An Interactive Ontology for the Intertidal Fish of the Cape Technical Report

Project team

Hilton Jacobson, Michael Kassa and Yan-Xun Hu

Supervisors

Anet Potgieter and Charlie Griffith

Abstract

This paper describes the design and implementation of an interactive identification guide for the intertidal fish of the Cape. Its aim is to create an interactive navigational system to improve the process of identifying fish species based on their feature. This system uses Topic Map as data structure to represent ontology information. It is separated into three parts: knowledge representation, Topic Map query langue and Visualization.

Key words: *topic maps, topic map query language, topic map visualizers*

1. Introduction

The aim of the project is to design better and more user-friendly identification keys system to marine invertebrate animals. More specifically it aims at creating an interactive identification guide to the intertidal fish of the Cape.

The project uses topic maps [1] to carefully capture the taxonomic characteristics, and the relationships associated with the species. An existing topic map visualizer [2] is then adopted and modified to display the topic map of the fish species to the user. Topic Map Query Language(TMQL) is also developed for users to be able to have a quick multiple entry point search for the identification of the fish species.

The project is modulated into three components, the first one deals with the data storage and representation the second one deals with querying the data and the last one is used to visualize queried results and the data by itself.

In this technical report section 2 gives a quick introduction to related research topic and topic maps. Section 3 discusses how topic maps, TMQL and topic map visualizers were applied to the fish species. Section 4 details the results acquired and makes an analysis of the distinction from the previous results.

2. Related Work

2.1 Binary Key Identification System

The existing system for the identification of the interdial fish species uses binary key system to identify a given fish species. The user is given a yes or no question, depending on the answer the user is then directed to other yes or no question until the fish that the person is looking for is found. However the problem with this is the person may not know the exact answer to the yes or no question therefore a mistake in one yes or no question will lead to a totally different result, the other problem with the binary key identification system is that, the binary key system allow only a single point entry making the tree transversal long.

2. 2 Topic Maps

This project depends mainly of the technology of topic maps for storing its data and hence a brief introduction to topic maps is given first before proceeding further.

Topic Maps are ISO standard used to represent and interchange knowledge by modeling the semantic network of the information, with the

emphasis of findability of information. In this paper knowledge is defined as the information and the information relationships that exists in an organization or any firm. Topic Maps are made of three building blocks of components.

- Topics
- Associations
- Occurrences

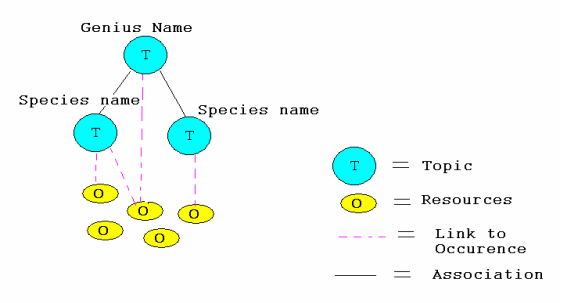
A topic in a topic map is used to represent any entity. Topics are used to represent *subjects* is a computer, both a topic and subject mean the same thing in most cases, and in this paper they do not have any distinction. The ISO/IEC 13250:2000 definition of a subject is:

"any thing whatsoever, regardless of weather it exists or has any other specific characteristics, about which any thing whatsoever may be asserted by any means whatsoever"

Associations are used to link topic with each other. They describe the knowledge relationships in topic maps.

Occurrences in topic maps are used to give the information that are relevant to a topic. They can be data string or URL to information that is related to the topic. Any resource that is relevant to a topic can be considered as an occurrence.

The following figure illustrates the overview of topic map's main components. In the figure below topic *Genius Name* is associated with topics *species name* and the species topics have occurrence, yellow ellipse



2.3 Visualization Babe System

Babe System [3] has developed software that is used for visual representation of Topic Maps. Their visual is a java applet, that when clicked on a diagram or a text, it brings to the image to the center and branches out the sub-branches of the images. Though like the topic maps, this paper do not explicitly talk about the fish species, its general structure is important for our visual representation and the querying part of the project.

