
Designing a knowledge-based tourism information system

Laurence Noël*

Department of Information and Communication Sciences,
Paris 8 University – MSH Paris Nord,
St Denis, France
E-mail: laurence.noel@mondeca.com
*Corresponding author

Olivier Carloni and Nicolas Moreau

Department of Computer Science,
LIRMM, Montpellier, France
E-mail: olivier.carloni@mondeca.com
E-mail: moreau@lirmm.fr

Stéphanie Weiser

Department of Applied Linguistics,
Paris 10 University – MoDyCO,
Nanterre, France
E-mail: sweiser@u-paris10.fr

Abstract: The goal of the Eiffel project is to design a knowledge-based information system that can be used by tourism organisations, both, to manage their resources and to promote them on a web portal. In this paper, we define and analyse two types of processes: first, the integration of data into the knowledge-base (internalisation process) and second, the communication of data from the knowledge base on a web portal (externalisation process). We describe how knowledge acquisition techniques and inference services can be used to alleviate tourism experts' annotation tasks and control the validity of the data imported into the knowledge base. Then, we analyse how semantic data can be used to provide targeted information and interactions so as to improve exploratory search on the web portal.

Keywords: knowledge management; information extraction; conceptual graphs; exploratory search; interface design; semantic web.

Reference to this paper should be made as follows: Noël, L., Carloni, O., Moreau, N. and Weiser, S. (2008) 'Designing a knowledge-based tourism information system', *Int. J. Digital Culture and Electronic Tourism*, Vol. x, No. x, pp.xxx–xxx.

Biographical notes: Laurence Noël is a research engineer at Mondeca, a company specialising in knowledge management systems. She is also a PhD student at the Department of Information Communication Sciences, Paris 8 University, France. Her current research focuses on interface design and semantic web technologies.

Olivier Carloni is a research engineer at Mondeca. He is also a PhD student at the Department of Computer Sciences, LIRMM, Montpellier, France. His current research interest is to provide reasoning capabilities to knowledge management systems using graph theory operations.

Nicolas Moreau is an engineer at the French National Center for Scientific Research (CNRS). He is also a PhD student at the Department of Computer Sciences, LIRMM, Montpellier, France. His current research focuses on knowledge representation and the use of conceptual graphs as a query language.

Stéphanie Weiser is a PhD student at the Department of Applied Linguistics, Paris 10 University, France. Her current research focuses on information extraction and knowledge acquisition techniques applied to spatio-temporal data.

1 Introduction

When describing tourism resources, National Tourism Organisations (NTOs) have to follow specific determined standards and norms. In France, for instance, NTOs have to respect the TourinFrance format.¹ This format takes into account 13 types of tourism resources (hotels, restaurants, natural resources, activities, etc.) and defines the different descriptors that can be used to characterise them. When respecting this format, a detailed objective description of tourism resources can be given, but the use of other data sources (like terminological resources, for instance) is also often required to optimise data search and management. If NTOs want to promote their resources on a tourism portal, they also have to describe the territory according to fluctuating marketing strategies and they have to provide data that are different from the ones used for management purposes so as to better answer end-users' expectations (instead of formulating their query according to strictly objective parameters; end-users may be interested in formulating it according to more subjective ones like searching for activities that are relaxing, extreme, original, etc.). Furthermore, if NTOs want to make potential tourists discover the '*Long Tail*' (Anderson, 2006) of tourism resources, they will need to be able to relate the different tourism objects to one another (they may try to attract end-users to a place by associating it to a famous sight-seeing spot, for instance).

In order to answer management and publications needs, the system has to give NTOs the possibility to integrate different data formats and to relate the data to one another. It has to give them the possibility to work at an organisational level and not a structural one. We base here our distinction between these two levels on the definitions of organisation and structure given by Marena and Varela (1998). Organisation denotes those relations that must exist among the components of a system for it to be a member of a specific class. Structure denotes the components and relations that actually constitute a particular unity, and makes its organisation concrete. Multiple data structures can be explicitly described and integrated into one common scheme of organisation (as long as they are coherent as regards one another) and, inversely, multiple potential data structures can be instantiated from the explicit data descriptions given in the organisational scheme (as long as they respect the organisational scheme). An organisational system thus gains flexibility in the two following processes:

- the internalisation process, which we will define as the controlled integration of new data into the organisational scheme
- the externalisation process, which we will define as the selective transmission of data from this organisational scheme according to a specific communication context.

For Nonaka and Takeuchi (1995), internalisation and externalisation processes correspond to a passage from explicit to tacit knowledge and from tacit to explicit knowledge. From an engineering point of view, however, integrating data into an organisational scheme requires that the relationships between those data should be explicitly described and formally represented, which makes the tacit/explicit opposition less relevant; when using a knowledge base, experts have first to define and express their knowledge in the domain, so as to provide a common framework for them to share and manage their resources, but this knowledge does not become tacit, it is formally represented so as to be machine-readable.

For the Eiffel project, an ontology-based application, called ITM², is used. This ontology gives a formal explicit description of the concepts used in the tourism domain, of the properties these concepts can have and of the relationships they share. The Eiffel ontology combines:

- *A domain ontology* which enables to describe the tourism resources and the territory. It is based on the TourinFrance format and integrates the RDF (W3C, 1999) descriptions of the INSEE³ geographical codes.
- *A functional ontology* which enables to take into account publication needs. It enables to define, for instance, tables of content and publication folders and to integrate resources like the WTO⁴ thesaurus, which gives a general vocabulary for indexing tourism resources and documents.

