Using XML-Topic Map on a PDA

Pascal Auillans   Olivier Baudon   Franck Jarry

15th October 2001

Contents

1 Authors

- Olivier Baudon, Assistant Professor, Labri, UMR 5800, Université Bordeaux 1, Talence, France baudon@labri.u-bordeaux.fr

Olivier Baudon, born in 1962, graduated from the Institute of Applied Mathematics of the University of Angers (France) and obtained his PhD at the University of Joseph Fourier (Grenoble, France) in 1989. Since 1990, he has been an assistant professor at the University of Bordeaux I. Baudon's main topic concerns graphs, interconnection networks and computer tools for graph theory. He is working with the company Mondeca since 1999 on graph clustering algorithms applied to graphs issued from Topic Maps.

- Pascal Auillans, R&D engeneer, Mondeca, Montrouge, France.

  pascal.auillans@mondeca.com

  Pascal Auillans, born in 1975, graduated from the University of Bordeaux I (France) in computer science. He started a PhD at the University of Bordeaux I Computer Science department (LaBRI) in 2000 with Olivier Baudon and Andre Raspaud in the field of graph theory. At the same time he is working for Mondeca, a firm involved in Topic Maps software developement. His PhD project deals with graph theory concepts (such as clustering) applied to semantic graphs.

- Franck Jarry, R&D engeneer, Mondeca, Montrouge, France.

  Franck.Jarry@scaramouch.com

  Franck Jarry, born in 1978, studied Computer Science from the University of Bordeaux I (France). He obtained the degree of MSc in Software Engineering from the University of Bordeaux I in 2001, completing his studies at Mondeca from March to September 2001.

2 Abstract

Palm Navigator is a shareware program that is designed to help import an XML/Topic-Map onto a PDA (Personal Digital Assistant), and to enable navigation, jumping from one topic to another as easily as Web surfing. Palm Navigator is fully compliant with the ISO Topic Maps standard (ISO/IEC 13250) which enables exchanges between Web sites.

Due to the lack of power, both in terms of memory and processing capacities, topic maps cannot be manipulated (parsing, navigating, querying) directly on the PDA. Moreover, the only way to stock data on a Palm Pilot is by using Palm Pilot Database files (PDB), which are archaical indexed files of fixed-length strings. In Palm Navigator, we propose an efficient, yet

complex architecture based on a data structure issued from graph theory. A prototype in Java has been available since March 2001 and a C version is planned for March 2002.

Palm Navigator is separated into two components. The first, located on the user's Personal Computer, is in charge of parsing and converting the XML-Topic Map file to a PDB file format and transferring it to the PDA, using the usual craddle. The second allows navigation and querying on the PDA.

Representation on the PDA is based on a graph structure, enabling efficient search and navigation on the topic map. As explained in our previous presentation at XML-Europe 2001 "Graph Clustering for Very Large Topic Maps", a topic map may be considered from a mathematical point of view, as a hypergraph, then represented by the bipartite graph representative of the hypergraph.

The graph structure is implemented with fixed size arrays which allow very fast navigation and search. If the maximum degree of the vertices is bounded, any basic operation can be performed in a constant time. In terms of topic maps, this means that our structure is efficient if the maximum number of roles for any association does not depend on the sum number of topics plus the number of associations.

In a future version, Palm Navigator will integrate new features such as the ability to attach Palm's standard objects to the topic map, in order to make it self-sufficient.

3 Knowledge Index Management (KIM) and XML - Topic Maps (XTM)

3.1 Mondeca's solution

Mondeca [MON01] provides Topic Maps-based Software Components and Solutions for Knowledge Index Management enabling companies to organize, navigate, visualize and access a large and rich collection of information

Mondeca's mission is to transform the way we all access and navigate content on the Web. Based on the principles that govern the way humans organize and interact with information, Mondeca allows Internet users to visualize information and its whole environment on an intelligent intuitive scale.

Mondeca's technology, referred to as Dynamic Content Navigation, is made of up two fundamental concepts: structuring content as a Topic Network, and allowing intuitive access to the content with Spatial Maps. The

Topic Network organization allows users to access content through a topic-to-topic navigation, providing a clear and efficient content browsing tool (figure 1). Mondeca has also developped Spatial Maps (figure 2) to enable users to access content as a whole by "flying over" these Topic Networks, similar to a zoom-in and zoom-out feature on a camera.