2.2 Topic Map Query Languages

TMQL are new ISO standard [4], that are used for querying topic maps. The ISO has specifies a list of use-cases that a TMQL should answer. Many TMQLs were developed some of them answer all the specification and some of them partly. Even though there are already implemented TMQLs, they are either expensive, therefore in this project the TMQL are developed from scratch though a reference is made to Tolog, a programming language that implanted TMQL fully.

3. Approaching the Problem

3. 1 Design

The design of the system was mainly focusing on the utilization of open source software. The system used MySQL for the topic map data backend. The topic map visualizer that is optimized for our project is also an open source.

3.2 Database Storage

A number of considerations were made in selecting a database that suits the needs of the application. The decision was made to use an RDBMS (Relational Database Management System); namely MySQL predominately for its high performance and small resource requirements compared to other databases systems. The schema of the database is designed to map to the visual view of the application, which allows the user to navigate through the characteristics of the fish.

3.2.1 XML Topic Map (XTM) representation

The next phase in the implementation was to generate a Topic Map based on the well-defined schema. The Topic Map is to be treated as a data structure that maintains all logical associations between the characteristics of the fish. The goal for the Topic Map creation is to accurately represent the ontology of fish.

The table below represents the transformation of the database into a Topic Map. Every aspect of

the database has some sort of relevance to the Topic Map.

Relational Database	Topic Map
Table	topic class
Row	topic instance of corresponding class
Column	name or occurrence
Key	topic id
foreign key	Association

3.2 Visualization

Graph visualization is a way of representing structural information as diagrams of abstract graphs and networks. Automatic graph drawing has many important applications in software engineering, database and web design, networking, and in visual interfaces for many other domains. Topic map structure is very similar to that of the semantic network, and there has been lots of wok done in visualizing semantic networks it is intuitive to use techniques of semantic networks visualization in visualizing topic maps.

A semantic network is a highly connected knowledge structure. A semantic network is similar to that of the ball and stick models used in chemistry. However, a semantic network can be far more complex, since not only adjacent nodes can be connected, in a semantic network any two nodes may be connected — regardless of the distance they are away from each other. The challenge in this scenario is to present a meaningful visualization of the underlying knowledge structure while providing intuitive understanding, eliminating visual clutter and maintaining aesthetic appeal.

There the system is written in java because many open source tools are available for visualizing semantic networks. This system is build from an open source tool called Link Browser, which uses Touch Graph as graph visualization. This tool has been modified to visualize XTM. Extra information related to a node can be displayed in hint box. If there are picture related to a node, it would be displayed in a panel.

3.3 TMQL

The direct application of the implemented TMQL in this project is for allowing queries with multiple entry points. However the TMQL has its own API that allows a user to query the topic map using the syntaxes defined using the grammar, and hence can be used independently for any topic map.

The TMQL syntaxes for this project are subsets of Tolog's TMQL syntax. The reason for choosing Tolog's TMQL is that the syntax are similar to the familiar SQL syntax, secondly Tolog has fully implemented the ISO TMQL use case[5] specification hence further expansion of this TMQL can make reference to the grammar of Tolog's TMQL specification and learn from their weakness and success.

3.3.1 TMQL Implementation

The program for this system is implemented in Java, the TMQL module has four main sections:

- Grammar validator (Query parser) : defines the grammar and validates the syntaxes
- XTMParser: Reads the XML topic map (XTM)
- QueryExcecution: Executes validated query
- Interface: Gives the user interface, returns executed results.

3.3.2 TMQL Syntaxes

Two sets of sublanguages were defined for this TMQL, *path expression* and *select Expression.*

The *path expression* defines the most basic syntaxes for accessing topic, occurrences and associations. The syntax of the path expression is more like XPath in XML, the path expressions start with a value and generate new values with steps until they finally produce a value or set of values that get returned.

Select expression is a superset of path expression and its syntaxes are similar to SQL.
The path expressions in this TMQL use mostly predicates, however the path expressions are used when some one wants to filter returned values of predicate variables.

3.4 System Integration

As explained above the system has three main components, data representation using topic

maps, topic map query language for querying the topic map, and the topic map visualization for visualizing the topic maps and the results returned from the TMQL. The following figure gives integration of the system.