If this application enables tourism experts to perform different management tasks (thesaurus management, terminology management, content classification, etc.) and to create and describe new instances so as to enrich the knowledge base, it does not solve the annotation problem raised by the quantity of resources that need to be described. Even if this task is distributed among different partners, it remains tedious and time-consuming. In the first part of this paper, we analyse two ways of enriching the knowledge base via semi-automatic processes and we present the different steps to follow so as to respect the internalisation process.

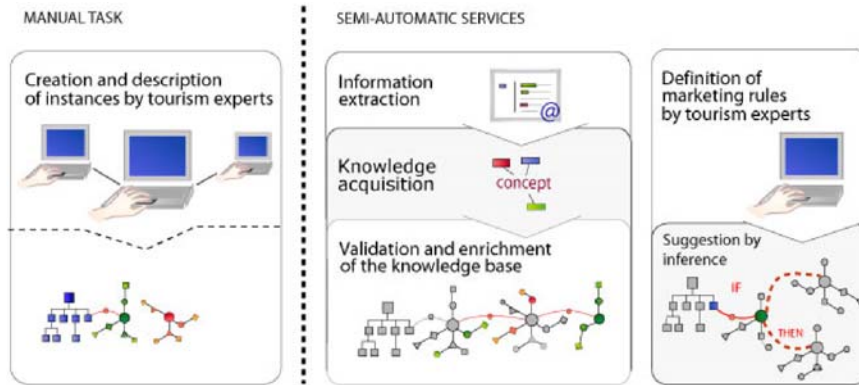
In a human-computer communication context, data that have been internalised also need to be externalised. To attract potential tourists, NTOs can present their territory via a web portal. When visiting this web portal, end-users need to have the means to discover the tourism resources that this territory has to offer. In current tourism information systems, several factors often slow down and limit end-users' exploration of the information space (tourism resources belonging to different categories are not related to one another, the content is not rich enough or cannot be filtered with precision, etc.). In the second part of this paper, we offer to analyse the externalisation process and to present a framework of analysis to characterise exploratory search. We describe how data related via explicit semantic relations can be translated into targeted information and interactions and articulated within a system of interfaces that can let end-users explore, thoroughly, the information space.

2 Internalising data via semi-automatic processes

To enrich the knowledge base, tourism experts can create and describe tourism instances manually, via ITM back-office interfaces. Tourism data, however, are often subject to modifications and need to be regularly updated. Furthermore, the web in itself is a source of data which experts do not always have the time to collect and analyse. One way of helping tourism experts populate the knowledge base is to find solutions to collect those data on the web and to integrate them into the knowledge base while controlling that the underlying organisational model is respected.

Another way of reducing the annotation task is to help tourism experts create dynamic links between the instances via the explicit definitions of Marketing Rules (MRs), instead of letting them manually define the different explicit associations that can be created between instances (Figure 1).

Figure 1 Internalisation of data into the knowledge base (see online version for colours)



2.1 From the web to the knowledge base

For the Eiffel project, we analysed how temporal and spatial data could be collected from the web and integrated into the knowledge base. Spatio-temporal data are pieces of information frequently encountered in tourism web pages. Temporal data are those related to periods and events and are characterised by different properties like frequency, duration, continuity. Spatial data are those related to areas and locations. They can be characterised via absolute references (addresses, spatial coordinates, etc.) or relative references (near the lake, close to the beach, etc.).

2.1.1 Information extraction

The type of spatio-temporal data that need to be extracted from web pages depends on the category of the tourism objects considered. In the case of a restaurant, for instance, what need to be extracted are the opening times and the address of the restaurant; whereas in the case of a concert, it is the date and place of the event that will have to be located. Identifying the string of data that has to be actually extracted can be based on a linguistic and a semiotic analysis.

In web pages, temporal and spatial data are often grouped as a list of attributes separated from the main textual description of the tourism resource and they are then introduced by a standard expression followed by a colon (*'Opening hours:'*). A list of these linguistic markers can be established and then used for pattern-matching purposes, both to locate terms introducing spatio-temporal data (*'Address:', 'From:',* etc.) or to locate actual spatio-temporal values (*'from July to December', 'all year long', 'in the forest',* etc.). As some of the linguistic markers are polysemic, it is also necessary to analyse their co-occurrence with other linguistic markers all as well as their semiotic context to look for clues allowing to disambiguate them (i.e., temporal values often include a numeric marker; a study of web page titles can provide clues about the nature of the data contained in the page).

2.1.2 Knowledge acquisition

At this stage, a mapping has to be defined between the results of the information extraction tool (structured data annotated via semantic tags) and the concepts modelled in the domain ontology (semantic data represented formally). As underlined by Amardeilh (2006), a single semantic tag can be used to map several concepts of the ontology and different semantic tags can also be used to map the same concept; knowledge acquisition rules are those defined to realise this mapping. They are based on an analysis between the way data are modelled in the domain ontology and the way they are annotated via the information extraction tool.

