SQL-TC: A Topic-Centric Query Language for Web-Based Information Resources†

İ. S. Altıngövde+, S. A. Özel+, Ö. Ulusoy+, G. Özsoyoğlu#, Z. M. Özsoyoğlu#

Bilkent University and Case Western Reserve University
(ismaila, selma, oulusoy)@cs.bilkent.edu.tr, (tekin, meral)@eecs.cwru.edu

Abstract

This report deals with the problem of modeling web information resources using expert knowledge and personalized user information, and querying them in terms of topics and topic relationships. We propose a "web information space" metadata model for web information resources, and a query language SQL-TC (Topic-Centric SQL) to query the model. The web information space model is composed of web-based information resources (XML or HTML documents on the web), expert advice repositories (domain-expert-specified metadata for information resources as XTM documents), and personalized information about users (captured as user profiles, that indicate users' preferences as to which expert advice they would like to follow, and which to ignore, etc., as XML documents).

Expert advice is specified using topics and relationships among topics (called metalinks), along the lines of the recently proposed topic maps. Experts attach importance values and domains to topics and metalinks that they specify. And, users declare to accept, reject or "don-t-care" about the choices that experts make.

The query language SQL-TC makes use of the metadata information provided in expert advice repositories and embedded in information resources, and employs user preferences to further refine the query output. Query output objects/tuples are ranked with respect to the (expert-judged and user-preference-revised) importance values of requested topics/metalinks, and the query output is limited by either top n-ranked objects/tuples, or objects/tuples with importance values above a given threshold, or both.Therefore, the query output of SQL-TC is expected to produce highly relevant and semantically related responses to user queries within short amounts of time.

1 Introduction

Due to the enormous growth of the world wide web in the last decade, today the web hosts very large information repositories containing huge volumes of data of almost every kind of media. However, due to the lack of a centralized authority governing the web and a strict schema characterizing the *data* on the web –which obviously promotes this incredible growth– finding relevant information on the web is a major struggle. Researchers in many fields of computer science including databases, information retrieval, data mining and AI investigate the problem of locating only the relevant information and locating it fast.

XML [10], one of the recent developments for web and adopted as a standard by W3C, provides a simple syntax for self-describing web data, and will likely become a data exchange model on the web [1]. Meanwhile, the topic maps model [8], another recently adopted standard, allows web administrators to define additional semantics for web documents in terms of topics, topic associations, etc., and, thus, provides a new way of navigating information resources. Basically, a topic map is a semantic data model describing the contents of web documents in terms of topics and topic associations, and therefore constitutes a "metadata" model. Thus, topic maps allow web users to benefit from semantic data modeling that may be employed in variety of ways, one of which is to improve the performance of search engines [29]. As one would anticipate, efforts are underway to combine XML and topic maps, and the result is the XML topic map (XTM) effort that intends to express a syntactic and functional subset of the topic map standard in XML [29].

---

† A shorter version of this paper appears in the Proceedings of DEXA 2001. For correspondence, please contact the first author. This research is supported by a joint grant from TÜBİTAK (grant no. 100U024) of Turkey and the National Science Foundation (grant INT-9912229) of the USA.

+ Affiliation: Department of Computer Engineering, Bilkent University, Ankara, Turkey

Affiliation: Department of Electrical Eng. and Computer Science, Case Western Reserve University, Cleveland, Ohio 44106

To illustrate the advantages of using metadata for an improved searching/querying paradigm, consider the movie database at www.movie-bank.com. Assume that we would like to locate movies listed at this site and are related to the novel “Carrie” and written by the novelist Stephen King, and are rated at least “very good” (i.e., with an importance value above 0.7 in a scale of 0 to1) by the movie critic (expert) Joe Siegel. Presently, such a task can be performed by browsing the movie pages or by a keyword-based search on a web search engine followed up by a lookup of (some of) the resulting hits, which may be ineffective as well as time-inefficient. Assume that we have an expert that provides a data model for this web site, where “novel”, “Carrie”, and “Stephen-King” are topics, *RelatedTo* and *WrittenBy* are relationships among topics (called *associations* in the topic map standard, and, in this report, referred to as *topic metalinks*), and for each topic, there are, perhaps X-Pointer-like pointers pointing to web documents containing “occurrences” of that topic, called *topic sources*. Then, we could formulate and evaluate the query "find movies *RelatedTo* novel Carrie, *WrittenBy* Stephen King, and rated above 0.7 by Joe Siegel" against the data model of the information source, and satisfy the user’s request in an efficient manner (assuming that the query optimization takes place). One can thus view the data model as a “knowledge (-based)-index” to the web information resource. Then such a knowledge-index can be the starting point for a semantic-based web search and controlled delivery of the response to users.

In this report, we describe a "web information space" data model for web information resources, and the query language SQL-TC, where TC stands for topic-centric, to query the data model and web information resources in an integrated manner. The information space is composed of

- Web-based information resources which are XML or HTML documents.
- Independent *expert advice repositories* that contain domain expert-specified model of information resources. We assume that the expert advice, modeled as topic maps, is stored and maintained as XTM documents.
- *Personalized information* about users, captured as user profiles, that contain users’ preferences as to which expert advice they would like to follow, and which to ignore, etc., and users’ knowledge about the topics that they are querying. We maintain user profiles as XML documents.

In this model, topics and topic metalinks are the fundamental concepts through which we model and query the contents of information resources. It is important to note that the expert advice repository is a *metadata* model, designed independently from the associated information resources (with the exception of topic source specifications) to model possibly multiple information resources, and capturing the expertise of a domain expert in a lasting manner. Therefore, the expert advice repository is *stable* (i.e., changes little), stays relevant (with the exception of topic sources) even when the information resource changes over time, and is much smaller than the information resource that it models. Another implication is that the topic metalinks are more likely to specify recursive relationships between topics, such as the relationships *RelatedTo, leadsTo, IsIn, Prerequisite, ComposedOf,* etc. Finally, SQL-TC query output objects/tuples are ranked with respect to the (domain-expert-judged and user-preference-revised) importance values of requested topics/metalinks. The SQL-TC query output sizes are kept small by returning either (a) top n importance value-ranked objects/tuples, or (b) objects/tuples with importance values above a given threshold, or (c) both.

Thus, the main advantage of our proposal for web search and querying are (a) incorporating expert advice and personalized information, and (b) controlled delivery of query outputs in terms of top-ranking objects/tuples above a given importance value threshold. The disadvantage is the cost of creating and maintaining expert advice and personalized user information. Note that the expert advice, being stable over time, is a one-time effort to create, amortized by its use over time and fast response to user queries. We also make the practical assumption that the modeled information resources, however large they may be, do not span the web; they are defined within *subnet*s such as the ACM SIGMOD Anthology sites, or the larger domain of Microsoft Developers Network sites [27], or the very large domain of Online Collections of the Smithsonian Institution [32]. And, different expert advice may be provided for the same web information resource to express varying viewpoints of different domain experts.

The query language SQL-TC allows users to query both the XTM-structured expert advice repositories, and the associated information resources. Thus, querying resources with respect to multiple expert advices simultaneously, coupled with the incorporation of personalized information, is expected to produce highly relevant and semantically related responses to users' queries within short time spans.

In the next section, we briefly discuss XML, topic maps and the related standard, and XTM. Section 3 is devoted to the web information space model with expert advice and user profiles. In Section 4, SQL-TC query language syntax and its features are covered, along with a number of examples. Section 5 describes the prototype implementation employing XTMs as expert advice repositories. Finally, we conclude and point out future research directions in Section 6.

2 Related work

Extensible Markup Language (XML) [10] is becoming a universal standard for data exchange on the web, recommended by the W3C Consortium. XML-Data [23] describes data in a self-describing format, either only through tags for elements and attributes (i.e., well-defined documents), or through separately defined schema (i.e., DTDs and valid documents). XML schemas specify metadata information, and allow one application on the web to receive data from another application without any prior built-in description of the data. It is claimed that the use of XML will bring a major change in the structure of web information [9], and XML will be the basis for data interchange on the Internet, i.e., electronic data interchange (EDI) format [15]. Recent research activity on XML and databases include, among others, storing and querying XML documents, XML views, XML architectures and others. (See as an example [3]).

We summarize the Topic Map data model, as described in [6, 8, 33]. Definition of a topic is very general: a topic can be anything about which anything can be asserted by any means. As an example, in the context of Encyclopedia, the country Spain, or the city Rome are topics. Topics are *typed,* (e.g., type of the topic Rome is city), and have names. Topic names are also typed; e.g., base name (required), display name (optional), etc. Topic names have *scopes,* e.g., language, style, domain, etc. Topics have *occurrences* within addressable information resources. For example, a topic can be *described* in a monograph, *depicted* in a video or a picture, or *mentioned* in the context of something else. Moreover, each occurrence is typed using the notion of *occurrence role.* A *topic association* specifies a relationship between two or more topics. For example, topic Rome *is-in* topic Italy; topic Tom Robbins *was-born-in* topic USA, etc. A *topic map* is a structure, perhaps a file or a database or an XML document, which contains a topic data model, together with occurrences, types, contexts, and associations. Publicly available example topic maps and topic map processors are provided in [39, 40]. Commercial topic map authoring tools are presented in [37, 38]. Topic-map embedded proceedings of GCA conferences is provided in [38] and their creation is discussed in [7]. A commercial topic map search engine is presented in [31] and a topic map query language is proposed in [22].

XTM (XML Topic Map) is an effort to represent topic maps as XML documents. The proposals and DTDs for XTM are publicly available in [42, 43]. An XTM Processing Model is provided in [44]. And, a number of technical issues on an XTM specification; such as linking issues (i.e., using Xlink [41]), addressing issues (i.e., among topics and occurrences), issue of including templates in the specification, specifying association properties and inference rules, and addressing the public topics are discussed in [29, 34, 35] as well as the XTM workgroup (WG) mailing list documents [43]. For the time being, none of these proposals are formally accepted, but we assume our models and tools would be flexible enough to support the standardized DTD version with minor modifications.

Recently, there has been significant activity on web modeling, and transforming model-based information. Semantic Web [5] is an RDF schema-based effort to define an architecture for web, with a schema layer, logical layer, and a query language. The semantic web workshop [16] contains various proposals and efforts for adding semantics to web.

While we propose a metadata-based search and querying for specific subnets on the web, the proposed web query languages in the literature have the broader goal of querying the web as a whole. One can view the world wide web as a directed graph, and formulate queries based on the contents of web pages [18]. Query languages for the web include WebSQL [25], STRUQL [17], Lorel[2], and UnSQL [11, 12]. There is also the issue of integrating information on different web sites (using "wrappers" [19]), and providing query capabilities over multiple web sites [24, 26]. More recently, a query language, called XML-QL [14] is proposed to the World Wide Web Consortium (W3C; http://www.w3.org) as a query language for XML.

3 Web Information Space Model

In this section, we present the topic-based web information space model, as illustrated in Figure 1. The three components of the model are information resource model, expert advice model, and user profile model.