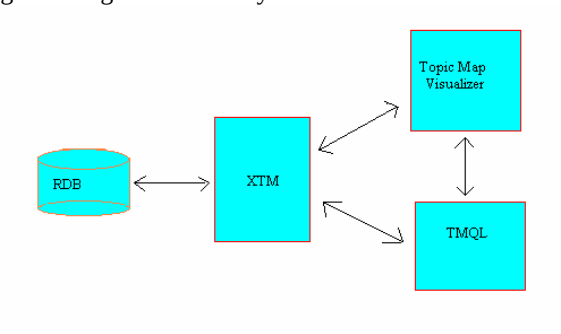


Figure 3. System integration,

4. Results

4.1 Statistical Results

A comparison made on the identification of a give fish species and the time it takes for a user to identify a fish was done on a number of users. The results showed that in the new system, users were able to identify the fish at a relatively short period of time when compared with the old existing binary key identification system. The results also show that more user were unable to identify a give fish using the old system that using the new system.

4.2 Usability testing

According an interview made with a few users of both systems the new system have proved to show a remarkable superiority over the previous binary key identification system. Users were able to use the new system with less difficultly, and enjoyed all the flexibility that it gave them in terms of the graphical user interface and data presentation. Most of all users were also able to discover more about the fish species. This is because while trying to identify the fish species they have at hand they also see all the other information that the fish is linked to, e.g. one can easily click on a location and see all the other fish that live in that location.

4.3 Accuracy test on TMQL

The TMQL was tested against opera.xtm, an XTM file that is used as a test data in many of Tolog's projects. The results of the query were accurate and returned in a relatively short period of time and the system.

5. Conclusion

The ability to encode arbitrarily complex knowledge structures and link them to information assets indicates a major role for topic maps in the realm of knowledge management

With topic maps a user can wander at leisure through a multidimensional topic space of knowledge before deciding which information resources are relevant, instead of wading through volumes or megabytes of data in order to find what he or she is looking for. In context of this project, topic maps are used as a navigational tool for identification. Users are able to navigate through the characteristics of fish deciding on what they feel is relevant.

Topic Maps lend themselves perfectly for use as a data structure. It is able to store all the information from a database along with the associations that connects it, producing a well structured knowledge base. This advantage along with the fact that it can be effectively generated from a well defined database schema provides a powerful tool for navigation

6. Future Work

PDA Implementation

Adapting the identification system for use on a PDA would provide portability for users. This would prove to be useful for zoologists to identify fish in their environment. A PDA implementation might require a number of adjustments due to its limitations in data storage, screen size and processing power.

Web Integration

Topic Maps lend themselves perfectly to the semantic web as described in the background section. It could then be possible to integrate this identification system to the web, enabling it to automatically harvest new data from other sources and merge with other topic maps.

Further Research in extending Topic Maps for databases

Further investigation can be conducted regarding the interface between a Topic Map and a database. This project concentrated on a relational database to store all information, which lends itself perfectly to Topic Maps because of its well-defined schema. This project can potentially be implemented with other database management systems, for example, an Object-Orientated Databases.

Extending TMQL for Updating Topic Maps

Currently the ISO standard for TMQL is limited to accessing topic map, updating has been left for the future, hence future expansion of TMQL should include updating as well.

Reference:

1. Empolis. 2003 Topic Maps Handbook. [Online].Available: http://www.geocities.com/xtopicmaps/mapping_xtm_on_wdatabases.html

2. Pepper, S. 2005. The TOA of Topic Maps Finding the way in the Age of Infoglut. [Online].Available: http://www.ontopia.net/topicmaps/materials/toa.html

3. Poitreger, A, Bishop, J. 2002. The Engineering of Emergence in Complex Adaptive Systems. [Online]. Available: http://www.cs.up.ac.za/cs/jbishop/Homepage/Pubs/Pubs2002/Potgieter-Emergence2002.pdf

4. ISO. 2005. Topic Map Query Language (TMQL). [Online]. Available: http://www.isotopicmaps.org/tmql/

5. Brata, R. & Garshol, LR. Topic Map Query Language, Use Cases. [Online]. Available: http://www.y12.doe.gov/sgml/sc34/document/0449.htm

6. Garshol, L, M. 2005. TMQL An Introduction.[Online].Available: http://www.emnekart.no/2005/forum-04-19/tmql-intro.pdf