

Integrating Heterogeneous Tourism Information in TIScover– The MIRO-Web Approach

M. Haller*, B. Proell*, W. Retschitzegger**, A M. Tjoa*, R.R. Wagner*

*Research Institute for Applied Knowledge Processing (FAW)
Hauptstraße 99, A-4232 Hagenberg, AUSTRIA
Tel: +43-7236-3343-770; Fax: +43-7236-3343-782
email: {mhaller, bproell, tjoa, rwagner}@faw.uni-linz.ac.at

**Institute of Applied Computer Science, Department of Information Systems (IFS)
Altenbergerstraße 69, A-4040 Linz, University of Linz, AUSTRIA
Tel: +43-732-2468-883; Fax: +43-732-2468-9511
email: werner@ifs.uni-linz.ac.at

1 Introduction

With the tremendous growth of the web, accessing information on the Internet has become less a question of determining whether the information is out there, but rather, in what form, and how to find it [Flor98]. This situation is especially true for the tourism industry where a broad spectrum of tourism information is already distributed over various web sites and stored using heterogeneous formats. It is obvious that this situation is very undesirable since the tourist is burdened with finding and visiting various web sites in order to gather all the desired tourism information and products. The situation gets even worse since web sites usually differ very much especially concerning information presentation and information access. What is therefore required is that the tourist is enabled to collect all desired tourism information and tourism products at one place, no matter whether, e.g., information about weather and traffic conditions, schedules of trains, planes and buses, or information about actual events like movies at the local cinema including movie reviews as published in different online newspapers are required.

To fulfill this requirement it is of course not feasible to store every kind of information a tourist might be interested in at one web site neither in terms of storage costs nor and even more important in terms of maintenance overhead. A more promising idea would be to take advantage of the huge amount of other relevant tourism information that is already distributed all over the web. In the course of the *ESPRIT project MIRO-Web* which was started at the end of 1997, the official Austrian tourism information and booking system TIScover is extended in order to federate multiple structured and non-structured tourism information sources on the Web [Fran98]. In particular, TIScover should be able to integrate *fully unstructured ASCII* data, *semi-structured data* in terms of HTML-pages *Excel-files* for managerial purposes and finally *different in-house databases* employed in the various tourism offices [Ebne98].

The rest of the paper is organized as follows: After a short overview about TIScover in Section 2, the basic concepts and technologies of MIRO-Web used in order to integrate heterogeneous information sources are presented in Section 3. Section 4 demonstrates the applicability of MIRO-Web to TIScover by discussing three different application scenarios comprising an *Event Agent*, a *Winter Sports Agent* and a *Golf Agent*. Section 5 concludes the paper by pointing to future work.

2 An Overview of TIScover

The development of TIScover has been started in 1996 based on the experiences made with the pioneering system TIS@WEB [Burg97]. The aim of TIScover is twofold [Proe98ab], [Proe99a]: first, tourists should be *supplied with comprehensive, accurate and up-to-date tourism information* on countries, regions, villages and all destination facilities they offer like hotels, museums or other places worth seeing. Second, it aims to *attract the tourist to buy certain tourism products* either offline or even more important to allow the tourist to buy them *online*. Originally, TIScover was realized to market the facilities of a certain region of Austria, namely Tyrol, only. Meanwhile, four other Austrian regions have joined TIScover [TISc99a]. Besides that, TIScover has been employed in Asia, presenting tourism information about Thailand [TISc99b]¹, it is currently used by the German company START Media Plus, a major player in the area of online reservation systems, to present tourism information about Germany [TISc99c], and it has just gone online in Switzerland as KISSswiss, employed by the companies Kümmerly+Frey and Basler Versicherungen [TISc99d].

The functionality provided by TIScover can roughly be categorized into three different components, the *public Internet* component, the *Extranet* and the *Intranet* [Proe98b]. The *public Internet* component comprises that functionality of the system that is accessible to the public, whereby the most important modules are *Atlas* and *Booking*. The module *Atlas* allows the customer to browse through all kinds of tourism information by navigating through a geographical hierarchy and to use a *full text search*. The module *Booking* allows for a *precise structured search* based on a subset of the tourism information presented by *Atlas*, like villages, hotels, available rooms, events and camping sites along the geographical hierarchy as well as *online booking* of these tourism products. Furthermore, TIScover provides an *Extranet* allowing authorized tourism information providers, no matter being a small guesthouse or a large local tourist office to update and extend their tourism information and products directly. Finally, the *Intranet* component of TIScover which is accessible at the system provider's side only allows to configure the whole system in various ways. It is, for example, possible to extend the geographical hierarchy, to specify expiration dates for reports and to define the default language for all system components.