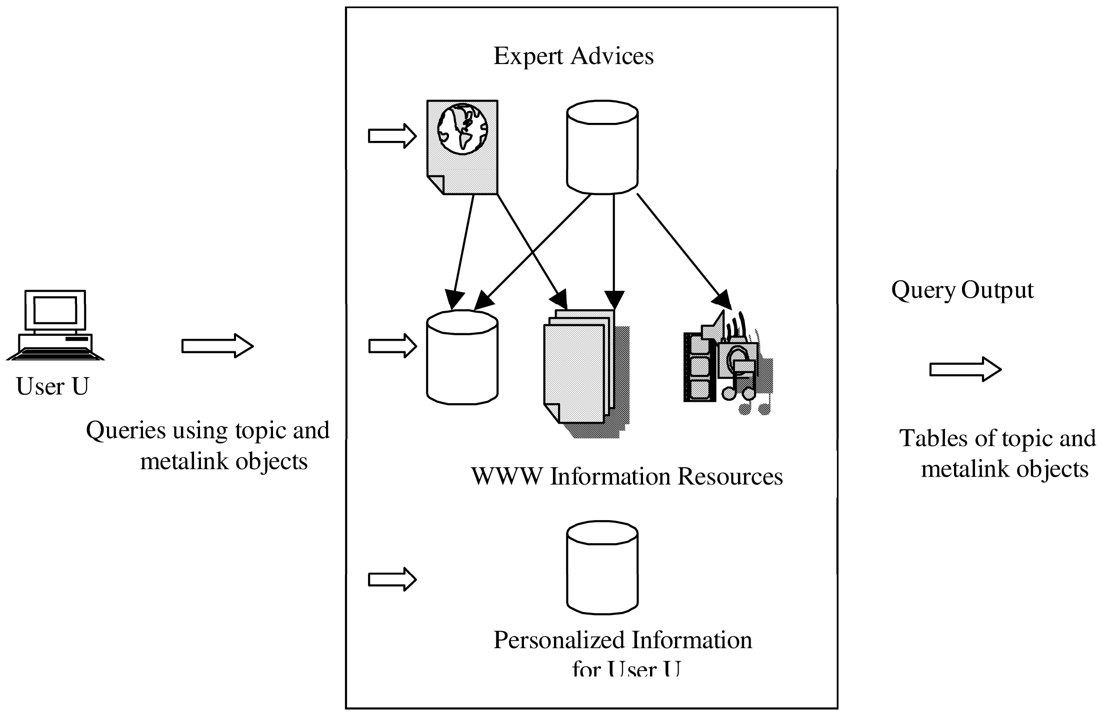

Figure 1. Web Information Space Model and Queries

3.1 Information Resource Model

Information resources are web-based documents, containing multimedia data of any arbitrary type. They may have bulk text in various formats (e.g., ascii, postscript, pdf, etc.), images with different formats (e.g., jpeg), audio, video, audio/video, etc. For the purposes of this research, we assume that information resources are in the form of XML/HTML documents.

Topic source represents an occurrence of a topic within an information resource. For example, the topic (with name) "Van Gough" occurs multiple times as HTML documents within the documents of the information resource "Online Collections of the Smithsonian Institution" [32], and each such HTML document occurrence constitutes a topic source. For XML-based web documents, we assume that a number of topic source attributes are defined within the XML document (using XML element tags) such as where the topic source starts within the document and where it ends, *LastUpdated, LastVisited, Author,* and *MediaType* attributes, etc. Also, the expert advice model, discussed next, has an entity, called Topic Source Reference, which contains (partial) information about a topic source (such as its web address, etc).

3.2 Expert Advice Model

3.2.1 Topic and Topic Source Reference Entity Types

We assume that the experts in the web information space model are registered and known either through the user profiles or each query explicitly, and we have n experts, E_i, $1 \le i \le n$. Each domain expert models an information resource in terms of topic and topic source reference entities and, metalink relationships. We start with the topic entity, which constitutes metadata, and has the following attributes.

- *T(opic-)Name* (of type string) contains either a single word (i.e., a keyword) or multiple words (i.e., a phrase). Topic names characterize the data (real-world subjects [42]) in information resources. Example topic names are "database" (a keyword) and "Understanding the United Nation's Global Warming Policies" (a phrase). Topic names are defined by domain experts, and can be arbitrarily specified phrases or words. Therefore, the issue of similarity between topic names is addressed. To check for the similarity of two topics on the basis of their names, we employ SimTName( ) function, which returns the name similarity of two topics with arbitrarily long topic names as a real value within the range [0, 1].
- *T(opic-)Type* and *T(opic-)Domain* attributes specify, respectively, the type of the topic and the domain within which the topic is to be used. For example, the topic "Hamlet" is of type "character" in the domain of "plays". The topic "Paris" may be of type "Greek god" in the domain of "mythology", whereas it is of type "city" in the domain "geography". And, the topic "diabetes" may be of type "chronic disease" in the domain of "medicine". Again, we allow different experts to use different words/phrases for topic types and topic domains.
- *T(opic-)Author* attribute defines the expert (name or id or simply a URL that uniquely identifies the expert) who authors the topic.
- *T(opic-)MaxDetailLevel.* Each topic can be represented by a topic source in the web information resource at a different *detail level*. Therefore, each topic entity has a maximum detail level attribute. As an example, assume that levels 1, 2 and 3 denote levels "beginner", "intermediate", and "advanced". Therefore, for a web-based information resource on finance (e.g., Microsoft MoneyCentral [28]), for example, a source for topic "stock market" can be at a beginner (i.e., detail level 1) level, denoted by $StockMarket^1$ (e.g., only "Basic Investing" and "Stock Market Indexes"). Or, it may be at an advanced (say, detail level n) level of $StockMarket^n$ (e.g., "Risk Analysis and Random Walk Theory"), etc. Note the convention that topic x at detail level i is more advanced (i.e., more detailed) than topic x at detail level j when

i>j. Also note that the detail level value of a topic source must be less than or equal to the maximum detail level attribute of the topic.

- *T(opic-)id.* Each topic entity has a T(opic-)id attribute, whose value is an artificially generated identifier, internally used for efficient implementation purposes, and not available to users.
- *T(opic-)SourceRef.* Each topic entity has a T(opic-)SourceRef attribute which contains a set of Topic-Source-Reference entities as discussed below.
- Topics also have other attributes such as roles, role-playing, etc. Some of these additional attributes are discussed in the topic map standard and described in detail in [42].

The attributes (TName, TType, TDomain, TAuthor) constitute a key for the topic entity. And, the Tid attribute is also a key for topics.

The expert E_i, $1 \leq i \leq n$, states his/her advice on topics as a Topic-Advice function TAdvice() that assigns an *importance value* to topics from one of $[0, 1] \cup \{\text{No, Don't-Care}\}$. The importance value is a measure for the importance of the topic, except for the cases below.

(a) When the value is "No", for the expert, the topic is rejected (which is different than the importance value of zero in which case the topic is accepted, and the expert attaches a zero value to it), and

(b) When the importance value is "Don't-Care", the expert does not care about the use of the topic (but will not object if other experts use it), and chooses not to attach any value to it. Don't-Care value is used when merging multiple expert advices.

We give an example.

Example 1. Assume that the expert E assigns the following topic advice:

TAdvice(E, TType="Diabetes", TName="*Diabetes Surgeries*", TDomain="New Patient Training") = 0.3
TAdvice(E, TName="Diabetes Management", TDomain="Patient Training") = 1
TAdvice(E, TName="*Kidney Complications*", TDomain="Nurse Training") = 0.7
TAdvice(E, TName="Advantages of Extreme Sports for Diabetes") = No
TAdvice(E, TName="Professional Sports and Diabetes") = Don't-Care

where * denotes a wildcard character that matches any string. The first topic advice states that, for training new patients, a topic of type diabetes and with a name containing the phrase "Diabetes Surgeries" is of low importance value. The second topic advice states that, for patient training, the topic name "Diabetes Management" of any topic type is of highest importance. The fourth topic advice states that, for any domain and any type, the topic "Advantages of Extreme Sports for Diabetes" is rejected. And, as the last topic advice, the expert does not care about the topic "Professional Sports and Diabetes" and does not object to its use as a topic name by other experts.

For the topic advice function TAdvice(), we use the Closed World Assumption with the "No" (or the "Don't'-Care") option, denoted as CWA-No (or CWA-Don't'-Care) that states that any TAdvice() choice that is not explicitly specified has the value "No" (or "Don't-Care", respectively).

Example 2. Using the CWA-Don't-Care assumption and TAdvice() specifications in example 1, the expert specifies:

TAdvice(E, TName="Drinking", TDomain="Professional Sports")="Don't-Care".

A *T(opic-)S(ource-)Ref(erence)*, also an entity in the expert advice model, contains additional information about topic sources. A topic source reference entity has the following attributes.

- *Topics* (set of Tid values) attribute that represents the set of topics for which the referenced source is a topic source.
- *Web-Address* (URL) of the document that contains the topic source.
- *Start-Marker* (address) indicating the starting address of the topic source relative to the beginning of the document. For topic sources in XML-based web documents, this attribute is redundant.
- *End-marker* (address) indicating the end address of the topic source relative to the beginning of the document. For topic sources in XML-based web documents, this attribute is redundant.
- *Detail level* (sequence of integers). Each topic source reference has a *detail level* describing how advanced the level of the topic source is for the corresponding topic. The detail levels are ordered using the same ordering of the corresponding topics in the attribute Topics.
- Other attributes such as *Mediatype, Role, Last-Modified, Last-Visited* etc.

The expert E_i, $1 \leq i \leq n$, states his/her advice on topic sources as a Source-Advice function SAdvice() that assigns an importance value to topic sources from one of $[0, 1] \cup \{No, Don't\text{-}Care\}$.

Example 3. Assume that the expert E assigns the following topic source advice:

SAdvice(E, TType="disease", TName="Cancer", TDomain="PatientCare",
Web-Address=www.mayo.org) = 0.5
SAdvice(E, TType="chronic disease", TName="Diabetes Management", TDomain="PatientCare",
Web-Address=www.ada.org) = 1
SAdvice(E, TName"Breaking Diabetes News", TDomain="Diabetes Research">},
Last-Modified=(Now – 2years)) = No

The last source advice states that, for the topic name "Breaking Diabetes News" (of any topic type) in the topic domain "Diabetes Research", any source that was last modified two years ago is rejected.

In addition to comparing topic entities by their names (as strings), we compare topics by their topic sources using the function SimTopicSource( ), which returns the similarity of two topics by their topic sources as a real value within the range [0, 1].

3.2.2 Metalink Types

Topic Metalinks represent relationships among topics. Metalinks may have as attributes types, roles, domains, etc. An example metalink type is *RelatedTo.* For example, the topic with name "Downhill Skiing" at level 1 is *RelatedTo* the topic with name "Super GS Ski Racing" at level 1, represented using the notation "Super GS Ski Racing^{1}" →*RelatedTo* "Downhill Skiing^{1}". The notation →*RelatedTo* represents an instance of the metalink type *RelatedTo*. As another example, consider learning-related metalink type *Prerequisite*, and the metalink instance "Diabetes Complications^{2}→ *Prerequisite* Diabetes^{1}" stating that "Understanding of the topic Diabetes at level 1 (or higher) is the prerequisite to understanding/learning the topic Diabetes Complications at level 2". Within the context of electronic books, we have given [30] a sound and complete set of axioms for the *Prerequisite* relationship. Similar to prerequisite metalinks, one may use the notion of *topic corequisites*, specifying the corequisites of the given topic. Yet another metalink relationship can be the *LeadsTo* relationship that states, for example, that "the topic relational model *LeadsTo* the topic query languages". Thus any relationship involving topics deemed suitable by an expert in the field can be a topic metalink.

SubTopicOf and *SuperTopicOf* metalink types together represent a topic composition hierarchy. As an example, the topic "database" is a super-topic (composed) of topics "data model", "query languages", "query processing", etc. And the topic "relational algebra" is a sub-topic of "query languages" and "relational model".

Metalinks represent relationships among topics, not topic sources. Therefore, they are "meta" relationships, hence our choice of the term "metalink". And, metalink types are usually recursive relationships.

The expert E_i, $1 \leq i \leq n$, states his/her advice (i) on metalink type signatures as the set *Metalinks,* and (ii) on metalink instances as a Metalink-Advice function MAdvice() that assigns an importance value to a metalink from one of $[0, 1] \cup$ {No, Don't-Care}. E_i.Metalinks denote the set of metalink types defined by the expert E_i. Similarly, E_i.Topics denote the set of topics defined by the expert E_i.

Example 4. Assume that the expert E states the following metalink signatures:

E.Metalinks={*RelatedTo:* topic → topic, *Prerequisite*: SetOf topic→ SetOf topic }

where the first signature states that the *RelatedTo* metalink type takes two topics of any type as arguments, and the second signature states that the *Prerequisite* metalink type takes two sets of topics of any type as arguments. Now, assume that the expert E states the following metalink (instance) advice:

MAdvice (E, Diabetes Care^1 →*RelatedTo* Diabetes Complications^1)=0.8
MAdvice (E, Diabetes Care^1 →*Prerequisite* Healthy Eating^1)=1
MAdvice (E, Diabetes Surgeries^3 →*Prerequisite* Diabetes Care^1)=No

The first metalink states that the importance value of the metalink "the topic Diabetes Complications at the beginner level (1) is related to Diabetes Care at the beginner level" is reasonably high (0.8) (There may be other causes for diabetes complications). The second metalink states that understanding healthy eating at a beginner level is a prerequisite to understanding diabetes care at a beginner level. And, the last metalink states that understanding diabetes care at beginner level or above is not a prerequisite to understanding the topic diabetes surgeries at an expert level.