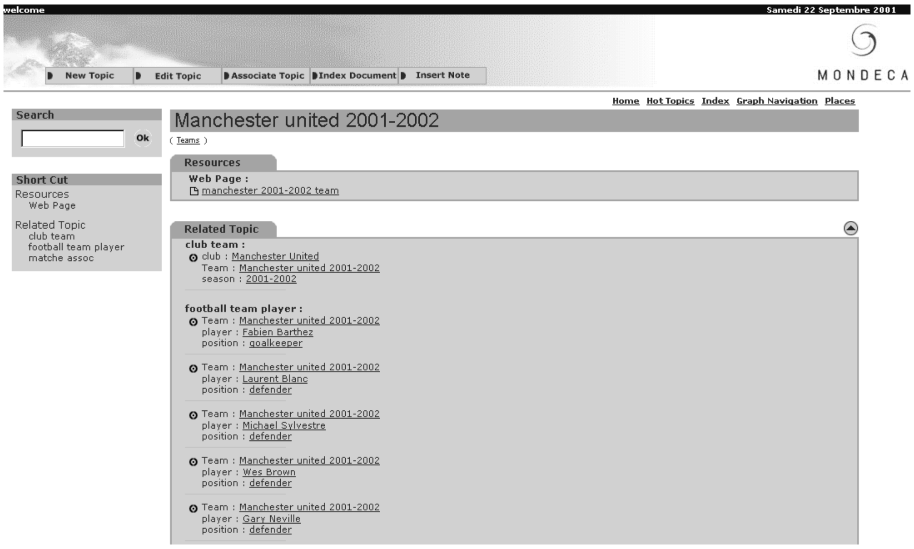


Figure 1: Topic Navigator - Screenshot of the centered topic screen

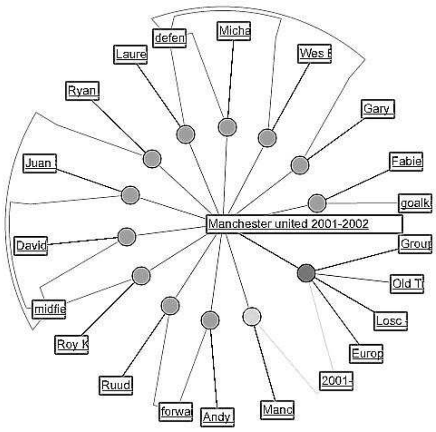


Figure 2: Topic Navigator - Polar drawing in XML SVG format

3.2 Topic Maps

Topic Maps [XTM01] are an ISO Standard (ISO/IEC 13250, January 2000) that enable vast information resources to be classified and navigated consistently. We have long understood the idea of creating style sheets to control the formatting and layout of information. Topic Maps introduce the concept of creating style sheets to control knowledge-based information access and navigation.

From a technical sense, Topic Maps describe what an information set is about by formally declaring topics and linking the relevant information sets to the appropriate topics (figure 3). Topic Maps can be applied from "above" the information set, rather than from "inside" them. They are superimposed views. A Topic Map expresses an opinion about what the topics are and their relevant relationships. There is no limit to the number of Topic Maps that can be created above the same information set.