Currently, TIScover stores all tourism information and tourism products within a central relational database. The common database schema of TIScover Austria consists of about 400 database tables and has been constructed on the basis of a domain data model which incorporates all conceptual entities gathered during the process of requirements definition with numerous tourism information providers and from the experiences with the predecessor system TIS@WEB [Burg97]. The database of TIScover Austria comprises about two gigabyte of data. To facilitate performant access, web pages are automatically generated out of the database every time one of the 4.500 tourism information providers ranging from hotels to local tourism offices maintains the content (cf. Section 2) covering among others 1.500 towns and villages and 20.000 accommodations [Proe99b]. As a result, there exist more than 400.000 web pages stored in about one million files. Per month, the system has to handle up to 5 million pageviews, 1.2 million visits as well as up to 20.000 requests for information and online bookings.

3 Basic Concepts and Technologies of MIRO-Web

The integration of heterogeneous data is a challenging problem when trying to utilize existing tourism information on the web. Most of current web sites do not allow for the integration of multiple data sources beyond simple links between them. Thus, the MIRO-Web project [Fran98] focuses on the development of a set of

¹ Note, that due to the economical crisis in asia, TIScoverasia is currently not operating.

middleware components providing integrated and transparent access from standard Web browsers to multiple data sources, ranging from databases to loosely structured files, located on different web sites. MIRO-Web builds on the results of the ESPRIT project IRO-DB [Kaps98]. However, in contrast to conventional database integration as developed in IRO-DB, integration of information sources on the web poses two fundamentally new challenges.

- (1) Tourism information is stored within distinct heterogeneous sources, under different formats. As an extreme we find tourism information coming from databases with a *completely known structure*. At another extreme, we have tourism information which is *fully unstructured*, such as images, sounds, and raw text. But most of the data falls somewhere in between these two extremes, called *semi-structured data* such as HTML-files [Flor98].
- (2) Information sources at the Internet *evolve at a much higher pace than databases* in a controlled business setting.

Thus MIRO-Web in particular focuses on tackling the *large variety of sources* and their *highly dynamic nature*. Interoperability is accomplished in MIRO-Web by means of three layers: *Database adapters* provide for a uniform access to individual databases, a *communication layer* supports remote data access, and an *interoperable layer* allows for integrated views on several databases [Huck98]. For a more detailed description it is referred to the full paper.

4 Applying MIRO-Web to TIScover

To demonstrate the applicability of MIRO-Web to TIScover in the following three different scenarios are presented. In all of these scenarios a *virtual agent* assists the tourist in querying heterogeneous tourism information sources in a transparent way. Such a virtual agent is, e.g., very useful for tourists having only a vague imagination of what they are looking for. The virtual agent uses MIRO-Web as the mediating layer between the heterogeneous sources and TIScover. The scenarios comprise an *event agent*, a *winter sports agent* and a *golf agent*. In the following, only the event agent is described briefly, for a description of the other ones it is again referred to the full paper. A common case is that a tourist wants to find a hotel, which should be near some event locations, such as a musical, an opera, or an action movie. Furthermore, the weather should be fine and the hotel should cost less than 50.000 Lire.

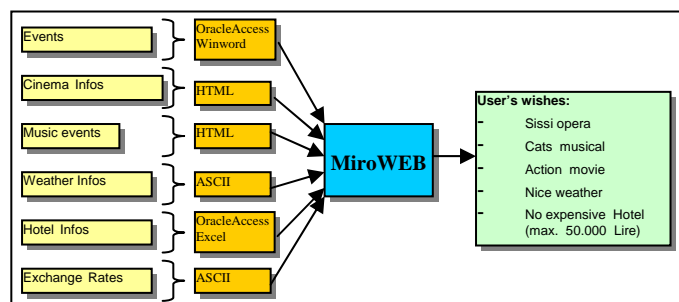


Figure 1: Event Agent Architecture

The different sources for answering this information request can be found on different web sites, stored in various formats (cf. Figure 1). The actual movies, for instance, can be found on a web site, comprising all cinema events of Austria (<http://www.film.at>). The weather information is available via file transfer protocol in form of an ASCII file and finally the exchange rates for calculating the requested foreign currency should be extracted out of an email. The join criteria of the query necessary to satisfy the information request would be the

name of the location. The information about the cinema events, and the music events can be joined with the weather information and the hotel information. Finally the price can be calculated and converted by means of the exchange rates and joined with the hotel information.