Let us take an example: a web page contains some information about a temporary exhibit *E* called *'Art and Sciences'* which takes place in the Beaubourg museum from March to July and this exhibit is open everyday from 2 pm to 10 pm. Those pieces of information have to be translated in the following way: *E* is an instance belonging to the class *event*. Its name is *Art and Science*. It is associated to an instance *M* of the class museum, which is identified as being the *Beaubourg museum*. *E* has different temporal properties: a property type *starting-point* which has for value *March*, a property type *end-point* which has for value *July*, a property type *opening time* which has for value *two* and a property type *closing time* which has for value *10*.

2.1.3 Knowledge validation and enrichment

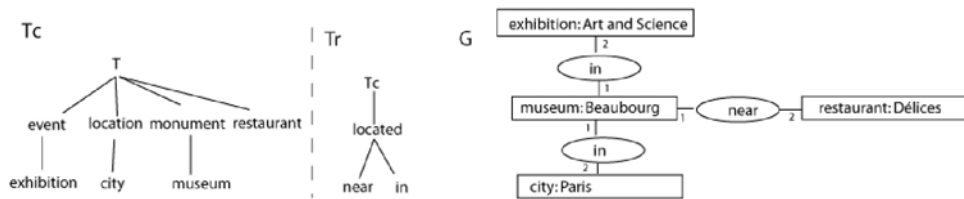
During the population of the knowledge base with spatial and temporal data, a Knowledge Validation and Enrichment Service (KVES) checks if the data imported into the knowledge base respect some constraints (validation step) and completes them by using inference rules (enrichment step).

If we take the previous example of the exhibition and we suppose that the instance corresponding to the Beaubourg museum already exists in the knowledge base and that it is associated with a restaurant *R*, the enrichment service can be used to add new properties to *E*: if *E* takes place in the museum *M*, then *E* can also be associated with *R* and *E* has the same spatial properties as *M*.

In order to do so, the *SG*-family (Baget and Mugnier, 2002) of conceptual graphs (Sowa, 1984) has been chosen as the knowledge representation and reasoning formalism upon which Eiffel project reasoning tasks are built since it is able to represent different kinds of knowledge: assertions (or facts) as Simple conceptual Graphs (SG), inference rules and constraints.

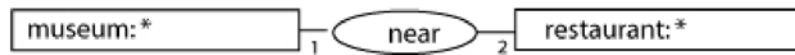
A *SG* is a bipartite labelled graph of concept nodes linked by relation nodes by labelled edges. A concept node may be generic (it represents an unspecified entity) or individual. All nodes are typed and their types are partially ordered in a structure called the ‘*support*’. Figure 2 shows a *SG* G defined via a support which presents the concept types and the relation types of the hierarchy T_C .

Figure 2 Representation of a support and a conceptual graph



Projection. To compare *SGs*, the *SG*-family provides a fundamental operation called ‘*projection*’ (in graph terms, it is a graph homomorphism). When there is a projection from Q to G , the knowledge represented by Q is included in the knowledge represented by G . Q may be considered as a query and its image by projection in G as an answer. A projection from one *SG* onto another is a mapping from the node set of the former onto the node set of the latter which preserves edges and may specify the labels. Figure 3 shows an example of a query Q , which searches for museums located near a restaurant. In this case, there is a projection from Q to the graph G defined in Figure 2.

Figure 3 Example of a query expressed via a conceptual graph



SGs and projections are the basic items upon which rules, constraints and their associated operations are built.

Inference Rules. An inference rule expresses a procedural type of knowledge, which has the following pattern “if *hypothesis* then *conclusion*”, where hypothesis and conclusion are both *SGs* linked by connection nodes. An example of a rule R could be:

“if a concept node $N1$ of the type exhibition is linked to another concept node $N2$ of the type museum by a relation of the type *in* and if $N2$ is linked to another concept node $N3$ by a relation of the type *near*, then $N1$ is also linked to $N3$ via a relation of the type *near*.”

If we follow our example, the application of the rule R would enable to infer that the exhibition ‘*Art and Science*’ takes place near the restaurant ‘*Délices*’ and a similar rule would enable to infer that it is located in Paris.

Constraints. Constraints have the same syntactical form as rules and can be either *positive* (“if the *condition* *SG* is found, so must the *obligation* *SG*”), or *negative*,

(“if the *condition* SG is found, the *interdiction* SG must not”). A SG G satisfies a positive constraint C^+ if *each* projection from the condition part of C^+ to G can be extended to a projection of its mandatory part (with corresponding connection nodes having the same image in G). G satisfies a negative constraint C^- if *no* projection from the condition part of C^- to G can be extended to a projection of its forbidden part.

2.1.4 Implementation

In order to facilitate the extraction of temporal data, a corpus of tourism web pages to be analysed is first constituted thanks to the Antidot Finder Suit (AFS)⁵ full-text indexation module. This module is also used to generate a corpus of XML documents from the web pages indexed, so as to facilitate their processing. In order to be annotated according to the text-mining rules defined with the linguistic and semiotic analysis, these documents are then analysed by transducers (Unitex)⁶ and transformed into conceptual trees. These trees are then used to populate the knowledge base via the OntoPop platform (Amardeilh, 2006).

In order to be able to use inference and constraint satisfaction services, the data stored in the ITM knowledge base are then represented via the *SG*-family (for more details on this formalism and on their mapping see Carloni et al., 2006). It is implemented in the CoGITaNT⁷ API dedicated to CGs-based applications development.

