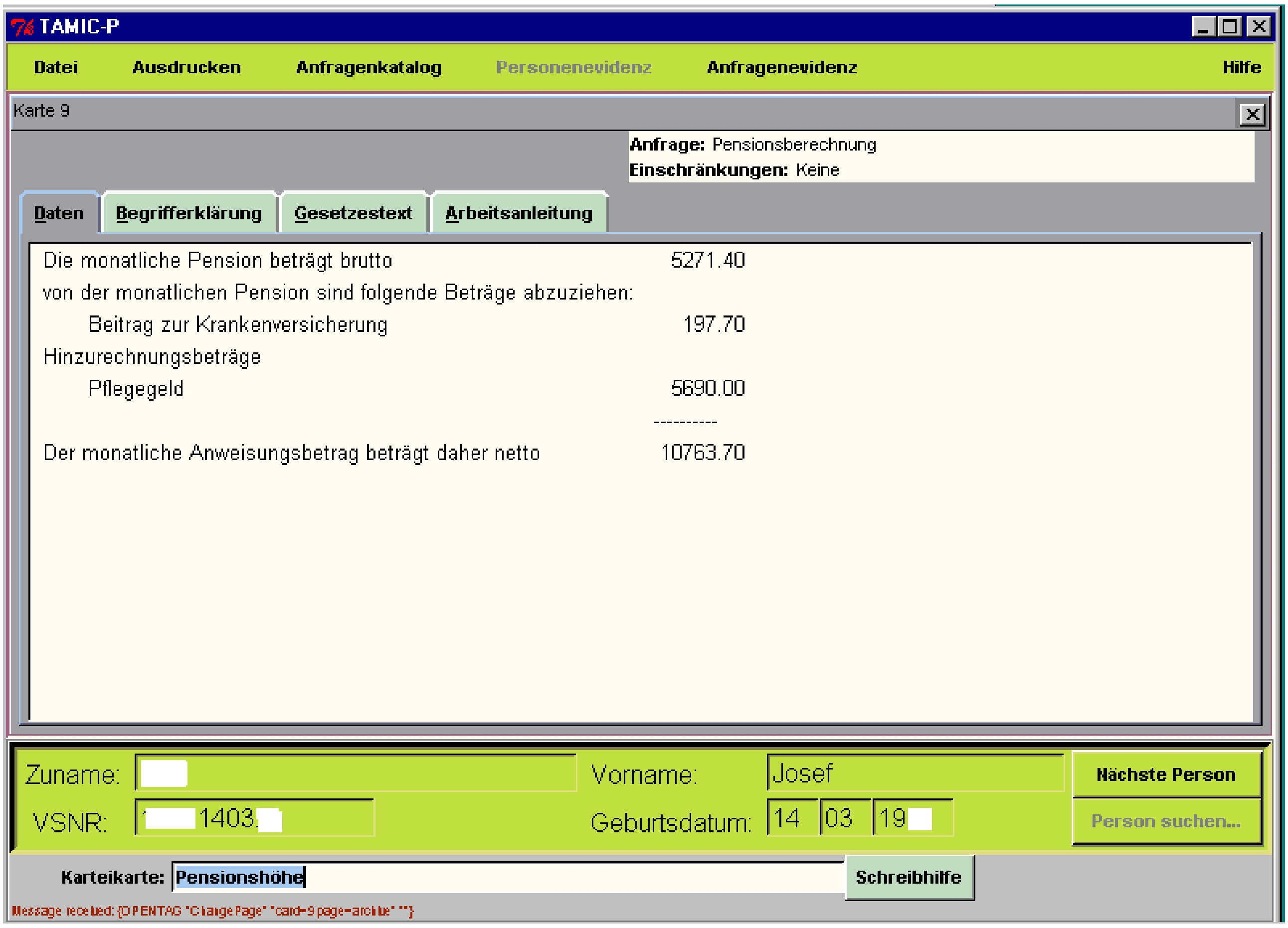
specific parts of the interface and the analysis mechanisms are completely different. There are also some domain-specific differences which had to be considered in the system. The user of the system for the Austrian application [2] is the Sozialversicherungsanstalt der Bauern (SVB, Social Insurance Institution for Farmers) which incorporates a pension insurance, a health insurance and an accident insurance system for citizens from the agricultural sector. Therefore, the SVB needs to use various heterogeneous databases related to contributions as well as pension, health and accident insurance. The domain is very complex, and so are the databases; consequently, only expert users who are familiar with the domain and the way the information is stored are able to issue requests.

2. System Functionality

Natural language processing plays an important role in the use of interface functionality as it can be used to translate the concepts the user is thinking of into the database access query and to decode the value found in the database. Thus, natural language access is the means to overcome the access barrier mentioned above. However, to enable the system to access the data it needs a unified model of the available information. The TAMIC-P system gathers distributed databases into one local data model, which is loaded whenever information concerning a specific person is requested. This local database can then be searched for a number of concepts which are common in information requests.

Apart from database access, the system provides additional knowledge sources depending on the concept contained in the query. Legal texts from the social insurance domain, help texts for internal use at the SVB, and explanations of the domain are provided which are relevant for the queries.