5 Conclusion and Future Work

This paper has demonstrated the MIRO-Web approach which provides the technical framework for creating a homogenous view on arbitrary tourism web servers which are semantically similar but realized in a heterogeneous way. As a next step towards electronic commerce, MIRO-Web could be extended towards the integration of several large *international CRS* (Computerized Reservation Systems), e.g. Amadeus, Sabre and Galilei or *local CRS*, such as Gulliver, Bokser and Bettenboerse, so that tourism products and packages can be booked without any human intervention.

References

- [Burg97] F. Burger, P. Kroiss, B. Proell, R. Richtsfeld, H. Sighart, H. Starck, *TIS@WEB - Database Supported Tourist Information on the Web*, Proc. of the Int. Conf. on Information and Communication Technologies in Tourism (ENTER'97), A. Min Tjoa (ed.), Springer, Edinburgh, 1997.
- [Ebne98] A. Ebner, M. Haller, K. Plankensteiner, P. Starzacher, E. Stauder, and A. M. Tjoa, *Tourism Application Requirements Definition (D7B-1/1)*, MIROWEB (EP 25208), April 1998.
- [Flor98] D. Florescu, A. Levy, A. Mendelzon, *Database Techniques for the World Wide Web: A Survey*, ACM SIGMOD Record, Vol. 27, No. 3, Sept. 1998.
- [Fran98] P. Fankhauser, G. Gardarin, M. Lopez, J. Munoz, A. Tomasic, *Experiences in Federating Databases: From IRO-DB to MIRO-Web*, Proc. of the 24th Int. Conference on Very Large Data Bases (VLDB'98), A. Gupta et al. (eds.), New York City, USA, Morgan Kaufmann, Aug. 1998.
- [Huck98] G. Huck, P. Fankhauser, K. Aberer, E. Neuhold, *Jedi: Extracting and Synthesizing Information from the Web*, Proc. of CoopIS'98, IEEE Computer Society Press, 1998.
- [Kaps98] E. Kapsammer, W. Retschitzegger, R. R. Wagner, *Meta Data-Based Middleware for Integrating Information Systems: A Case Study*, Proc. of the 4th Int. Conference on Information Systems Analysis and Synthesis (ISAS'98), Orlando, July 12-16, 1998.
- [Proe98a] B. Proell, W. Retschitzegger, R.R. Wagner, A. Ebner, *Beyond Traditional Tourism Information Systems - TIScover*, Journal of Information Technology in Tourism (ITT), Vol.1, Inaugural Volume, Cognizant Corp., USA, 1998.
- [Proe98b] B. Proell, W. Retschitzegger, R.R. Wagner, *TIScover - A Tourism Information System Based on Extranet and Intranet Technology*, Proc. of the 4th Americas Conf. on Information Systems (AIS'98), Baltimore, Maryland, 1998.
- [Proe99a] B. Proell, W. Retschitzegger, R.R. Wagner, *Holiday Packages on the Web*, Proc. of the Int. Conf. on Information and Communication Technologies in Tourism (ENTER'99), D. Buhalis et al. (eds.), Springer, Innsbruck, 1999.
- [Proe99b] B. Proell, W. Retschitzegger, H. Sighart, H. Starck, *Ready for Prime Time - Pre-Generation of Web Pages in TIScover*, accepted for publication at the 8th Int. ACM Conference on Information and Knowledge Management (CIKM), Kansas City, Missouri, Nov. 2-6, 1999.
- [TISc99a] Homepage of TIScover, <http://www.tiscover.com>, TIS Innsbruck, FAW Hagenberg, 1999.
- [TISc99b] Homepage of TIScover ASIA, <http://www.tiscoverasia.com>, GoThailand, 1999.
- [TISc99c] Homepage of TIScover Germany, <http://www.deutschlandreise.de>, START Media Plus, 1999.
- [TISc99d] Homepage of TIScover Switzerland, <http://www.kissswiss.com>, Kümmerly+Frey, Basler Versicherungen, 1999.