We assume that there are multiple experts (and, thus multiple expert advices) on information resources, with each expert specifying (a) possibly different topic entities with similar names, (b) overlapping topic sources, and (c) possibly different metalink types and instances. Thus, the system may need to merge the advices from multiple experts and resolve the possible conflicts among them. An example illustrating this situation along with a user preference-based solution attempt is provided in the example query 9 of Section 4.

In this work, we assume that the expert advice described here may either be embedded in information resources or stored independently; in which case, we assume that the expert advice is in the form of an XTM document. The prototype system described in Section 5 is developed using XTM documents as expert advice repositories. Note that XTM is a recently initiated effort, and there are many issues with XTM to be resolved, conceptually or practically, such as namespaces, topic map merging and processing model [44] implementations. In this sense, our research also contributes to resolve some of these issues as it provides a real life application for the use of XTMs.

3.2.3 Topic and Metalink Closures

As stated before, metalink types are usually recursive. For example, *RelatedTo* is both transitive and reflexive. *IsIn* is transitive, but not reflexive; *SubTopicOf* is transitive. Therefore, when a user lists a set X of topics, and asks for topic sources of topics in X as well as others that are *RelatedTo* topics in X, we need to take the "topic closure" of the topic set X with respect to the recursive metalink type *RelatedTo*. We emphasize the notion of *Topic Closures* with respect to recursive metalink types, in order to return query results that satisfy all the axioms of the associated metalink types. Given a set X of topics, the query response will include the topic closure X^+, which is formed of all topics that are logically implied by the initial set X.

Assume that the expert E specifies four topic entities A (with name "Diabetes Care"), B (with name "Insulin Shot Plan"), C ("Diabetic Food Plan"), and D ("Carbohydrate Counting"), and the metalinks A→*Prerequisite*(B, C), and C→*Prerequisite* D. (For the sake of simplicity, assume that all detail level values are 1, and ignored). And, the user U asks for topic sources of topic A, subject to the advice of expert E. Since the *Prerequisite* metalink is transitive, the user's request about topic A needs to be expanded by a *topic closure* of A with respect to the *Prerequisite* metalink instances specified by E. Also note that we can always decompose the right-hand side (RHS) of a *Prerequisite* metalink. That is, the metalink

A→*Prerequisite*(B, C) is equivalent to the metalinks A→*Prerequisite* B and A→*Prerequisite* C.

Thus, the correct response should include the topics A, B, C, D, and their resources. In this section, we briefly discuss how to compute *Topic Closures*, i.e., given a set X of topics, obtaining the closure X^+ (the set of topics that are also logically implied and thus are in the response). Clearly, computing topic closures requires a sound and complete set of axioms for the metalink types deployed by the expert E, and a polynomial-time algorithm that computes the topic closure using the axioms.

Consider the *Prerequisite* metalink type. In our earlier work [30], we gave the following set of sound and complete axiomatization for the *Prerequisite* metalink type.

Case **1.** Prerequisite metalinks are not left-hand-side (LHS) decomposable (that is, A, B→*Prerequisite* C is not equivalent to the metalink A→*Prerequisite* C and the metalink B→*prerequisite* C, and are allowed to be cyclic.

Axioms: Let X, Y, and Z denote sets of topics.

- Subset-Reflexivity. If $Y \subseteq X$ then X→ *Prerequisite* Y
- Augmentation. If X→*Prerequisite* Y then XZ →*Prerequisite* YZ for any Z
- Transitivity. If X→*Prerequisite* Y and Y→*Prerequisite* Z then X→*Prerequisite* Z

These are the so-called Armstrong's axioms [45].

Case **2**. Prerequisite metalinks are not LHS-decomposable and are acyclic.

Axioms: Let X, Y, Z and W denote sets of topics.

- Pseudo-transitivity. If X→*Prerequisite* Y and WY→*Prerequisite* Z then WX→*Prerequisite* Z
- Split/join. If X→*Prerequisite* YZ then X→*Prerequisite* Y and X→*Prerequisite* Z, and vice-versa

In [30], we have proven that these axioms are sound and complete.

Case **3**. Prerequisite metalinks are LHS-decomposable.

We first decompose the LHS of all metalinks so that all metalinks have a single topic in the left and the right hand sides. And, then the only axiom is

- Transitivity. If A→*Prerequisite* B and B→*Prerequisite* C then A→*Prerequisite* C where A, B, and C are topics.

Note that in all three cases, the topic closure X^+ of a set X of topics can be found by using an O(n.l) topic closure algorithm, specified in Figure 2 below, where n is the number of prerequisite metalinks, and l is the length of the encoding for a prerequisite metalink [30].

```
Input: Set of topics X, Set M of metalink types:
          Var X: SetOf topic;
              M: SetOf metalink;
              Change: Boolean;
              OldX : SetOf topic;

Output: Closure set X+ of topics X with respect to
          metalink types M:
          Var X+ : SetOf topic;
begin
X+ := X;
OldX := X+;
Change := True;
while (Change) do
    X+ := X+ ∪ {y} such that there is a metalink instance of type T
          in M of the form Z →T {y} where Z ⊆ X+ and y ∉ X+;
    if (X+= OldX)
      then Change := False
      else OldX := X+;
 endwhile
 end
```

Figure 2. Topic closure algorithm (involving multiple metalink types)

For each new metalink type added into the expert advice model, sound and complete axioms for all metalink types, including those that apply to multiple metalink types are found. To illustrate this, consider the *RelatedTo* metalink type and the cyclic and nondecomposable *Prerequisite* metalink type. Note that, from its signature, all *RelatedTo* metalink instances have a single topic in the LHS and the RHS. Then we have the following axioms:

RelatedTo Axioms:

- Reflexivity. If A →*RelatedTo* B then B →*RelatedTo* A
- Transitivity. If A →*RelatedTo* B and B →*RelatedTo* C then A→*RelatedTo* C

Prerequisite Axioms: Armstrong's axioms (Case 1 above).

RelatedTo and *Prerequisite* mixed axioms:

- If X →*Prerequisite* A and A →*RelatedTo* B then C →*RelatedTo* B for all C where $C \in X$.
- If X → *RelatedTo* A and A →*Prerequisite* B then C →*RelatedTo* B for all C where $C \in X$.

With these axioms, we can find the topic closure X^+ of a set X of topics by using the O (n.l) closure algorithm in Figure 2, where n is the number of *Prerequisite* and *RelatedTo* metalinks, and l is the max length of the encoding for a *Prerequisite* or a *RelatedTo* metalink.

3.3 Personalized Information Model: User Profiles

The user profile model maintains for each user his/her preferences about experts, topics, sources, and metalinks as well as the user's knowledge about topics.

3.3.1 User Preferences

In this report, we employ user preference specifications, along the lines of Agrawal and Wimmers [4]. The user U specifies his/her preferences as an ordered set of Accept-Expert, Accept-Expert-Metalink-Importance-Threshold, etc. statements, as listed below.

(Accept-)Expert (U)= < E_i,…, E_j> $1 \le i, j \le n$ (sequence of experts whose advice is to be satisfied)

(Accept-Expert-)T(opic-)Imp(ortance-Threshold) (U) = {(E_i, ValueThreshold),…, (E_j, ValueThreshold)}
(set of topic importance value thresholds, one for each expert)

(Accept -Expert-)M(etalink-)Im(portance-Threshold) (U) = (E_i, ValueThreshold),…, (E_j, ValueThreshold)}
(set of metalink importance value thresholds, one for each expert)

(Accept -Expert-)S(ource-)Imp(ortance-Threshold) (U) = {(E_i, ValueThreshold),…, (E_j, ValueThreshold)}
(set of source importance value thresholds, one for each expert)

Reject-T(opic-Attribute) (U) = {Attribute=value | < E_i, Attribute=value>}
(set of topic attribute values to be rejected)

Reject-S(ource-Attribute) (U) = {Attribute=value| < E_i, Attribute=value>}
(set of source attribute values to be rejected)

(Expert-)Conflict-R(esolution) (U) = Ordered-Accept | Accept-All | Manual
(Accept advices in an ordered manner, as listed by the EXPERT function; or accept all advice, ignoring all conflicting advice; or let user choose which advice to accept in the case of a conflict)

We illustrate these preference functions with an example.

Example 5. Assume that we have three experts W-Clinton, A-Gore, and GW-Bush. The user John-Doe specifies the following preferences:

Expert (John-Doe) = <GW-Bush, W-Clinton>
(Accept the advices of GW-Bush and W-Clinton; reject any advice from A-Gore)
TImportance (John-Doe) = {(GW-Bush, 0.5), (W-Clinton, 0.9)}
(Accept the topics from GW-Bush if GW-Bush-assigned importance is above 0.5; accept the topics from W-Clinton if W-Clinton-assigned importance is above 0.9)
MImportance (John-Doe) = {(W-Clinton, 0.9)}
(Always accept the metalinks from GW-Bush; accept the metalinks from W-Clinton if W-Clinton-assigned importance is above 0.9)
SImportance (John-Doe) = {(GW-Bush, 0.5)}
(Always accept the sources from W-Clinton; accept the sources from GW-Bush if GW-Bush assigned importance is above 0.5)
Reject-T (John-Doe) = {name="*Lewinski*", <W-Clinton, Name="Gift-Taking">}
(Always reject topics with names containing the word "Lewinski" (regardless of the expert); reject advice from W-Clinton on a topic with name "Gift-Taking")
Reject-S (John-Doe)={Web-Address=www.dirtypolitics.com)

Conflict-R=Ordered-Accept

(Follow the order as specified by "expert": always accept the advice of GW-Bush; accept the advice of W-Clinton only when it does not conflict with the advice of GW-Bush)

3.3.2 User Knowledge

For a given user and a topic, the knowledge level of the user on the topic (zero, originally) is a certain detail level of that topic (and less than the maximum detail level attribute of the topic). The set U-Knowledge (U) = {(topic, detail-level-value)} contains users' knowledge on topics in terms of detail levels. As in other specifications, topics may be fully defined using the three key attributes TName, TType and TDomain, or they may be partially specified in which case the user's knowledge spans a set of topics satisfying the given attributes. We give an example.

Example 6. Assume that the user John-Doe knows topics with names "Racquetball" at an expert (3) level, and "Tennis" at a beginner (1) level, specified as

UKnowledge (John-Doe)={(TName="Racquetball", 3), (TName="Tennis", 1)}

Besides detail levels, we also keep the following history information for each topic source that the user has visited: web addresses (URLs) of topic sources, their first/last visit dates and number of times the source is visited. We use the information on user's knowledge while evaluating query conditions and computing topic closures, in order to reduce the size of the information returned to the user. In the absence of a user profile, the user is assumed to know nothing about any topic, i.e., the user's knowledge level about all topics is zero. Here in this report, we assume that the user profiles are XML documents with single/multiple user profiles.

4 Topic-Centric Query Language: SQL-TC

4.1 Overview and Basic Features

SQL-TC is an integrated SQL-like topic-centric language for querying web-based information resources, expert advice repositories and personalized user information. In this section, we outline the syntax of the language and its special constructs. Then, we illustrate the features of our language, with queries over information resources, incorporating multiple expert advice and personalized user information. The queries demonstrate the notion of closure computation on particular metalinks and search of information resources. Next we provide a few queries that are posed solely on the expert advice information in order to illustrate additional features of SQL-TC.

Note that the query language presented here is a multi-database (multiple expert advice and user profiles, and multiple information resources) query language with heterogeneous data access. SQL-TC is strongly typed with each variable having a well-defined type. Furthermore, the language is general enough to be operational on any underlying expert data model, as long as the model supports metadata objects and their attributes.

First, we specify the syntactic constructs of SQL-TC. The formal syntax in an extended Backus-Naur format is given in Appendix C.

select [*topic {.attribute}* | *metalink {.attribute}*] *as T*
from resources *XML: url1, …*
using experts *Topic Map1: url1, …* **as** E1, …
with user profile *XML: URL* **as** *U*
where (i) conditions on topics and metalinks of experts
(ii) content-based conditions on sources,

(iii) conditions on user profile information.
order by [topic] **importance**
stop after n **most important| when importance below** m
| **after** n **most important and when importance below** m

Variables are prefixed by $ symbol, constants are in quotes, and metalinks are in italics. **Stop after** clause is adapted from [13].