During the initialisation step, the ITM domain ontology is mapped into a support S , a rule set R and constraint sets C^+ and C^- . The ITM knowledge base (which is stated valid) is mapped into the SG G . Semantically annotated data coming from ITM are translated in a normal SG A (a SG is in a normal form if no pair of different nodes in G represents the same entity) and the following operations are performed:

- The KVES checks if A satisfies all negative constraints from C^- set.
- The KVES merges A with G (the result is written $G + A$). This is done by merging all the different concept nodes representing identical entities. As this merging process may have generated a forbidden pattern, it checks if the merged part satisfies all negative constraints.
- The KVES computes the *rules closure* G' of $G + A$ with regard to R , and checks if the inferred knowledge satisfies all negative constraints. The rule closure is done by recursively applying rules until no more new knowledge may be inferred (e.g., when each application of a rule leads to some data which are already present in the knowledge base).
- The KVES checks if G' is valid with respect to all positive constraints from C^+ set.

Steps 2–4 are carried out if the previous one did not lead to a negative constraint violation. At the end of the third step, if no negative constraint has been broken, G' is kept as the new knowledge base. In the other case G' is deleted and G continues to be considered as the knowledge base. The fourth step detects missing mandatory knowledge without considering it as a critical incoherency. At the end of this process, the KVES knowledge base is kept in normal form, closed with regard to the rule set R and valid with regard to constraints sets C^+ and C^- . If a constraint is broken, an error report is sent to ITM to indicate the part breaking the constraint. Otherwise, the knowledge inferred during the second step is returned to ITM.

Associations and properties are here inferred to enrich the knowledge base via the data that have been collected and extracted from the web. But the KVES platform is also used to provide inference services enabling the NTO to recommend tourism resources according to MRs.

2.2 A suggestion mechanism based on Marketing Rules

Recommender systems are used to display suggestions to end-users according to pre-defined parameters. As listed by Berka and Plössnig (2004), there are different types of recommender systems: they can be based, for instance, on collaborative filtering (suggestions are made according to what other end-users have also selected), they can be content-based (suggestions depend on the nature of the item chosen) or they can combine different approaches (hybrid systems).

For the Eiffel project, the suggestion mechanism had to be designed so as to support both NTO goals (NTOs want to be able to promote different tourism resources according to fluctuating marketing strategies) and end-users' expectations (end-users are the targets of NTOs' marketing strategies: for these marketing strategies to be successful, the end-users' own preferences and tastes also need to be taken into account). It is, thus, based on a hybrid approach: the suggestions made are based both on MRs defined by NTO and on the type of the tourism resources that end-users have selected.

Basically, the recommender system takes some types of resources selected by an end-user, and returns a set of suggestion patterns, according to the rules defined by NTO (Figure 4). Conceptual graphs are a formalism that fits the needs of such a system since their graphical representation enables NTO agents to easily create a set of suggestion rules, and they support the different operations (projection, inference, control) needed for our suggestion mechanism.

Figure 4 Suggestion mechanism (left) and Marketing Rule (right)

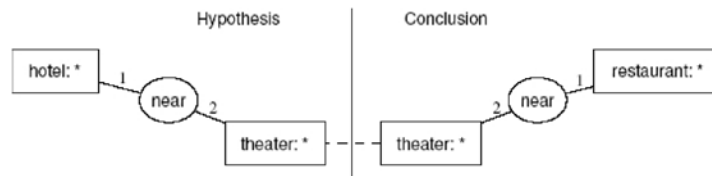


The recommender system takes as a parameter a list of resource types (defined via the elements the end-user has selected) and also clearly identifies the query. An example of an item of the list may be “(end-user chooses) a three-star hotel close to a theatre”.

Then, the system uses MRs to compute suggestions. MRs are a couple of connected CGs in which the first graph is the *condition* on the selected resource, and the second is a *pattern of suggestion* which will be completed with data from the knowledge base. An example of a MR is “if the user chooses a hotel near a theatre; then suggests restaurants near this theatre” (Figure 5). As the same ‘*theatre*’ is present in both parts of the MR, it defines a connection. The first part of the MR is searched (with the projection operation) and if it is present, the pattern is filled with the data from the knowledge base. The less a MR contains connections, the more the suggestions will

be based on NTOs' commercial policy, and the less they will take into account user interactions. When an end-user selects an item, it can activate multiple MR, which will generate one or more suggestions. The system finally returns a set of data corresponding to the different tourism resources that can be recommended to the end-user.

Figure 5 An example of a marketing rule expressed via two connected conceptual graphs



In the first part of this paper, we have analysed how data could be integrated into the knowledge base via semi-automatic processes in order to alleviate tourism experts' annotation task. As the semantics of the data is made explicit, inference services can be offered, which enables not only to control the data but also to enrich the knowledge base.

Having a rich description of the tourism resources a territory has to offer should also allow for the rich exploration of this territory. In the second part of this paper, we will analyse how the data stored in the knowledge base can be externalised on a web portal so as to improve exploratory search.