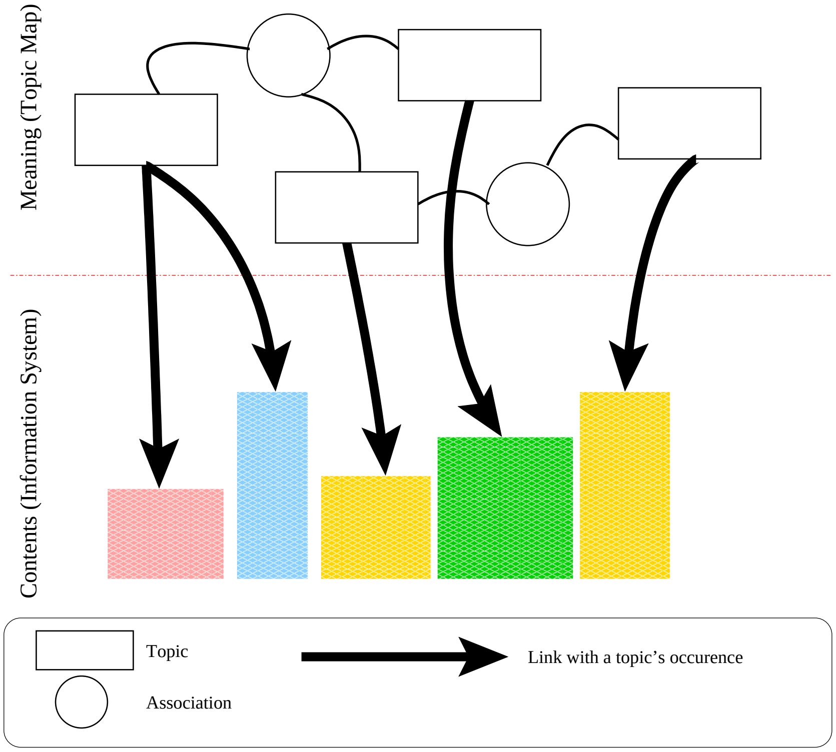


Figure 3: Separation Contents - Meaning

4 Topic Maps and PDA

4.1 Palm Navigator

Mondeca would like to provide XTM users with a shareware software for using Topic Maps on their PDA. This shareware, called Palm Navigator, is developped for Palm OS with the help of MSc students from the University of Bordeaux 1, coordinated by Pascal Auillans and Olivier Baudon. A first version has been available since March 2001. A second is planned for March 2002. This paper will present the current version, as well as improvements planned for the second version.

Programming for a PDA presents three major difficulties :

- small capacity of RAM
- slow processor
- archaical file system

To solve these limitations, we used an efficient data structure issued from graph manipulations. Moreover, the parsing and analysis of the topic map is deported to a PC (figures 4 and 5).