4.2 Querying Web-based Information Resources

In this section, we illustrate SQL-TC by example queries. We assume that we have two experts whose advices are at www.sql-tc.com/king.xtm (expert E1) and www.horror-books.com/books.xtm (expert E2), respectively. The information resources are at http://www.stephenkinglibrary.com and http://www.stephen-king.net. As the expert advice and the user profile information, we use the instances provided in Appendices A and B, respectively. For example, from Appendix B, the user U prefers to accept first the advice of expert E1, and then, if there are no conflicts, the advice of expert E2.

Example 7. (*Topic and source variables, and detail levels*) Using only the advice at www.sql-tc.com/king.xtm, find two highest-ranked novels that are written by the novelist Stephen King and the novels' detail level 4 reviews from the two information resources.

select [$topic.name, $sourceRef.web-address] **as** T
from resources http://www.stephenkinglibrary.com, http://www.stephen-king.net
using experts www.sql-tc.com/king.xtm **as** E
where *WrittenBy* **in** E.Metalinks **and**
$topic = **any** (*WrittenBy* ("Stephen King", "horror novelist", "literature", E,)) **and**
$sourceRef = SourceOf($topic, 4, E) **and**
"review" **in** $sourceRef.roles
order by $topic **importance**
stop after 2 **most important**

This query returns novel names (i.e., topic names) and the web addresses of their reviews (i.e., topic sources) from two information resources on the web. The result of the query is a 2-column table. The first atomic formula in the where clause states that *WrittenBy* is a metalink type declared by expert E. Assume that the metalink type *WrittenBy* has the signature:
WrittenBy(E): SetOf author → novel

In the second where clause statement, the variable $topic is instantiated by one of the novel entities returned by the *WrittenBy()* metalink where each selected novel is authored by the topic that has TName of "Stephen King", TType of "horror novelist" and TDomain of "literature", and specified by the expert E. This query illustrates two types of variables, namely, $topic which is a topic variable, and $sourceRef which is a topic source reference variable. SourceOf() is a function that takes in the triple <topic entity, detail level, expert>, and returns a set of topic source reference (TSRef) entities at the given detail level as specified by the given expert. Thus, in the above query, the value of $sourceRef.web-address expression is, according to expert E, the web addresses of topic sources at detail level 4 obtained from the topic reference entities for the topic $topic.

Please note that this query does not employ the user profile. Thus, using the expert advice in Appendix A, this query produces 4 tuples; however, only the two highest ranked tuples (one for Carrie with the importance value of 1, and another for the Stand with the importance value of 0.8) are returned as shown in Table1.

Table 1. Output of the SQL-TC query in Example 7

Tname	SourceRef.Web-address
“Carrie”	www.critics.com/carrie.html
“The Stand”	www.critics.com/stand.html

Example 8. (*Topic closure computation and user profiles*) Using only the advice of expert E and excluding the novels read by the user, find the highest ranked novel and its detail level 4 reviews where the novel is written by Stephen King and related to the novel “Wizard and The Glass”.

> **select** [$topic.name, $sourceRef.web-address] **as** T
> **from resources** http://www.stephenkinglibrary.com, http://www.stephen-king.net
> **using experts** www.sql-tc.com/king.xtm **as** E
> **with user profile** www.myprofile.com
> **where** *WrittenBy, RelatedTo* **in** E.Metalinks **and**
> $topic = **any** (*WrittenBy* ("Stephen King", ”horror novelist”, ”literature”, E,) **and**
> *RelatedTo** (“Wizard and The Glass”, ,”literature”, E,)) **and**
> $sourceRef = SourceOf($topic, 4, E) **and**
> “review” **in** $sourceRef.roles **and**
> $topic **not in** GetTopics(U.UserKnowledge)
> **order by importance**
> **stop after** 1 **most important**

We assume for this query that the metalink type *RelatedTo* of expert E has the signature

RelatedTo(E): novel → novel

Note that *RelatedTo*() metalink in this query uses the topic variable $topic, which in turn has all four key attributes (i.e., TName, Ttype, TDomain, TAuthor) specified. In this query the user asks for the highest-valued tuple, not the highest-valued novel. Derived importance value computation of output tuples [2] takes place, and the tuple in Table 2 is chosen.

Let us discuss the interpretation of this query using the expert repository and user profile instances in the Appendix. The novels that are related to the novel “Wizard and The Glass” are recursively located. From Appendix B, the output returns only those novels that are not known by the user. For instance, according to the expert advice in Appendix A, the topics that are related to “Wizard and The Glass“ are “The Wasteful Lands”, “Drawings of Three” and “Dark Tower”. However, since the novel “Dark Tower” is already known according to the user profile (given in Appendix B), it is not included in the final result, and the tuple (NOT the novel) with the highest importance value is selected.

Table 2. Output of the SQL-TC query in Example 8

Tname	SourceRef.Web-address
"The Wasteful Lands"	www.critics.com/dark3.html

Example 9. (*User preferences, user knowledge and multiple experts*) Using first the expert www.sql-tc.com/king.xtm, and then, if there are no conflicts, the expert www.horror-books.com/books.xtm, find all novels and their summaries such that the main characters of the selected novels are influenced from "Jack Park", the main characters of the novel "The Stand", and retrieve only those sources that have not been visited by the user in the last 30 days.

select [$topic.name, $sourceRef.web-address] **as** T
from resources http://www.stephenkinglibrary.com, http://www.stephen-king.net
using experts www.sql-tc.com/king.xtm **as** E1, www.horror-books.com/books.xtm **as** E2
with user profile www.myprofile.com as U
where *NovelsOfNovelCharacters*, *InfluencedBy* **in** (E1, E2). Metalinks **and**
$topic=**any** *NovelsOfNovelCharacters* (
*InfluencedBy** ("Jack Park", hero, novel characters, ,), ,) **and**
$sourceRef=**any** SourceOf($topic, ,) **and**
"summary" **in** $sourceRef.roles **and**
$sourceRef.web-address **in** GetSourceAddresses (U.UserKnowledge) **and**
GetLastVisitedDays(U.UserKnowledge, $sourceRef.web-address) > "30"

The second where clause assigns a novel to the topic variable $topic where the novel has a main character influenced by a main character of the novel "The Stand" in the domain of "literature". For both experts, we assume that the signatures of the metalink types *InfluencedBy* and *NovelsOfNovelCharacters* are the same, and each is defined as

InfluencedBy (E): novel-character → novel-character and
NovelsOfNovelCharacters (E): novel-character → SetOf novel

where E denotes either of the two experts. Note that, in the query, the selection of the expert for the above metalinks (and the expert of the function SourceOf()) is not specified in the query, and deferred to the user's preferences. Also, in the SourceOf() function, a topic source at any detail level is accepted.

For this example, we assume that the *InfluencedBy* metalink is binary, transitive, and cyclic, and we apply the corresponding topic closure computation algorithm for this case. According to the advice of expert www.sql-tc.com/king.xtm (E1 in Appendix A), the novel "The Stand" has the main character "Jack Park", who influences the character "John Smith". As "John Smith" is claimed to be the main character of the novels "Scream" and "Maniac" by expert www.horror-books.com/books.xtm (E2 in Appendix A), the topic closure computation will bind each of "Scream" and "Maniac" to the $topic variable. Thus, $sourceRef.web-addresses will be assigned to the corresponding sources www.books.com/scream.html and www.books.com/maniac.html. The function GetSourceAddresses() returns addresses of visited sources and the function GetLastVisitedDays() retrieves the days since the last-visit of a given source from the user profile database U (in Appendix B). Subsequently, the entire query will return www.books.com/maniac.html as it is the only source that is visited by the user and not in the last thirty days.

Note that as this query employs more than one expert advice, the issue of possible conflicts among different expert advice comes up. In the user preferences (given in Appendix B), first the advice of E1 and then, if there are no conflicts, the advice of E2 are to be accepted. Assume that the following metalink advice instances are encountered during the topic closure computation with respect to the *InfluencedBy* metalink type:

MAdvice(E1, "Jack Park" →*InfluencedBy* "John Smith") = 0.8

MAdvice (E2, “Jack Park” →*InfluencedBy* “John Smith”) = “No”

The query evaluation relies first on E1 and includes the character “John Smith” in the closure set, or relies on E2 and discards the character “John Smith” (and thus all other topics that may possibly be added to the closure because of the inclusion of “John Smith”) from the closure. To resolve the conflict, the query engine consults the metalink-importance-threshold statements declared in the user preferences, and discards the advice with a lower importance value than the given threshold. The user preferences (of Appendix B) declare threshold values 0.5 and “Don’t-Care”[1] for experts E1 and E2 respectively. And, the conflict-resolution statement of the user’s preferences declares an ordered acceptance of advices. Thus, we add “John Smith” into the topic closure set.

Example 10. (*Multilevel nested queries and the aggregation operators*) Find the advice, at any level, of the expert www.sql-tc.com/king.xtm for a novel written by Stephen King such that the novel has the highest importance value according to the reviews at www.king-review.com.

```
select [$sourceRef.web-address] as T
from resources http://www.stephenkinglibrary.com, http://www.stephen-king.net
using experts www.sql-tc.com/king.xtm as E1
where $topic in E1.Topics and
        $sourceRef=SourceOf($topic,,E1) and
        $topic.TName in
           select [$topic1.TName]
           using experts www.king-review.com as E2
           where WrittenBy in E2.Metalinks and
                 $topic1= any (WrittenBy (SimTName(“Stephen King”, “horror novelist”, “literature”, E2,)))
                 and TAdviceMatch (E2, $topic1) =
                     select max (TAdviceMatch (E2, $topic2))
                      where $topic2 in E2.Topics
```

The function TAdviceMatch (expert, topic) returns the expert’s importance value “matched” to topic. If no matching takes place, the value of 0 is returned.

In the above query, three levels of select-subquery nesting are employed. The innermost query finds the highest topic importance value for any topic, assigned by the expert www.king-review.com by using the aggregate operator **max** in a way similar to SQL. The second-level select subquery returns the names of Stephen King novels that are assigned the maximum importance value according to the expert www.king-review.com. And the outermost query finds all sources, at any level, for the selected novels, advised by the expert www.sql-tc.com/king.xtm.

Example 11. (*Text search in the source*) Find all sources, at any detail level, that are about Stephen King and contain the string "accident".

```
select [$sourceRef.web-address] as T
from resources http://www.stephenkinglibrary.com, http://www.stephen-king.net
using experts www.sql-tc.com/king.xtm as E
where $topic in E.Topics and $topic.TName = "Stephen King" and
        $sourceRef = SourceOf($topic,,E) and Contains($sourceRef, "accident")
```

[1] The threshold value Don’t Care indicates that all metalink instances are accepted from this particular expert, regardless of their importance value.

Assume that the information resource located at www.newsweek.com/17-10-2000/ mentions the topic "Stephen King" and contains the string "accident". In this case, the result of the query would be as in the Table 3.

Table 3. Output of the query in Example 11.

SourceRef.Web-address
"www.newsweek.com/17-10-2000/tragic-accident.html"

4.3 Querying Expert Advice Repositories

In this section, we provide two examples that solely query the expert advice repositories.

Example 12. (*Multiple metalink closures*) Find all topic names of any type, which are related to the novels based upon Stephen King novels.

select [$topic.TName] **as** T
using experts www.sql-tc.com/king.xtm as E
where *WrittenBy*, *RelatedTo*, *BasedUpon* **in** E.Metalinks **and**
$topic1 **= any** (*WrittenBy* ("Stephen King",,"literature", E,)) **and**
$topic **=any** ((*RelatedTo* (*BasedUpon*))* ($topic1.TName, $topic1.TType, $topic1.TDomain, E,))

For this query, we assume the following metalink signatures.
WrittenBy (E): SetOf author → novel
RelatedTo (E): topic → topic
BasedUpon (E): SetOf topic → topic

Note that novels are topics, and the *RelatedTo* and *BasedUpon* metalinks return novels, movies, musicals, and persons. In the above query, the result is a table with rows that contain topic names for each topic entity qualifying the conditions. As in the previous examples, the notation ((*RelatedTo* (*BasedUpon*))* enforce topic closure computation. The algorithm in Figure 2 represents the interpretation semantics for this case.