3 Externalising data on a web portal

The externalisation process corresponds to the selective transmission of data according to a specific communication context. The interface is the medium through which human-computer communication occurs: for this communication to be successful, the interface has to offer targeted information (that is, data that can potentially match an end-user's query and can thus be considered as relevant to him or her) and also, the targeted possibilities of interaction with the system (that is, interactions that the end-user may potentially want to perform given the set of data that is displayed): in short, it has to offer targeted *inform-action*.

Transforming semantically annotated data into *inform-action* is not a straightforward task. Uniform data display, irrelevant display of data infrastructure or properties and lack of usability, are some of the criticisms that have been addressed in semantic web applications (Schraefel and Karger, 2006) and these critics are all the more accentuated as the web 2.0 has brought focus on enhancing user experience (McClelland, 2005; Mahlke, 2005). When semantically annotated data are displayed on an interface, they first need to be grouped, ordered and prioritised (van Ossenbruggen and Hardman, 2002). They not only have to be given a form, they also have to be given a function. The point here is that what should be displayed on the interface is not all that is known but only what can have an informational and/or a procedural function in a specific communication context. To define these functions, the first step to take is to analyse the tasks that end-users may want to perform (Battle, 2006). In this section, we will concentrate on how using data related via explicit semantic relations can help support exploratory search.

3.1 *Exploratory search*

A web portal is the entrance door to an information space, that is a set of data that can be potentially selected so as to answer different end-users' queries. In the case of a tourism web portal, the information space contains data about an area and the objects/events/activities it has to offer. When consulting such a portal, end-users can have a very precise idea of what to look for or, on the other hand, they can expect the information system to give them ideas about where to go and what to do because they themselves do not really know what to search for and what there is to see. End-users will, thus, need help not only to search the information space – that is, to access a precise part of the information space quickly – but also to explore it – that is, to construct their way towards different parts of the information space, have an overview of what is available and decide if they want to know more about this part or if they want to change directions.

Exploratory search is the term coined to define this second type of information-seeking behaviour. It has been characterised by Marchionini (2006) as blending querying and browsing strategies and by White et al. (2006) as a way of presenting relationships and mechanisms for discovering new insights. Different papers present interface solutions to facilitate this activity (Alonso et al., 2007; Clarkson and Foley, 2007; Rose and Raju, 2007). We present here a common framework of analysis to describe exploratory search via the definition of a three-axis referential, which is based on the possibilities of interaction an end-user has, given a set of data to be explored.

3.2 *Vertical axis of exploration*

To allow for the rich exploration of a territory, it is important to take into account all the different types of tourism resources it has to offer (hotels, sport activities, cultural resources, natural resources, patrimonial resources, etc.). If the description of the territory is exhaustive, the system also becomes data-intensive and the risk is the loss of end-users due to an overload of data.

The vertical axis of exploration is the one that enables end-users to determine the selection of data that can potentially be transmitted to them by the system. It enables them to reduce or enlarge their focus on a specific part of the information space and to construct their way towards a point of interest (which can be one particular instance in particular or a collection of instances). Implementing a vertical axis of exploration not only requires giving end-users the means to retrieve a selection of data, it also requires to give them explicit indications about how they can specify their request so as to let them decide on the path they want to take to explore a part of the information space further.

In the case of the Eiffel project, vertical exploration is designed to be implemented via the combined use of a faceted navigation system and a keyword search component. Faceted navigation is a type of navigation that has already been described as facilitating exploratory search (Marchionini, 2006; White et al., 2006) since it enables end-users to access objects according to multiple paths, which correspond to the different properties according to which these objects are characterised in the knowledge base. The faceted navigation system relies on the definition of multiple navigation taxonomies, expressed in SKOS (W3C, 2005), and which concern only a selection of some property types used to characterise the instances in the knowledge base: at this stage, tourism experts have to discuss with knowledge engineers and web designers about how they think property types should be grouped, ordered and prioritised so as to better answer end-user needs.

Indeed, all the property types defined in the knowledge base for management purposes are not relevant for the web portal's end-users and if the number of facets displayed is high, it is also important to order and prioritise them so as to facilitate end-users' access to those that are more frequently used (in the case of hotels, for instance, priority display will be given to the price and the hotel category).

This faceted navigation is combined with the use of a keyword search component: the textual value entered by the end-user can be added to the other query parameters defined via the facet values. End-users have then the possibility to express their search goal according to their own terms – even if it is only poorly defined – and to refine it via the use of the facet values. For instance, an end-user might start a search by entering 'exhibition' in the keyword search component. The facets will then be actualised according to his query and he will have the opportunity to specify a date or a location or a domain, etc. via the different facets displayed.

3.3 Horizontal axis of exploration

End-users not only need to construct their own path in the information space, they also need to actually discover what's on the path. When navigating along the horizontal axis, end-users go from one displayable portion of the set of results matching their current request to another displayable portion. Indeed, the set of data that can actually be displayed to end-users does not always correspond to the set of data that can potentially be returned to them according to their query; it is constrained both by the space of the screen and by the quantity of data that an end-user can actually process and analyse.

Here, it is important to underline the dependence between the horizontal axis and the vertical axis since the more an end-user is focused on a part of the information space, the more information about each component of this information space it will be relevant to give to him or her. Facilitating horizontal exploration (enabling end-users to browse quickly a set of data) mainly relies on the implementation of appropriate information visualisation modes. Most tourism information systems only offer one visualisation mode to view a set of results. As each way of representing the data has the characteristics of making some properties more directly accessible to the viewer, it can help end-users to propose different visualisation modes so that they can choose the one that they find most appropriate for comparing the properties of the object they are interested in.

