PERSONALIZED QUERY GENERATION IN RELATIONAL DATABASE SYSTEMS

Silvia Noemí Schiaffino

Prof. Dra. Analía Amandi
Advisor

Thesis report submitted to evaluation as partial fulfillment of the Master degree requirements

Tandil, March 2001
Abstract

The volume and variety of information distributed over distant sites make the manipulation of data a progressively more complex process. A big part of these data is recorded in relational databases, which are manipulated by information systems through query languages. In addition, with the growing of the WWW, several companies are leveraging these systems in this context. Then, managing this information is becoming a big challenge both for companies and for users of these database systems.

Considering users’ point of view, querying a database through the Internet or an Intranet may become a hard task. The problem with these systems is that users generally have difficulties in getting the information they need in a timely manner and they have to wait for it more than they would want to. Besides, the tasks they perform with these systems are generally repetitive and routine as users always query about the information that is relevant for them. In consequence, users spend their time performing repetitive and time-consuming tasks, instead of working productively with these systems.

In this work we propose a solution to this problem using intelligent agents. Intelligent agents are a quite new alternative to assist users with repetitive tasks, as they can learn the ways users prefer to perform those tasks. This work presents QueryGuesser, an intelligent agent developed to assist users who frequently query a database system in order to get the information they need. The QueryGuesser agent has the capability of managing personalized queries to a database system according to the users’ information needs, habits and preferences. This agent observes a user's behavior while he is querying the database and it builds a model of his preferences. QueryGuesser agent uses a technique that integrates two Machine Learning techniques, Case-Based Reasoning and Bayesian Networks, to acquire knowledge about users and to build then user profiles. This technique was also developed as part of this work. Users' profiles are used to handle users' most relevant queries in advance and offer them the information they need when they enter into the system.
Acknowledgements

I would like to take this opportunity to thank all those who have contributed to this master thesis, directly or indirectly.

I would first like to express my sincere gratitude to my advisor Analía Amandi for her constructive criticism, guidance and support. I would also like to thank her and Marcelo Campo for providing me the appropriate environment for research.

I would like to thank Alejandro, Andrés and Daniela for providing me with valuable comments and reading material, and for giving me their opinion on topics related to this thesis.

I also owe my gratitude to all the members of ISISTAN for helping make work at this institute a pleasure.

I would like to thank CONICET for providing financial support for the work in this thesis.

I would also like to thank Mario Zito and Analyte - Lab Information Technologies for providing us information to test the QueryGuesser agent in a LIMS application domain.

Thanks to Alicia and Santiago Bargas for helping me with my English handwriting.

I am deeply indebted to my parents for their encouragement, support and sacrifices without which I certainly would not have reached where I have. I would especially like to thank Diego for his love and his understanding during these last five months.

Silvia
# List of Figures

1.1 – General Schema of QueryGuesser’s functionality      4  
2.1 – How interface agents work         8  
2.2 – Some existing interface agents        11  
2.3 – Summary of content and construction of profiles     13  
3.1 – External view of a QueryGuesser agent       16  
3.2 – Pseudo-code of QueryGuesser’s main tasks      16  
3.3 – Information recorded during observation process  18  
3.4 – Similar queries                          18  
3.5 – Temporal ranges                          19  
3.6 – Bayesian model of a user’s interests        19  
3.7 – User profile: topics of interest           20  
3.8 – General schema of a QueryGuesser agent      21  
3.9 – The QueryGuesser agent’s capabilities      21  
4.1 – CBR process                             25  
4.2 – QueryGuesser’s case representation        26  
4.3 – A case representing a particular query     27  
4.4 – Ranges used to detect the user’s routine   28  
4.5 – Case representation taking into account a user’s routine  28  
4.6 – Ranges for work shifts                     29  
4.7 – Implementation of cases in Prolog           30  
4.8 – Finding correspondences between cases       30  
4.9 – Importance values of dimensions and matching rules  31  
4.10 – Portion of matching algorithm            32  
4.11 – Importance values in situation matching    32  
4.12 – Matching algorithm to determine week range similarity  32  
4.13 – Matching algorithm to determine day range similarity  33  
4.14 – Matching algorithm to determine month range similarity  33  
4.15 – A numerical evaluation function           33  
4.16 – Matching algorithm to determine information needs similarity  34  
4.17 – Matching algorithm to determine temporal data similarity  34  
4.18 – Clusters representing different topics of interest  35  
4.19 – Comparing querying situations            36  
5.1 – Example of a BN                           40  
5.2 – Attributes used as filters in a user query   42  
5.3 – BN representing John Smith’s query space model  43  
5.4 – Conditional Probability Tables belonging to John Smith’s Bayesian model  43  
5.5 – Attributes used as filters in John Smith’s new query  43  
5.6 – BN representing John Smith’s new query space model  44  
5.7 – Conditional Probability Tables belonging to John Smith’s Bayesian model  44  
5.8 – QueryGuesser’s case representation        45  
5.9 – BN including labels                       45  
5.10 – Example of a BN                          46  
6.1 – A user’s profile                          49  
6.2 – Related Systems                           51  
6.3 – Technique overview                        51  
6.4 – Pseudo-code of the user profile building algorithm  52  
6.5 – Bayesian model of a user’s topics of interest  53  
6.6 – CBR representation of a user’s topics of interest  53  
6.7 – Example of a BN                           54  
6.8 – Bayesian model of John Smith's preferences  56
List of Tables

- Table 5.1 – Attributes and attribute values   41
- Table 5.2 – Dependencies between attributes  42
- Table 5.3 – Dependencies between values      42
- Table 6.1 – Feedback provided by the user    62
- Table 8.1 – CPT example in a LIMS domain     81