We now briefly discuss the query evaluation using the instances in the Appendix. According to the expert advice at www.sql-tc.com/king.xtm (i.e., expert E1 in Appendix A), the set of novels {"Carrie", "The Stand", "Wizard and The Glass", "The Wasteful Lands", "Drawings of Three", "Dark Tower"} is written by Stephen King. The input to the multi-closure algorithm is these topics and the metalink set M that contains metalinks *BasedUpon* and *RelatedTo*. At the first iteration of multi-closure computation, X^+ is augmented by the movie "Story of Carrie" which is *BasedUpon* the corresponding novel "Carrie". Then the closure set is further expanded by "John Carpenter", which is *RelatedTo* this movie. There is no contribution of other novels to the closure set. At the second iteration of the while loop, the algorithm first retrieves the topics that are based upon the current set, which has been extended by {"Story of Carrie", "John Carpenter"} in the previous step. As "Carrie: The Musical" is based upon the movie "Story of Carrie", it is added to X^+, and no further topic is based upon the topics in this set. Then, the topics that are related to "Carrie: The Musical" are added to the set, and "Broadway Season 1980" is the only such topic to be added. Again, no other topics are added to the closure set, as there is no topic that is RelatedTo the topics in current set. The algorithm stops when the closure does not change in the next iteration. The result of the query is given in Table 4.

Table 4. Output of the query in Example 12.

Tname
“Carrie”
“The Stand”
“Wizard and the Glass”
“Wasteful Lands”
“Drawings of Three”
“Dark Tower”
"Story of Carrie"
“John Carpenter”
“Carrie: The Musical”
“Broadway Season 1980”

Example 13. (*Metalink attributes*) Find top 30-ranked metalinks in the domain of literature and having an importance value of at least 0.7 for the expert www.sql-tc.com/king.xtm such that, in each such metalink, Stephen King is a participator.

> **select** [$metalink] **as** T
> **using experts** www.sql-tc.com/king.xtm **as** E
> **where** $metalink **in** E.Metalinks **and**
> $metalink**= any** (MetalinksWithTopic ("Stephen King", , , E)) **and**
> $metalink.domain = "literature"
> **order by importance**
> **stop after** 30 **most important and when importance below** 0.7

The function MetalinksWithTopic() takes a topic (either fully identified by TName, Ttype, TDomain, and TAuthor in the given order, or partially identified), and returns metalink instances.

The function MAdviceMatch () returns the importance value of a given metalink instance for the given expert. Using the expert advice in Appendix A, the query yields the output specified in Table 5.

Table 5. Output of the query in Example 13.

Mid	MType	MDomain	Antecedent players	Consequent players	M-Advice()
M1	*WrittenBy*	{"literature", "horror"}	[Stephen King, horror novelist, literature, -]	[Carrie, novel, literature, -]	1

5 Implementation Issues

5.1 System architecture

In Figure 3, we provide the architecture of our system in its most general form. In what follows, we briefly explain the key components of this system. Implementation-specific parts are described in the next section.

- *Visual query interface* is a graphical user interface provided to users. Although the examples illustrated up to this point are textual, the implementation, when finished, will provide the user a GUI to ease the use of SQL-TC in the web environment.
- *Expert advice access interface* provides a uniform set of interface functions for accessing expert advice repositories with many possible kinds of underlying models or implementations.
- *User profile database access interface* is similar to the above one, providing a uniform set of interface functions to access underlying heterogeneous databases.
- *Metadata filter* is intended to extract expert advice information embedded to resource documents and providing a compatible interface to the expert advice access interface.

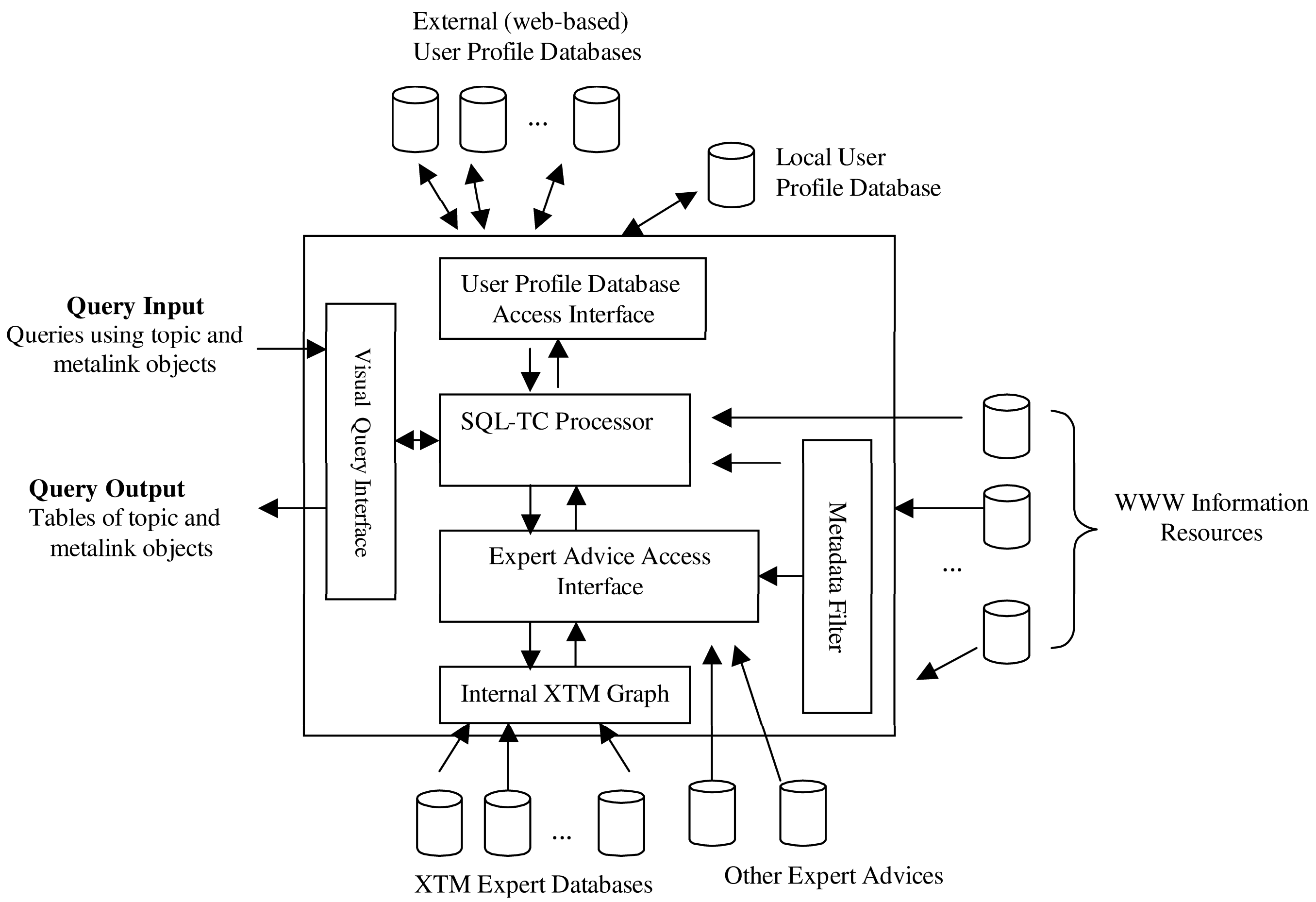

Figure 3. SQL-TC System Architecture

5.2 Prototype implementation

We have implemented a prototype system on the PC-Windows platform using Java 2 SDK™, Sun Java servlet engine and a web server provided within JSWDK™ (JavaServer Web Development Kit). In the prototype system, XTMs serve as expert advice repositories. Thus, the prototype system includes some of the core components described above along with XTM specific tools, which we discuss next.

- *Visual query interface* is implemented as a Java *servlet*, which may be accessed at [36]. When a user is first connected to our site, (s)he has the option to login in order to access local user profile (Figure 4). Then, the user specifies the address of the web domain(s) to be queried and/or one or more XTM expert advice repositories. In the restricted demo in [36], we provide a (virtual) domain and an XTM document of Appendix A as defaults.

Next, query design page (Figure 5) is presented where the user may pose queries by using the topic and metalink objects available in the specified advice repositories. At the moment, we provide a web page with a query template that allows designing queries of one particular type. Providing a functional and user-friendly GUI supporting all features of SQL-TC will need more work, and we will report the results later.

In the prototype demo [36], we return query outputs as interactive tables, where the metadata objects and/or the sources in the tables are hyperlinks that may be followed and navigated.

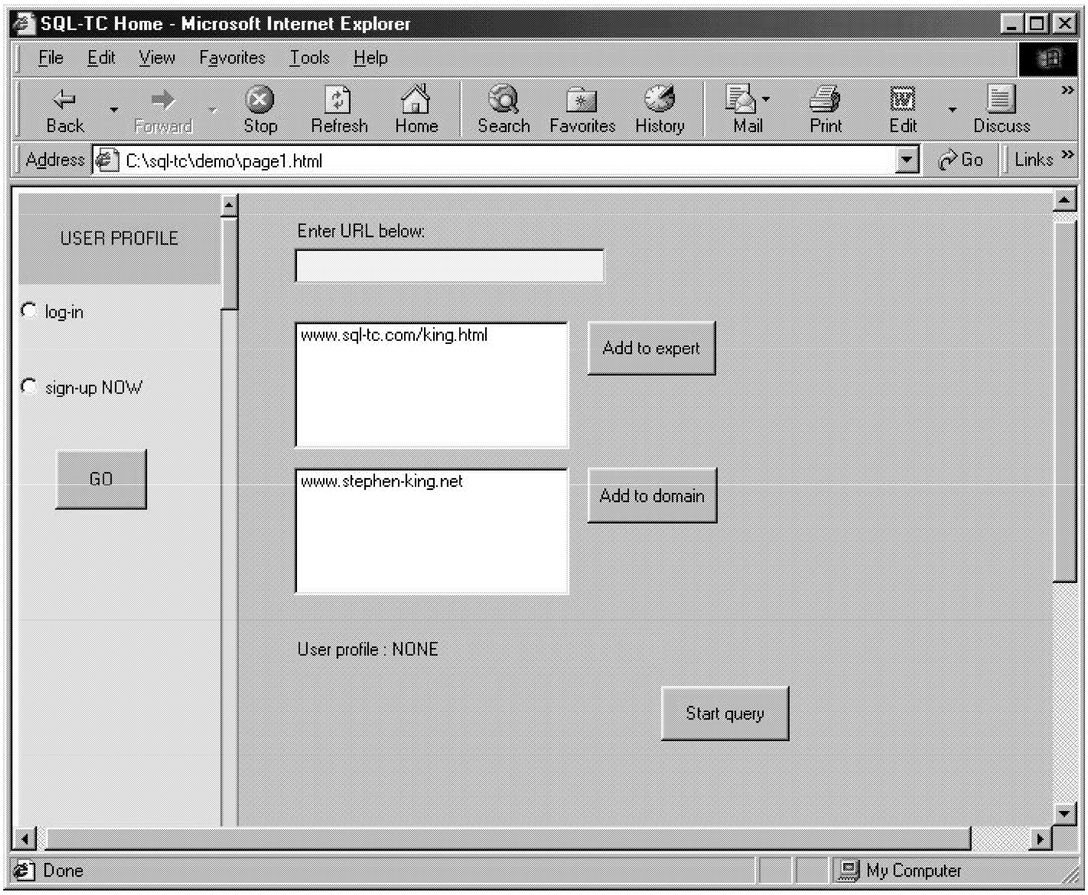

Figure 4. Visual Query Interface

- *XTM Processor v1.0*: This is a fully conformant implementation of [44] that parses an XTM document and creates internal topic map graph, which is said to be "*reconstituting the information captured in the XTM syntax"* [44]. At the moment, the graph is stored in the memory using Java Hashtable data objects, but we are also investigating to store it in an actual database to be able to handle XTMs including large volumes of data.

- *XTM Access Functions*: This is actually a set of Java functions which implements the expert advice access functions, mentioned in Section 5.1, for a specific expert advice database type, namely XTM. Note that, the interface mechanism supported by Java enables us to declare an interface class, which may be implemented by many other actual classes for different types of expert databases, and then whenever such an interface function is called, the appropriate function implementation will be bound according to the expert database type [21]. Thus, we can add new expert database models to our system by just providing the implementation of access functions declared in the interface.

In the prototype system, the user profile access interface and the metadata filter are not yet implemented, and the query processing algorithms are currently being studied.