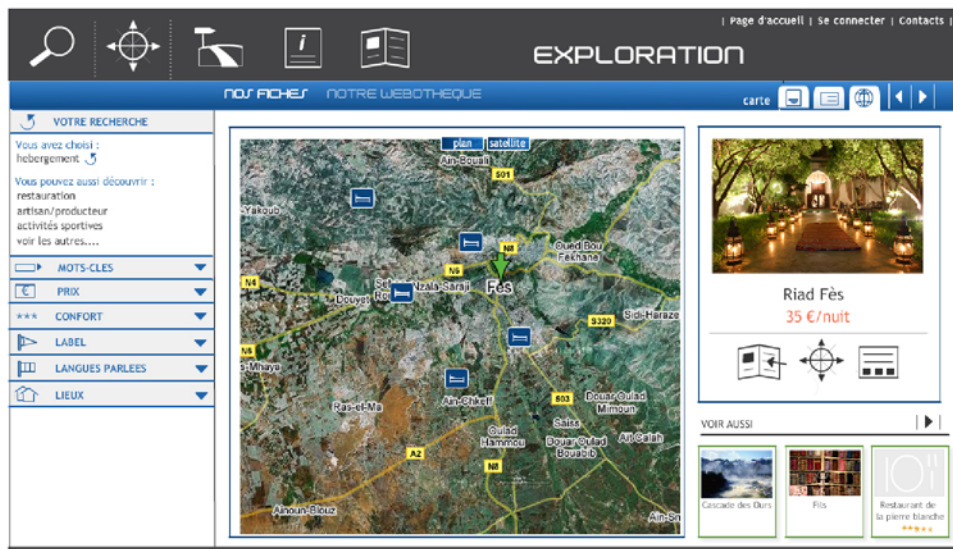
The Eiffel tourism information system is designed to implement four different visualisation modes to browse a set of results. By default, end-users can view the results in the form of image/summary cards. It contains data that have an identification function (the title of the tourism object and a photo of the object) and data that have a comparison function (three attributes maximum, determined and selected according to the class of the object).

The second visualisation mode is similar to the first one except for the inclusion of a short text that has a descriptive function (the end-user gains in informational content as regards each result but there is a loss in the number of results that he can see within the same surface).

The third visualisation mode enables end-users to view results on a map (Figure 6). End-users' lack of knowledge about the area results in a lack of knowledge about how the tourism resources are located with regard to one another and a map will give them the means to better estimate geographical distances than if they had to browse a list of geographical property values. It will also give them the means to compare different spatial areas in function of the quantity of tourism resources they offer.

Time, like space, has the characteristics of being used as a reference framework that has commonly accepted representations (timeline, agenda, week planner, etc.). The advantage of letting end-users view the results on an agenda (which, here, is the temporal representational mode that is more adapted to view results spanning within a year period) is comparable to that of using a map for geographical data: it facilitates both the comparison between objects according to their temporal properties since they are grouped into time slices, but it also facilitates the comparison between the different temporal periods as end-users can see the number of events they contain; so, they can discover at which period of the year they are more things to do.

Figure 6 A set of results displayed on a map. Resources are identified by an icon. On rollover on this icon, an overview of the tourism resource is displayed (see online version for colours)



The particularity of each of the visualisation mode proposed is that, by rollover on the different results displayed, the end-user has also the means to visualise this result in the context of the instances it is related to in the knowledge base (Figure 7). As stated by Latour (2005) in his actor-network theory, it is via the relations an actor (that is, a performing object in a network) has with other actors that this actor can be defined. From an information communication perspective, showing a tourism resource in the environment of the tourism resources it is related to has a contextualisation function (it gives information about how an object is situated in relation to other objects, not only geographically but also thematically and temporally) which will help end-users better identify the data presented to them.

Figure 7 A set of results in image-card mode. On rollover, an image-card gains more visual importance and the resources related to it are displayed in the bottom part (see online version for colours)



3.4 Transversal axis of exploration

Displaying tourism objects that are related to another tourism object also has a suggestion function: it can help end-users discover new resources and give them the means to access these resources directly, without having to formulate a query. When advancing along the transversal axis, end-users do not adjust their focus on the part of the information space they are interested in (i.e., vertical axis); they actually change their focus and access another part of the information space via the use of the relational properties of the data on which they had first been focused. Tourism information systems generally do not implement this transversal axis of exploration, or they only implement it partially, by relying only on the geographical properties of the tourism resources to offer suggestions, thus leaving out other possibilities of relating the data. This prevents end-users from gaining new insights into the information space but also slows down their exploration of the information space (instead of having a direct access to objects that are related to one another, they will have to formulate a new query to access them).

As we have seen in a previous part of this paper, data from the knowledge base can be related to one another in two ways: either via static associations or via dynamic links that are created on-the-fly via the inference mechanism. The suggestions displayed by rollover on a specific item of a set of results are those issued from static associations (the display has to be immediate and the quantity of information also has to remain limited: the rollover presentation only has to give an overview of the tourism resource). Tourism resources issued from dynamic links are displayed by default when the end-user views the set of results corresponding to his query (when there is no rollover). They are also displayed – in combination with those issued from static associations – when he consults the complete set of information of a tourism resource (Figure 8), according to the query formulated to access this tourism resource.