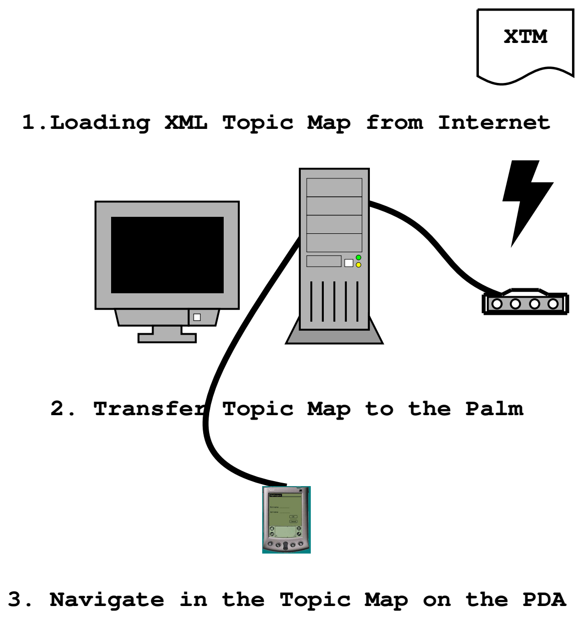


Figure 4: General principle of Palm Navigator

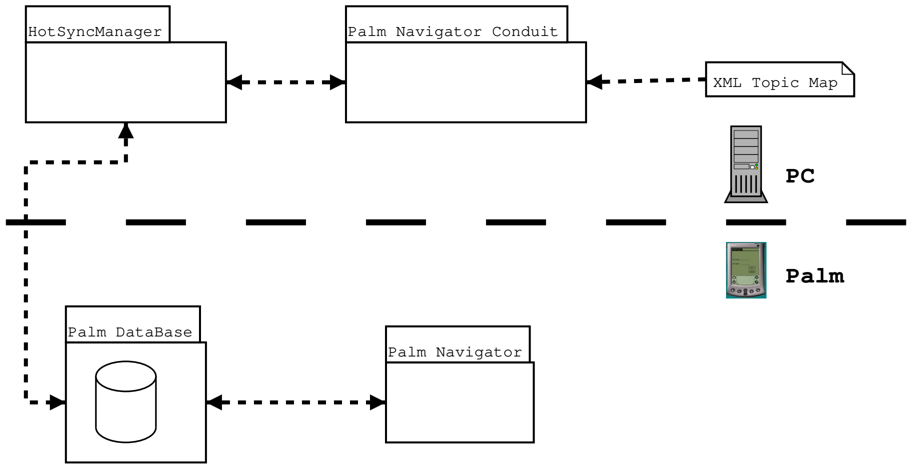


Figure 5: General architecture of Palm Navigator

4.2 From Topic Maps to Bipartite Graphs

From a mathematical point of view, a topic map may be considered as a hypergraph and thus, as a bipartite graph [AB01].

4.2.1 Graph

A **graph** G is a couple (V, E), where V is the set of **vertices** and E the set of **edges**. An edge e is a pair of vertices $\{x, y\}$. We usually say that x and y are **incident with** e, or that they are the **endpoints** of e. x and y are said to be **adjacent** or **neighbors**.

In this paper we will consider that all the graphs are simple, meaning there exist neither multiple edges nor self-loops, ie each pair $\{x, y\}$ in E is unique and with $x \neq y$.

The **order** of a graph, that we will note n, is its number of vertices, and its **size**, that we usually note m, is its number of edges. The **degree** of a vertex v, denoted by $deg(v)$, is the number of edges that are incident with it.

4.2.2 Example

The simple graph in figure 6 is of order 6, and size 8. The degree of the vertex labeled 0 is 2. Its neighbors are the vertices 1 and 5 and its incident edges are $a = \{0,1\}$ and $f = \{0,5\}$.

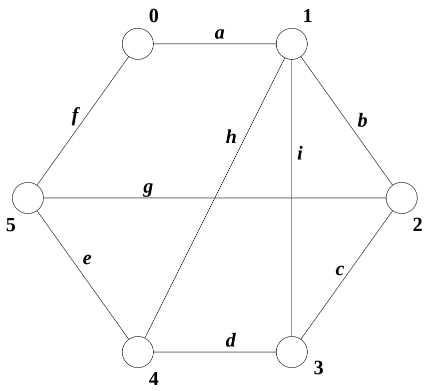
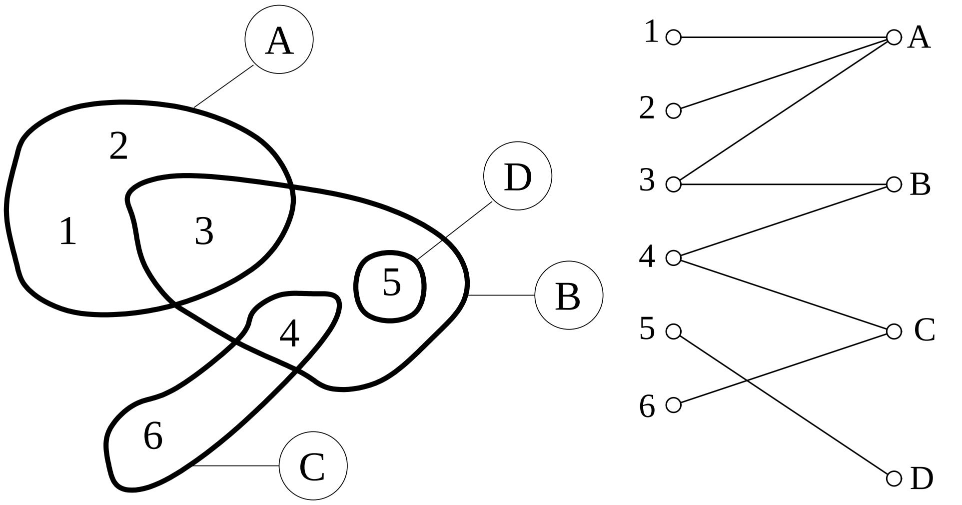


Figure 6: A simple graph G

4.2.3 Hypergraphs

Hypergraph is a generalization of graph concept, where an edge is incident with an unspecified number of vertices. In this case, we will use the term **hyperedges**. The representing graph of a hypergraph, $H = (V_H, G_H)$ is a graph G where the set of vertices is the union of the set of vertices and the edges of the hypergraph, and there exists an edge between two vertices, if and only if, one of the vertices is a hyperedge incident with the other vertex : $G = (V_H \bigcup E_H, \{\{v,e\}, e \in E_H, v \in e\})$.