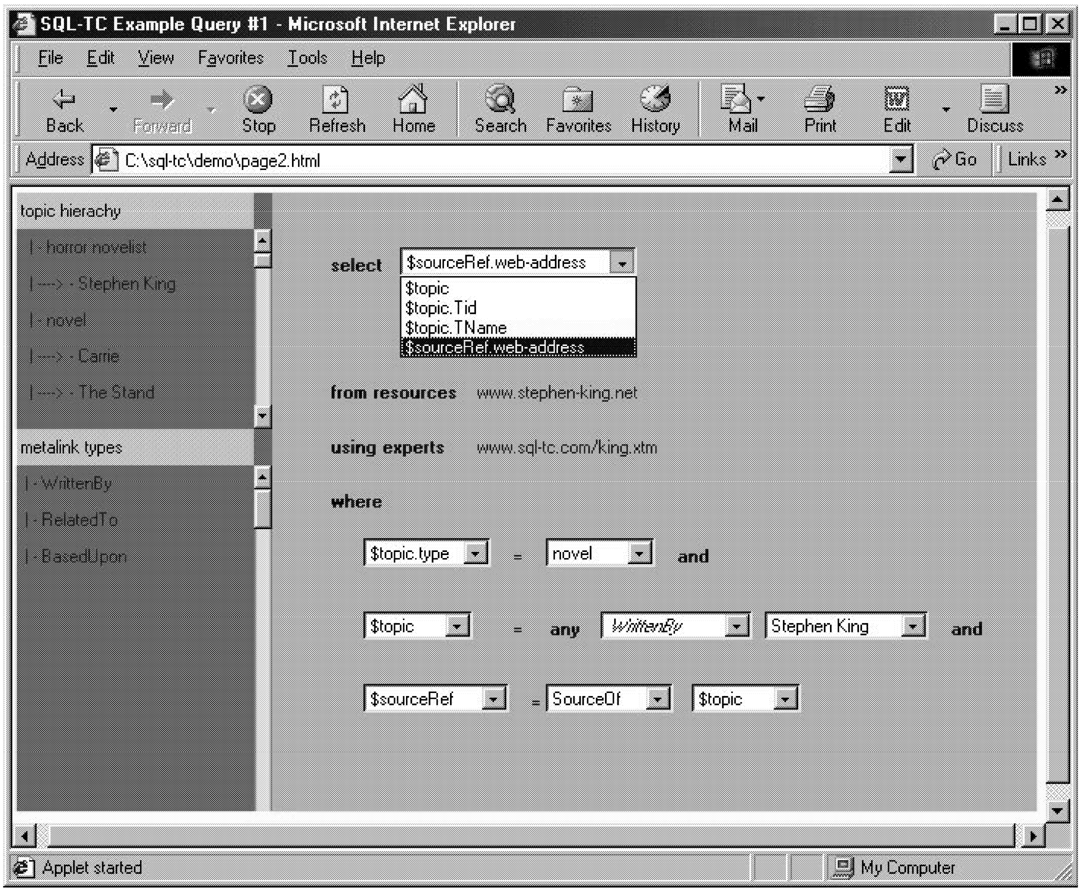

Figure 5. Query Design Interface

6 Conclusion

In this study, we develop a web information space model and its query language to allow sophisticated queries/searches over the web resources. Proposed information model has three major components: (i) information resources that are representing the web-based documents, (ii) expert advice model that is contributing to a topic-centric knowledge index over the resources, and (iii) personalized information model that is capturing the user preferences and knowledge. In this study, information resources are assumed to be HTML/XML documents, and the user profile is captured in XML documents. XTMs are used to serve as expert advice repositories that are imposed over information resources. Expert advice repositories contribute to a semantic index as they identify the topics and their relationships in a resource set, and provide links to the actual occurrences of topics in these resources. We make the practical assumption that the web resource domains associated with expert advice repositories do not span the web, although they may be arbitrarily large.

The SQL-like query language SQL-TC is designed to operate over this information space and query the information resources by incorporating the expert advices and users' personalized information. SQL-TC queries are expressed using topics, metalinks and their sources, and they are capable of querying both the associated information resources and the expert advice repositories themselves in an integrated manner. In this sense, SQL-TC provides multi-database access as well as heterogeneous data access. One distinguishing feature of SQL-TC is the concept of topic closures, which retrieves a set of all topics that are logically implied by the queried topic(s) with respect to a particular metalink type. Pruning topic closure sets by consulting user knowledge- that is kept in the user profile- is possible, which may otherwise grow too large. As SQL-TC is capable of using multiple expert advice repositories simultaneously along with personalized information databases, the issue of expert advice conflicts and their resolution is important. In this report, we discuss a user-preference based resolution approach, which may further be improved in the future.

Finally, the report presents the system architecture and a prototype implementation using XTMs as expert advice repositories. In this sense, our work serves as a real-life application of XTM effort, and may also contribute to identify and resolve many issues with the currently evolving XTM standard.

The future work will include query processing and optimization for SQL-TC queries.

References

[1] Abiteboul, S., Buneman, P., and Suciu, D., “Data on the Web”, Morgan Kaufmann Publishers, San Francisco, 258p., 2000.

[2] Abiteboul, S., Quass, D., McHugh, J., Widom, J., Wiener, J., “The Lorel query language for semistructured Data”, Jour. of Digital Libraries, 1998.

[3] ACM SIGMOD Workshop on The Web and Databases (WebDB'99), June 3-4, 1999, available at http://www.acm.org/sigmod/dblp/db/conf/webdb/webdb1999.html#LudascherPV99

[4] Agrawal, R., Wimmers, E.L., “A Framework for Expressing and Combining Preferences”, ACM SIGMOD Conf., pp. 297-306, 2000.

[5] Berners-Lee, T., “Semantic Web Roadmap”, W3C draft, Jan 2000, available at http://www.w3.org/DesignIssues/Semantic.html

[6] Biezunski, M., “Topic Maps at a glance”, at http://www.infoloom.com/tmsample/bie0.htm

[7] Biezunski, M, "A Topic Map of This Conference's Proceedings", Proc. of GCA, 1996, http://www.infoloom.com/IHC96/mb214.htm

[8] Biezunski, M., Bryan, M., Newcomb, S., editors, ISO/IEC 13250, Topic Maps, available at http://www.ornl.gov/sgml/sc34/document/0058.htm.

[9] Bosworth, A., Brown Jr., A. L., “Microsoft's vision for XML”, IEEE Data Engineering Bulletin 22(3): 35-43, 1999.

[10] Bray, T., Paoli, J., Sperberg-McQueen, C. M., “Extensible Markup Language 1.0 Specification”. World Wide Web Consortium (W3C), February 1998.

[11] Buneman, P., Davidson, S., Hillebrand, G., Suciu, D., “A Query Language and Optimization Techniques for Unstructured Data”, ACM SIGMOD Conf., 1996.

[12] Buneman, P., “Semistructured data”, ACM PODS Conf., 1997.

[13] Carey, M.J., Kossmann, “On Saying “Enough Already” in SQL”, ACM SIGMOD 1997.

[14] Deutsch, A., Fernandez, M., Florescu, D., Levy, A., Suciu, D., “XML-QL: A Query Language for XML”, submission to W3C, Aug. 1998, available at http://www.w3.org/TR/NOTE-xml-ql

[15] Deutsch, A., Fernandez, M. F., Suciu, D., “Storing Semistructured Data with STORED”, SIGMOD Conference 1999.

[16] ECDL Workshop on the Semantic Web, Sept. 21, 2000, Lisbon, Portugal, electronic proceedings available at http://www.ics.forth.gr/proj/isst/SemWeb

[17] Fernandez, M., Florescu, D., Levy, A., Suciu, D., “A Query Language for a web-site management System”, SIGMOD Record, Sept. 1997.

[18] Florescu, D., Levy, A., Mendelzon, A., “Database Techniques for the World-Wide Web: A Survey”, ACM SIGMOD Record, 27, 3, Sept. 1998.

[19] Gruser, J-R., Raschid, L., Vidal, M.E., Bright, L., “Wrapper Generation for Web-Accessible Resources” CoopIS Conf., 1998.

[20] ISO 13250 Topic Map Standard available at http://www.w3c.com
[21] Java Documentation at http://java.sun.com
[22] Ksiezyk, R., "Answer is Just a Question [of matching Topic Maps], Proc. of XML Europe 2000, GCA, Alexandria, VA, 2000.
[23] Layman, A., et al, "XML-Data", Proposal for a Standard, World Wide Web Consortium, Jan. 1998.
[24] Levy, A., Rajaraman, A., Ordille, J.J., "Querying Heterogeneous Information Sources using source descriptions", VLDB Conf., 1996.
[25] Mendelzon, A., Mihaila, G., and Milo, T., "Querying the WWW", International Journal on Digital Libraries", Vol.1, No.1, pp.54-67, April 1997.
[26] Mihaila, G. A., Raschid, L., Tomasic, A., "Equal Time for Data on the Internet with Web Semantics", EDBT Conf., 1998.
[27] Microsoft MSDN Online Support, available at http://support.microsoft.com/servicedesks/msdn.
[28] Microsoft MoneyCentral, available at http://moneycentral.msn.com/home.asp
[29] Newcomb, S., Biezunski, M., "Topic Maps go XML", XML Europe 2000, June 2000.
[30] Ozsoyoglu, G., Balkir, N.H., Cormode, G., Ozsoyoglu, Z.M., "Electronic Books in Digital Libraries", IEEE Advances in Digital Libraries Conf., Washington, D.C., May 2000.
[31] Ontopia Topic Map Technology, available at www.ontopia.net
[32] Online Collections of the Smithsonian Institution, available at http://www.si.edu.
[33] Pepper, S., "Euler, Topic Maps, and Revolution", available at http://www.infoloom.com/tmsample/pep4.htm
[34] Rath, H.H., Pepper, S., "Topic Maps at Work", Goldfarb, C.F., Prescod P. (eds): XML Handbook, 2nd edition, Prentice Hall, 1999.
[35] Rath, H.H., "Topic Maps Self Control", Extreme Markup Languages 2000, Montreal, PQ, Canada, 2000.
[36] SQL-TC servlet, available at http://139.179.21.49:8080/servlet/SQL-TC
[37] The K42 Topic Map Engine, available at http://k42.empolis.co.uk
[38] Topic Maps available at www.infoloom.com
[39] Topic Map Samples, available at http://www.techquila.com/tmsamples/
[40] tmproc: A Topic Maps implementation, available at http://www.ontopia.net/software/tmproc/index.html
[41] XML Linking Language (XLink) version 1.0, available at http://www.w3.org/TR/xlink/
[42] XML Topic Maps (XTM) 1.0 available at http://www.topicmaps.org/xtm/1.0
[43] XTM: XML Topic Maps- working documents, available at http://www.doctypes.org/xtm/home.html
[44] XTM Processing Model 1.0, TopicMaps.Org AG Review Specification, 4 Dec 2000, available at http://topicmaps.org/xtm/1.0/xtmp1.html
[45] Ullman, J.D., "Principles of Database and Knowledge-Base Systems", Vol. 1, 1988.

Appendix A: Expert Advice Repositories

In the following, we provide expert advices in a tabular form for the ease of illustration. Clearly, the expert advice repositories may be in the form of text files, XML files and/or any conventional databases. Note that, as we use XTMs to serve as expert advice repositories in the prototype implementation, an XTM document corresponding to some of the following tabular forms is available in Appendix D. In order to save space, we only provide a portion of XTMs that are used as expert advices.

In this report, we provide examples with expert advice repositories that are assumed to be available at the virtual web locations www.sql-tc.com/king.xtm and www.horror-books.com/books.xtm. The former web site provides expert advices about the horror writer Stephen King and includes a number of concepts and relationships about Stephen King himself, his books, movies etc., and provides pointers to various kinds of resources for these concepts (topics) or relationships (metalinks)[2]. The latter web site is more general including expert advices about horror writers, books, characters of books, etc.

To save space, we only provide topics, sources and metalinks that are illustrated in the examples throughout the report. Top-level topic types such as person, novel, movie, musical, horror novelist, etc.

[2] Please note that, most of the novels, characters, web locations etc. given in this Appendix and in the examples throughout the report are imaginary.

are examples of topics that are not explicitly specified as topics in the tables given below. Similarly, metalink signatures that are mentioned in Section 3 are also not included. In case of the XTMs, they are assumed to be in XTM templates.