Figure 8 Local view on a specific tourism resource with suggestions displayed on the bottom part (see online version for colours)



3.5 Exploratory search, semantic data and web design

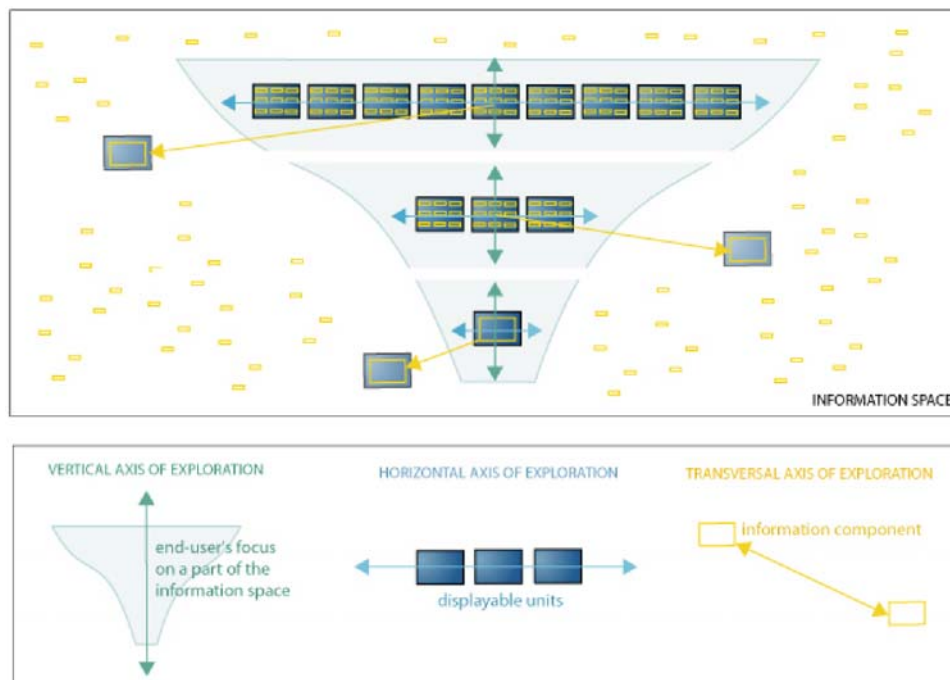
As a reminder and to show the articulation between the different axes of exploration, we will give here an example of a short scenario: an end-user has to spend a night in Paris for his work. He first searches for a hotel and uses the faceted navigation system (vertical exploration) to specify his query (Paris + three star). He browses the different results (horizontal exploration) and chooses to get more information about a hotel *H*. As he consults the local view corresponding to this resource, he discovers that it is close to a theatre. He clicks on this tourism resource (transversal exploration), consults the actual website of the theatre to check the programme via the link provided on the local view display, but finds nothing interesting. However, as he returns to the local view of the theatre, he notices a restaurant in the suggestions so he checks the address and as he does so, he notices – again in the suggestions – that there is an exhibition about *Art and Sciences* near this restaurant and he decides he will go there too.

As shown by this example, and as we have already explained in the previous paragraphs, the three axes of exploratory search (Figure 9) entertain dependency relations:

- When advancing along the vertical axis of exploration, an end-user gets more and more (or less and less) focused on a part of the information space. This modifies the information granularity level to be given, and influences the way data should be represented so as to facilitate horizontal exploration.
- To access a resource via the transverse axis of exploration, those resources first have to be displayed. As they do not correspond to the end-user's current focus on the information space as determined by the vertical axis, they have to be presented to him or her in an alternative way (and with less visual importance than the actual results of the query) as he or she advances along the horizontal axis of exploration, etc ...

We do not intend here to describe exhaustively those dependency relations. Our goal is to underline that this inter-dependency requires that the implementation of the three axes should be thought of in a combinatory way; it is through the coherent articulation of its components that the system can eventually gain in usability.

Figure 9 Exploratory search: a three-axes referential (see online version for colours)



Using data that are related via explicit semantic relations gives the potential means to combine the three axes of exploratory search: as data are not contained within fixed structural blocks, they can be assembled into multiple sets and accessed via multiple property types or association types. Managing to combine the three axes of exploration search within each interface so as to articulate these interfaces within a coherent system is then a question of interface design (regarded as a domain covering information visualisation and architecture and interaction design). As stated by Wilson et al. (2007), “the challenge is not to simply add more features, but to combine them to produce synergetic designs”.

4 Conclusion

In this paper, we have defined and analysed two processes at stake when designing a knowledge-based tourism information system: the internalisation process and the externalisation process. Using a knowledge base enables tourism experts to give a rich description of a territory and of the way the tourism resources are organised within this territory. The richer the description is, the more the annotation task becomes important. However, as the semantics of the data is made explicit, inference services can be used

- to control the integration of data automatically collected and extracted from the web
- to create dynamic associations between data and thus provide a suggestion mechanism.