In order to come up with a perspicuous presentation, the interface simulates a card index [5] to provide the user with a familiar organization of the data. If a query has been posed, the user can view the corresponding passages from legal texts, from help texts and from the domain model representation by simple point-and-click operations on the index cards. Information requests in natural language are issued by the user by simply naming the card she wants to see (frequently used card names can be accessed via a pull-down menu to avoid the need for typing, see Fig. 1 for a screen-shot of the interface).



Anfrage: Pensionsberechnung
Einschränkungen: Keine

Die monatliche Pension beträgt brutto	5271.40
von der monatlichen Pension sind folgende Beträge abzuziehen:	
Beitrag zur Krankenversicherung	197.70
Hinzurechnungsbeträge	
Pflegegeld	5690.00

Der monatliche Anweisungsbetrag beträgt daher netto	10763.70

Zuname: [redacted]
Vorname: Josef
VSNR: 1[redacted]1403[redacted]
Geburtsdatum: 14 03 19[redacted]

Karteikarte: Pensionshöhe

Message received: {OPENTAG "ChangePage" "card=9 page=archive" ""}

Fig. 1: Screen shot of the Tamic-P interface

Thus, differently as in other NL database interfaces (such as DB-DIALOG [4]), queries to the system are posed as complex NPs, consisting of clusters of NPs and PPs, and they refer to the entities represented in the database. Some examples for queries are:

- *Ausgleichszulage* (‘compensatory supplement’)
- *landwirtschaftliches Pauschale* (‘agricultural income’)
- *Beitragsmonate vor 1980* (‘contribution months before 1980’)
- *Versicherungsmonate als Betriebsführer nach 1960* (‘insurance months as head of business after 1960’)
- etc.

3. Understanding Natural-Language Queries

The system has to analyze these queries in order to come up with a card filled with the data instances of the concept in the database the queries refer to. The following knowledge sources are used for this analysis and understanding process:

- Grammatical knowledge bases
- a morphological lexicon which is used for the analysis of the word forms
- a grammar for complex German nominal phrases employing context-free phrase-structure rules augmented with feature structures for the handling of subcategorization and semantic restrictions
- a lexical knowledge base built around a Wordnet-style hierarchy [3] of synsets which provides the links to the semantic concepts of the conceptual data model and to the relevant pieces of textual information.
- scenario-specific knowledge bases
- the conceptual data model consisting of a hierarchy of the concepts which occur in the databases and their relations
- the logical data model describing the unified view of the data of the domain

A unification-based, memoized parser derives a representation in a quasi-logical form which is constructed from the rules in the grammar and from the entries in the lexical knowledge base. This compositional approach has to be slightly modified in cases where simple entities and relations in the databases, which are represented as objects, attributes or attribute values, are being referred to by complex nominal phrases. For these cases, a filter mechanism was introduced which applies a set of substitutions to the quasi-logical form, transforming the complex description into the corresponding simplistic one contained in the conceptual data model. The quasi-logical-form expression which is obtained from the parse and possibly from an application of the filter mechanism is then mapped onto the appropriate data access query.

4. Conclusion

In this paper, we have presented a system for access to heterogeneous databases. Users are able to address entities in the various databases by means of a coherent interface. The interface simulates a card index which simultaneously provides information concerning personal records, legal texts, help texts and domain knowledge. Entities in the databases as well as the texts can be accessed by natural-language queries. For the insurance clerks, this type of access facilitates their tasks as they no longer need knowledge about the structure of the databases and the way the information is coded. A first prototype of the system has been implemented and is currently being evaluated by the SVB clerks.

5. References

[1] Bagnasco, Clara, Paolo Bresciani, Bernardo Magnini, Carlo Strappavara. „Natural Language Interpretation for Public Administration Database Querying in the TAMIC Demonstrator“. In: R.P. van de Riet, J.F.M. Burg, A.J. van der Vos (eds*), Applications of Natural Language to Information Systems*. Amsterdam: IOS Press, 1996.

[2] Klein, Alexandra, Johannes Matiasek, Harald Trost. „The Treatment of Noun Phrase Queries in a Natural Language Database Access System“. In: Busa F. et al.(ed), *Proceedings of the COLING-ACL98 workshop 'The Computational Treatment of Nominals'*, Universite de Montreal, Canada, 1998, pp. 39-45.

[3] Miller, George A., Richard Beckwith, Christiane Fellbaum, Derek Gross, Katherine J. Miller. `Introduction to WordNet: an on-line lexical database.' In*: International Journal of Lexicography* 3 (4), 1990, pp. 235 - 244.

[4] Trost, Harald, Ernst Buchberger, Wolfgang Heinz, Christian Hoertnagel, Johannes Matiasek. „'Datenbank-DIALOG' - A German Language Interface for Relational Databases“. In*: Applied Artificial Intelligence* 1.2, 1987, pp. 181-203.

[5] Zancanaro, Massimo. *WP7: Interface and Utilities - Intermediate Synthesis Report*.TAMIC-P Technical Report. February 1998.

Acknowledgement:

Research in the TAMIC-P project is funded by the European community under contract no. LE-4253. Financial support for ÖFAI is provided by the Austrian Federal Ministry of Science and Transport.