A.1 Expert Advice provided in www.sql-tc.com/king.xtm (Expert E1)

Table A.1 Topics of Expert E1

Tid	TDetail level	TType	Tname	TDomain	T-Advice	Source
T1	-	horror novelist	"Stephen King"	literature	1	{S1, S2}
T2	-	novel	"Carrie"	literature	1	{S3, S4}
T3	-	novel	"The Stand"	literature	0.8	{S5, S6}
T4	-	movie	"Story of Carrie"	cinema	0.7	-
T5	-	musical	"Carrie: The Musical"	leisure	0.2	-
T6	4	novel	"Wizard and The Glass"	literature	0.3	-
T7	3	novel	"The Wasteful Lands"	literature	0.4	S7
T8	2	novel	"Drawings of Three"	literature	0.6	S8
T9	1	novel	"Dark Tower"	literature	0.8	-
T10	-	movie director	"John Carpenter"	cinema		-

Note that, in the below table, the attributes may be set-valued. A player topic in a metalink instance is identified with the triple TName, TType, TDomain along with the topic detail level. For instance, the topic "Carrie", of type "novel" in the domain "literature" and with no specified detail level, is one of the player topics in the metalink instance M1 of type "*WrittenBy*". Note that, the player topics may be simply referred by their internal ids in the implementation, and we fully specify them in the above table for the sake of illustration.

Table A.2 Metalinks of Expert E1

Mid	Mtype	Mdomain	Antecedent players	Consequent players	M-adv
M1	*WrittenBy*	{literature, horror}	[Stephen King, horror novelist, literature, -]	[Carrie, novel, literature, -]	1
M2	*WrittenBy*	{literature, horror}	[Stephen King, horror novelist, literature, -]	[The Stand, novel, literature, -]	0.6
M3	*WrittenBy*	{literature, horror}	[Stephen King, horror novelist, literature, -]	[Wizard and The Glass, novel, literature, -]	0.6
M4	*WrittenBy*	{literature, horror}	[Stephen King, horror novelist, literature, -]	[The Wasteful Lands, novel, literature, -]	0.6
M5	*WrittenBy*	{literature, horror}	[Stephen King, horror novelist, literature, -]	[Drawings of Three, novel, literature, -]	0.6
M6	*WrittenBy*	{literature, horror}	[Stephen King, horror novelist, literature, -]	[Dark Tower, novel, literature, -]	0.6
M7	*BasedUpon*	-	[Carrie, novel, literature, -]	[Story of Carrie, movie, cinema, -]	0.6
M8	*BasedUpon*	-	[Story of Carrie, movie, cinema, -]	[Carrie: The Musical, musical, -]	
M9	*RelatedTo*	-	[The Wasteful Lands, novel, literature, 3]	[Wizard and The Glass, novel, literature, 4]	0.3
M10	*RelatedTo*	-	[Drawings of Three novel, literature, 2]	[The Wasteful Lands, novel, literature, 3]	0.5
M11	*RelatedTo*	-	[Dark Tower, novel, literature, 1]	[Drawings of Three novel, literature,2]	0.6
M12	*NovelCharacters*	-	[The Stand, novel, literature, 1]	[Jack Park, hero, novel characters, -]	0.7
M13	*InfluencedBy*	-	[Jack Park, hero, novel characters, -]	[John Smith, character, novel characters, -]	0.8
M14	*RelatedTo*	-	[Story of Carrie, movie, cinema, -]	[John Carpenter, movie director, cinema, -]	-
M15	*RelatedTo*	-	[Carrie: the Musical, musical, leisure, -]	[Broadway Season 1980, musical, leisure, -]	-

Table A.3 Sources of Expert E1

Sid	Web-address	Role	MediaType	LastUpdated	Detail level	S-Advice
S1	www.king.com/	Website	multimedia	16.01.2001	-	1
S2	www.newsweek.com/17-10-2000/tragic-accident.html	Mentions	Text	20.01.2001	-	0.7
S3	www.books.com/carrie.html	Summary	Text	-	-	0.5
S4	www.critics.com/carrie.html	Review	Text	-	4	0.8
S5	www.books.com/stand.html	Summary	Text	-	-	0.4
S6	www.critics.com/stand.html	Review	Text	-	4	0.7
S7	www.critics.com/dark3.html	Review	Text	-	4	0.8
S8	www.critics.com/dark2.html	Review	Text	-	4	0.3

A.2 Expert Advice provided in www.horror-books.com/books.xtm (Expert E2)

Table A.4 Topics of Expert E2

Tid	TDetail level	TType	Tname	Tdomain	T-Advice	Source
T11	-	novel	"Scream"	Literature	0.3	S9
T12	-	novel	"Maniac"	Literature	0.4	S10
T13	-	hero	"Jack Park"	Novel characters	-	-
T14	-	character	"John Smith"	Novel characters	-	-

Table A.5 Metalinks of Expert E2

Mid	MType	MDomain	Antecedent players	Consequent players	M-Advice
M16	*InfluencedBy*	-	[Jack Park, hero, novel characters, -]	[John Smith, character, novel characters, -]	No
M17	*NovelsOfNovelCharacters*	-	[John Smith, character, novel characters, -]	[Scream, novel, literature, -]	0.6
M18	*NovelsOfNovelCharacters*	-	[John Smith, character, novel characters, -]	[Maniac, novel, literature, 1]	0.2

Table A.6 Sources of Expert E2

Sid	Web-address	Role	MediaType	LastUpdated	Detail level	S-Advice
S9	www.books.com/scream.html	Summary	Text	12.02.2001	-	0.6
S10	www.books.com/maniac.html	Summary	Text	13.02.2001	-	0.7

Appendix B: Personalized Information for User U

In the following, we provide personalized information in terms of user preferences and user knowledge for a typical user U. Assume that user profile is kept in the virtual web location www.myprofile.com.

User-Preferences (U) contains a set of statements as follows:

Expert (U) = <www.sql-tc.com/king.xtm, www.horror-books.com/books.xtm>

TImportance(U) = {(www.sql-tc.com/king.xtm, 0.5),
(www.horror-books.com/books.xtm, 0.3)}

Mimportance(U) = {(www.sql-tc.com/king.xtm, 0.5),
(www.horror-books.com/books.xtm, "Don't-Care")

Simportance(U) = {(www.sql-tc.com/king.xtm, 0.5),
(www.horror-books.com/books.xtm, 0.3)

Reject-S (U) = {www.sking-fanatics.com}

Conflict-R (U) = Ordered-Accept

User-Knowledge (U)

TName	Detail level	Source address	Source role	Source media type	First visit	Last visit	Visit No
Scream	-	www.books.com/scream.html	summary	text	-	12.02.2001	2
Maniac	-	www.books.com/maniac.html	summary	text	-	13.02.1999	3
Dark Tower	1	www.books.com/dark1.html	review	text	-	-	3

Appendix C. BNF for SQL-TC

<SQL-TC_query> :=<select_clause>
<from_clause>
<where_clause>
[<order_by_clause>]
| <SQL-TC_query> **union** <SQL-TC_query>
| <SQL-TC_query> **minus** <SQL-TC_query>
| <SQL-TC_query> **intersect** <SQL-TC_query>

<select_clause> := **select [** <topic_metalink_source_variable> **]** [**as** <variable>]
| **select** <aggregate_function>

<topic_metalink_source_variable> := <a_variable>[**.**<attribute>]
| <a_variable>[**.**<attribute>]**,** <topic_metalink_source_variable>

<a_variable> := **$**<identifier>
<attribute> := <identifier>
<variable> := <letter>
<underscore> := _
<identifier> := <letter>{<letter> | <digit> | <underscore>}
<letter> := **A**..**Z** | **a**..**z**
<digit> := **0**..**9**

<aggregate_function> := <aggregate_function_name>(<argument>)
<aggregate_function_name> := **max** | **min** | **avg** | **sum** | **count**
<argument> := <a_variable>[**.**<attribute>]
| <importance_functions>

<importance_functions>:=
TAdviceMatch(<expert_var>**,**(<a_variable>[**.**<attribute>] |
<string_constant>)) |
MAdviceMatch(<expert_var>**,**(<a_variable>[**.**<attribute>] |
<string_constant>)) |
SAdviceMatch(<expert_var>**,**(<a_variable>[**.**<attribute>]| <url>))

<built_in_function> := <importance_functions> |
GetSourceAddresses(<profile_var>**.UserKnowledge**) |
GetLastVisitedDays(<profile_var>**.UserKnowledge,** <a_variable>**.web-address**) |
Contains(<a_variable>,<string_constant>) |
MetalinksWithTopic(<topic_name>**,**<topic_type>**,**<topic_domain>**,**
<expert_name>)

<from_clause> := [**from resources** <url_list>]
[**using experts** <expert_list>]
[**with user profile** <profile_list>]

<url_list> := <url> | <url>**,**<url_list>
<url> := a valid URL

<expert_list> := <url> [**as** <expert_var>]
| <url> [**as** <expert_var>]**,**<expert_list>

<expert_var> := **E**{<digit>}

<profile_list> := <url> [**as** <profile_var>]
| <url> [**as** <profile_var>]**,**<profile_list>

<profile_var> := **U**{<digit>}

<where_clause> := **where** <variable_declaration_clause> **and** <condition_clause>

<variable_declaration_clause> := <topic_variable_declaration>[**and**
<variable_declaration_clause>]
|<metalink_variable_declaration>[**and**
<variable_declaration_clause>]

<topic_variable_declaration> := <topic_var>{**,** <topic_var>} **in** <expert_var>**.Topics**
| <topic_var>{**,** <topic_var>} **in (**<expert_var>{**,**<expert_var>}**).Topics**

<metalink_variable_declaration> := (<metalink_var>|<metalink_type>){**,**
(<metalink_var>|<metalink_type>)} **in** <expert_var>**.Metalinks**
| (<metalink_var>|<metalink_type>){**,** (<metalink_var>|<metalink_type>)} **in**
(<expert_var>{**,**<expert_var>}**).Metalinks**

<topic_var> := <a_variable>
<metalink_var> := <a_variable>
<metalink_type> := <identifier> | <identifier>*****

<condition_clause> := <metalink_closure_clause>
| <source_of_clause>
| <a_condition_clause>
| <SQL-TC_query>
| <condition_clause> **and** <condition_clause>
| <condition_clause> **or** <condition_clause>

<metalink_closure_clause> := <topic_var> **=any (** <metalink_closure_expression> **)**

<metalink_closure_expression> :=
<metalink_closure_expression> **or** <metalink_closure_term>
| <metalink_closure_term>

<metalink_closure_term> := <metalink_closure_term> **and** <metalink_closure_factor>
| <metalink_closure_factor>

<metalink_closure_factor> := **(**<metalink_closure_expression>**)**

```
                    | <metalink_closure_element>

<metalink_closure_element> := [<bool_not>] <metalink_type>(<closure_arguments>)
                        | ( [<bool_not>] <metalink_type>(<multi_metalink_closure>))*

<multi_metalink_closure>  :=  [<bool_not>] <metalink_type>(<multi_metalink_closure>)
                                      | [<bool_not>] <metalink_type>

<closure_arguments>  := <topic_name>, <topic_type>, <topic_domain>,
                             <expert_name>, <detail_level>
                            | ( [<bool_not>] <metalink_type> | SimTName) (<closure_arguments>)

<topic_name>  := "<identifier>" | ∈
<topic_type>  := "<identifier>" | ∈
<topic_domain>  := "<identifier>" | ∈
<expert_name>  := <expert_var> | ∈
<detail_level> := <digit>{<digit>} | ∈

<source_of_clause>  :=
                    <source_var> =any SourceOf(<topic_var>, <detail_level>, <expert_name>)

<source_var>  := $<identifier>
<detail_level>  := {<digit>}

<a_condition_clause>  :=  <topic_metalink_source_variable> <set_op> <built_in_function>
                             | <topic_metalink_source_variable> = <constant>
                             | <topic_metalink_source_variable> <set_op> <SQL-TC_query>
                             | <built_in_function> <op> <constant>
                             | <built_in_function> <set_op> <SQL-TC_query>
                             | <built_in_function>
                             | <constant> <set_op> <topic_metalink_source_variable>

<op>  :=   = | <  |  <=  |  >  |  >=
<set_op>  := <op> | in | =any | =all
<bool_not> := not
<constant>  := <string_constant> | <numeric_constant>
<string_constant> := "{<digit> | <letter> | <underscore>}"
<numeric_constant> := <digit>{<digit>}[. | ,]{<digit>}

<order_by_clause> := order by [<topic_var>] importance
                      ( stop after {<digit>} most important  |
                        when importance below <numeric_constant> |
                        after {<digit>} most important and
                             when importance below <numeric_constant>
                      )
```

Appendix D. XTM for Expert Advice Repository

In the following, we provide a "pruned" version of a topic map of expert E1 in Appendix A. In this map,

- Top level (meta) topic types: horror novelist, novel, and movie.
- Top level (meta) topic and metalink domains: literature, cinema, and horror.
- Top level (meta) source (occurrence) roles: website, mentions, summary, and review.
- Metalink (association) types: writtenby, and based-upon.