Such a rich description of the territory is required if tourism organisations want to be able to manage their resources efficiently. In this paper, we have analysed how the data stored in the knowledge base could also be used to provide information on a web portal. We have presented an analysis framework enabling to describe exploratory search according to a three-axes referential and we have shown that the use of data which are related via explicit semantic relations enabled us to provide different interface solutions which can be used to improve end-users' exploration of the information space by end-users.

As part of the Eiffel project, we also intend to study end-users' interactions with the system via web usage mining techniques. Apart from enabling us to evaluate the interfaces of the web portal and giving tourism experts feedback about end-users' centres of interests, these techniques will also be used to prioritise the data when a set of results is returned so as to provide personalised interfaces.

References

- Alonso, A., BaezaYates, R. and Gertz, M. (2007) 'Exploratory search using timelines', *Proceedings of the ACM SIGCHI 2007 Workshop on Exploratory Search and HCI*, pp.23–26.
- Amardeilh, F. (2006) 'Ontopop or How to populate an ontology and annotate documents', *Proceedings of the ESWC 2006 Workshop on Mastering the Gap: From Information Extraction to Semantic Representation*, <http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-187/11.pdf>
- Anderson, C. (2006) *The Long Tail: Why the Future of Business is Selling Less of More*, Hyperion, New York.
- Baget, J-F. and Mugnier, M-L. (2002) 'Extensions of simple conceptual graphs: the complexity of rules and constraints', *JAIR*, Vol. 16, pp.425–465.
- Battle, L. (2006) 'Preliminary analysis of users and tasks for the semantic web', *Proceedings of the 3rd International Semantic Web User Interaction Workshop*, Athens, Georgia, USA.
- Berka, T. and Plössnig, M. (2004) 'Designing recommender systems for tourism', *Proceedings of ENTER 2004*, Cairo.
- Carloni, O., Leclere, M. and Mugnier, M-L. (2006) 'Introducing graph-based reasoning in a knowledge management tool', *Proceedings of IEA-AIE*, Springer, Berlin/Heidelberg, pp.590–599.
- Clarkson, E.C. and Foley, J.D. (2007) 'ResultMaps: search result visualisation for hierarchical information spaces', *Proceedings of the ACM SIGCHI 2007 Workshop on Exploratory Search and HCI*, pp.32–35.
- Latour, B. (2005) *Reassembling the Social: An Introduction to Actor-Network Theory*, Oxford University Press, New York.
- Mahlke, S. (2005) 'Understanding users' experience of interaction', *Proceedings of the 2005 Annual Conference on European Association of Cognitive Ergonomics*, Chania, Greece, pp.251–254.
- Marchionini, G. (2006) 'Exploratory search: from finding to understanding', *Communications of the ACM*, Vol. 49, No. 4, pp.41–46.
- Marena, H.R. and Varela, F.J. (1998) *The Tree of Knowledge*, Shambhala, Boston, MA.

- McClelland, I. (2005) 'User experience design: a new form of design practice takes shape', *Proceedings of ACM CHI 2005 Conference on Human Factors in Computing Systems*, pp.1096–1097.
- Nonaka, I. and Takeuchi, H. (1995) *The Knowledge-Creating Company*, Oxford University Press, New York.
- Rose, D.E. and Raju, S. (2007) 'Encouraging exploration with Elroy: a new user experience for web search', *Proceedings of the ACM SIGCHI 2007 Workshop on Exploratory Search and HCI*, pp.56–59.
- Schraefel, m.c. and Karger, D. (2006) 'The pathetic fallacy of RDF', *Proceedings of International Workshop on the Semantic Web and User Interaction (SWUI)*, Athens, Georgia.
- Sowa, J.F. (1984) *Conceptual Structures: Information Processing in Mind and Machine*, Addison-Wesley Systems Programming Series, Boston, MA.
- van Ossenbruggen, J. and Hardman, L. (2002) 'Smart style on the semantic web', *Semantic Web Workshop WWW2002*, Hawaii.
- W3C (1999) *Resource Description Framework (RDF) Model and Syntax Specification*, W3C Recommendations are available at <http://www.w3.org/TR>, 22 February.
- W3C (2008) Miles, A. and Brickley, D. (Eds.) *SKOS Simple Knowledge Organization System Reference*, W3C Working Draft 25 January 2008, available at <http://www.w3.org/TR/2008/WD-skos-reference-20080125/>
- White, R., Rules, B., Drucker, S. and Schraefel, m.c. (2006) 'Supporting exploratory search', *Communications of the ACM, Special Issue*, Vol. 49, No. 4, pp.36–39.
- Wilson, M.L., Schraefel, m.c. and White, R.W. (2007) *Evaluating Advanced Search Interfaces using Established Information-seeking Models*, Technical Report, available at <http://eprints.ecs.soton.ac.uk/13737/>

Notes

¹TourinFrance: <http://www.tourinfrance.net/tourinfrance22/index.htm>

²Mondeca Intelligent Topic Manager (ITM) – www.mondeca.com

³INSEE: Institut National de la Statistique et des Etudes Economiques – www.insee.fr

⁴WTO: World Tourism Organization – www.unwto.org

⁵AFS: <http://www.antidot.net/fr/technologie/afs>

⁶Unitex: <http://www-igm.univ-mlv.fr/~unitex/>

⁷Cogitant Web Site. <http://cogitant.sourceforge.net> (David Genest).