This map includes a number of concepts and relationships about Stephen King's books, movies etc., and provides pointers to various kinds of resources for these concepts (topics) or relationships (associations).

```
<?xml version="1.0" encoding="iso-8859-1"?>
<!DOCTYPE topicMap SYSTEM "xtm1.dtd">
<topicMap xmlns="http://www.topicmaps.org/xtm/1.0/"
          xmlns:xlink="http://www.w3.org/1999/xlink">
     <topic id="MT1">
             <baseName>
                   <baseNameString> horror novelist</baseNameString>
             </baseName>
     </topic>
     <topic id="MT2">
             <baseName>
                   <baseNameString> novel</baseNameString>
             </baseName>
     </topic>
     <topic id="MT3">
             <baseName>
                   <baseNameString> movie</baseNameString>
             </baseName>
     </topic>
     <topic id="MT4">
             <baseName>
                   <baseNameString> literature<baseNameString>
             </baseName>
     </topic>
     <topic id="MT5">
             <baseName>
                   <baseNameString> cinema</baseNameString>
             </baseName>
      </topic>
      <topic id="MT6">
             <baseName>
                   <baseNameString> horror</baseNameString>
             </baseName>
      </topic>
      <topic id="MT7">
             <baseName>
                   <baseNameString> website</baseNameString>
             </baseName>
      </topic>
      <topic id="MT8">
             <baseName>
                   <baseNameString> mentions</baseNameString>
             </baseName>
      </topic>
      <topic id="MT9">
             <baseName>
                   <baseNameString> summary</baseNameString>
             </baseName>
      </topic>
```

```
<topic id="MT10">
	<baseName>
		<baseNameString> review</baseNameString>
	</baseName>
</topic>
<topic id="MT11">
	<baseName>
		<baseNameString> WrittenBy</baseNameString>
	</baseName>
</topic>
<topic id="MT12">
	<baseName>
		<baseNameString> BasedUpon</baseNameString>
	</baseName>
</topic>
<topic id="MT13">
	<baseName>
		<baseNameString> antecedent</baseNameString>
	</baseName>
</topic>
<topic id="MT14">
	<baseName>
		<baseNameString> consequent</baseNameString>
	</baseName>
</topic>
<topic id="MT15">
	<baseName>
		<baseNameString> HasTAdvice</baseNameString>
	</baseName>
</topic>
<topic id="MT16">
	<baseName>
		<baseNameString> HasMAdvice</baseNameString>
	</baseName>
</topic>
<topic id="MT17">
	<baseName>
		<baseNameString> HasSAdvice </baseNameString>
	</baseName>
</topic>
<topic id="MT18">
	<baseName>
		<baseNameString> 1 </baseNameString>
	</baseName>
</topic>
<topic id="MT19">
	<baseName>
		<baseNameString> 0.8</baseNameString>
	</baseName>
</topic>
<topic id="MT20">
	<baseName>
		<baseNameString> 0.7 </baseNameString>
	</baseName>
</topic>
<topic id="MT21">
	<baseName>
		<baseNameString> 0.6 </baseNameString>
	</baseName>
</topic>
<topic id="MT22">
	<baseName>
		<baseNameString> 0.5 </baseNameString>
	</baseName>
</topic>
<topic id="MT23">
```

```
            <baseName>
                  <baseNameString> hasSDetailLevel </baseNameString>
            </baseName>
      </topic>
      <topic id="MT24">
            <baseName>
                  <baseNameString> 4 </baseNameString>
            </baseName>
      </topic>
      <topic id="MT25">
            <baseName>
                  <baseNameString> source </baseNameString>
            </baseName>
      </topic>
      <topic id="MT26">
            <baseName>
                  <baseNameString> source detail level
      </baseNameString>
            </baseName>
      </topic>

      <topic id="T1">
            <instanceOf>
                  <topicRef xlink:href="#MT1"/>
            </instanceOf>
            <subjectIdentity>
                  <subjectIndicatorRef
                        xlink:href="http://www.stephenking.com"/>
            </subjectIdentity>
            <baseName>
                  <baseNameString>Stephen King</baseNameString>
                  <scope>
                     <topicRef xlink:href="#MT4"/>
                  </scope>
            </baseName>
            <occurrence id="S1">
                  <instanceOf>
                        <topicRef xlink:href="#MT7"/>
                  </instanceOf>
                  <resourceRef xlink:href="http://www.king.com"/>
            </occurrence>
            <occurrence id="S2">
                  <instanceOf>
                        <topicRef xlink:href="#MT8"/>
                  </instanceOf>
                  <resourceRef xlink:href="http://www.newsweek.com/17-10-2000/tragic-
accident.html "/>
            </occurrence>
      </topic>

      <topic id="T2">
            <instanceOf>
                  <topicRef xlink:href="#MT2"/>
            </instanceOf>
            <baseName>
                  <baseNameString>Carrie</baseNameString>
                  <scope>
                     <topicRef xlink:href="#MT4"/>
                  </scope>
            </baseName>
            <occurrence id="S3">
                  <instanceOf>
                        <topicRef xlink:href="#MT9"/>
                  </instanceOf>
                  <resourceRef
                      xlink:href="http://www.books.com/carrie.html"/>
```

```
            </occurrence>
            <occurrence id="S4">
                  <instanceOf>
                        <topicRef xlink:href="#MT10"/>
                  </instanceOf>
                  <resourceRef
xlink:href="http://www.critics.com/carrie.html "/>
            </occurrence>
      </topic>

      <topic id="T4">
            <instanceOf>
                  <topicRef xlink:href="#MT3"/>
            </instanceOf>
            <baseName>
                  <baseNameString>Story of Carrie</baseNameString>
                  <scope>
                     <topicRef xlink:href="#MT5"/>
                  </scope>
            </baseName>
      </topic>

      <association id="M1">
            <instanceOf>
                  <topicRef xlink:href="#MT11"/>
            </instanceOf>
            <scope>
               <topicRef xlink:href="#MT4"/>
                <topicRef xlink:href="#MT6"/>
           </scope>
            <member>
                  <roleSpec>
                        <topicRef xlink:href="#MT13"/>
                  </roleSpec>
                  <topicRef xlink:href="#T1"/>
            </member>
            <member>
                  <roleSpec>
                        <topicRef xlink:href="#MT14"/>
                  </roleSpec>
                  <topicRef xlink:href="#T2"/>
            </member>
      </association>

      <association id="M7">
            <instanceOf>
                  <topicRef xlink:href="#MT12"/>
            </instanceOf>
            <member>
                  <roleSpec>
                        <topicRef xlink:href="#MT13"/>
                  </roleSpec>
                  <topicRef xlink:href="#T2"/>
            </member>
            <member>
                  <roleSpec>
                        <topicRef xlink:href="#MT14"/>
                  </roleSpec>
                  <topicRef xlink:href="#T4"/>
            </member>
      </association>

     <association id="MM1">
            <instanceOf>
                  <topicRef xlink:href="#MT15"/>
            </instanceOf>
```

```
        <member>
              <roleSpec>
                    <topicRef xlink:href="#MT13"/>
              </roleSpec>
              <topicRef xlink:href="#MT18"/>
        </member>
        <member>
              <roleSpec>
                    <topicRef xlink:href="#MT14"/>
              </roleSpec>
              <topicRef xlink:href="#T1"/>
        </member>
 </association>

<association id="MM2">
        <instanceOf>
              <topicRef xlink:href="#MT15"/>
        </instanceOf>
        <member>
              <roleSpec>
                    <topicRef xlink:href="#MT13"/>
              </roleSpec>
              <topicRef xlink:href="#MT18"/>
        </member>
        <member>
              <roleSpec>
                    <topicRef xlink:href="#MT14"/>
              </roleSpec>
              <topicRef xlink:href="#T2"/>
        </member>
 </association>

<association id="MM3">
        <instanceOf>
              <topicRef xlink:href="#MT15"/>
        </instanceOf>
        <member>
              <roleSpec>
                    <topicRef xlink:href="#MT13"/>
              </roleSpec>
              <topicRef xlink:href="#MT20"/>
        </member>
        <member>
              <roleSpec>
                    <topicRef xlink:href="#MT14"/>
              </roleSpec>
              <topicRef xlink:href="#T4"/>
        </member>
 </association>

<association id="MM4">
        <instanceOf>
              <topicRef xlink:href="#MT16"/>
        </instanceOf>
        <member>
              <roleSpec>
                    <topicRef xlink:href="#MT13"/>
              </roleSpec>
              <topicRef xlink:href="#MT18"/>
        </member>
        <member>
              <roleSpec>
                    <topicRef xlink:href="#MT14"/>
              </roleSpec>
              <subjectIndicatorRef
                xlink:href="http://www.sql-tc.com/king.xtm#M1"/>
```

```
        </member>
</association>

<association id="MM5">
        <instanceOf>
                <topicRef xlink:href="#MT16"/>
        </instanceOf>
        <member>
                <roleSpec>
                        <topicRef xlink:href="#MT13"/>
                </roleSpec>
                <topicRef xlink:href="#MT21"/>
        </member>
        <member>
                <roleSpec>
                        <topicRef xlink:href="#MT14"/>
                </roleSpec>
                <subjectIndicatorRef
                   xlink:href=" http://www.sql-tc.com/king.xtm #M7"/>
        </member>
</association>

<association id="MM6">
        <instanceOf>
                <topicRef xlink:href="#MT17"/>
        </instanceOf>
        <member>
                <roleSpec>
                        <topicRef xlink:href="#MT13"/>
                </roleSpec>
                <topicRef xlink:href="#MT18"/>
        </member>
        <member>
                <roleSpec>
                        <topicRef xlink:href="#MT14"/>
                </roleSpec>
                <subjectIndicatorRef
                   xlink:href=" http://www.sql-tc.com/king.xtm#S1"/>
        </member>
</association>

<association id="MM7">
        <instanceOf>
                <topicRef xlink:href="#MT17"/>
        </instanceOf>
        <member>
                <roleSpec>
                        <topicRef xlink:href="#MT13"/>
                </roleSpec>
                <topicRef xlink:href="#MT20"/>
        </member>
        <member>
                <roleSpec>
                        <topicRef xlink:href="#MT14"/>
                </roleSpec>
                <subjectIndicatorRef
                   xlink:href=" http://www.sql-tc.com/king.xtm#S2"/>
        </member>
</association>

<association id="MM8">
        <instanceOf>
                <topicRef xlink:href="#MT17"/>
        </instanceOf>
        <member>
                <roleSpec>
```

```xml
                        <topicRef xlink:href="#MT13"/>
                  </roleSpec>
                  <topicRef xlink:href="#MT22"/>
            </member>
            <member>
                  <roleSpec>
                        <topicRef xlink:href="#MT14"/>
                  </roleSpec>
                  <subjectIndicatorRef
                     xlink:href=" http://www.sql-tc.com/king.xtm#S3"/>
            </member>
      </association>

      <association id="MM9">
            <instanceOf>
                  <topicRef xlink:href="#MT17"/>
            </instanceOf>
            <member>
                  <roleSpec>
                        <topicRef xlink:href="#MT13"/>
                  </roleSpec>
                  <topicRef xlink:href="#MT19"/>
            </member>
            <member>
                  <roleSpec>
                        <topicRef xlink:href="#MT14"/>
                  </roleSpec>
                  <subjectIndicatorRef
                    xlink:href=" http://www.sql-tc.com/king.xtm #S4"/>
            </member>
      </association>

      <association id="MM10">
            <instanceOf>
                  <topicRef xlink:href="#MT23"/>
            </instanceOf>
            <member>
                  <roleSpec>
                        <topicRef xlink:href="#MT26"/>
                  </roleSpec>
                  <topicRef xlink:href="#MT24"/>
            </member>
            <member>
                  <roleSpec>
                        <topicRef xlink:href="#M25"/>
                  </roleSpec>
                  <subjectIndicatorRef
                     xlink:href=" http://www.sql-tc.com/king.xtm#S4"/>
            </member>
      </association>

</topicMap>
```