An important property of a representative graph of a hypergraph is that the vertices can be decomposed into two sets such that there exist no edges between any couple of vertices belonging to the same set. These types of graphs are therefore named **bipartite** graphs.



(a) Hypergraph H

(b) Representing graph of H

Figure 7: Hypergraph H

4.2.4 Graph Coding

To represent the Topic Map on the PDA, we use its graph representation. A graph is mainly represented using two arrays of integers. Each vertex is represented by an integer from 0 to $n-1$. Because the graph is simple, we only have to code the list of neighbors for each vertex. This is done in a first array $a1$, of size $2m$ where the adjacency lists are given in a contiguous order. The second array $a2$ is of size $n+1$ and contains at each index v the starting index of the neighborhood of vertex v. The last index n contains $2m$. In that way, the degree of each vertex v may be easily obtained because it is equal to $a2[v+1] - a2[v]$.

Note that if each neighborhood is sorted, knowing if two vertices are adjacent or not may be obtained in a logarithmic time, using dichotomy.

Example The graph in figure 6 is coded by the two follwing arrays :

i	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
a1[i]	1	5	0	2	3	4	1	3	5	1	2	4	1	3	5	0
a2[i]	0	2	6	9	12	15	18									
i	16	17														
a1[i]	2	4														

5 Palm Navigator

5.1 Using Palm Navigator

The figure 8 presents the sequence of screens in Palm Navigator. An example of each screen is given in figures 9, 10 and 11.

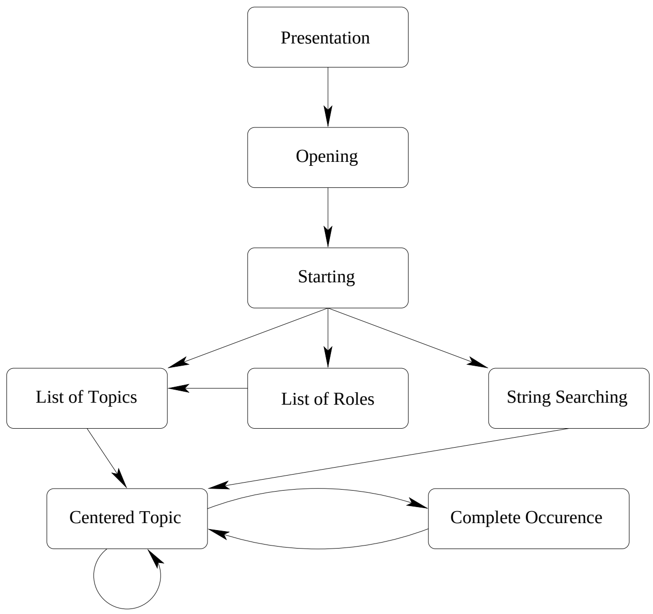
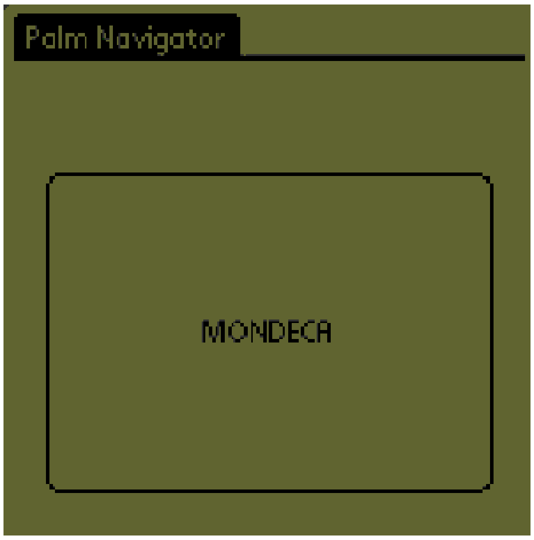


Figure 8: Sequence of screens on the PDA



(a) Presentation

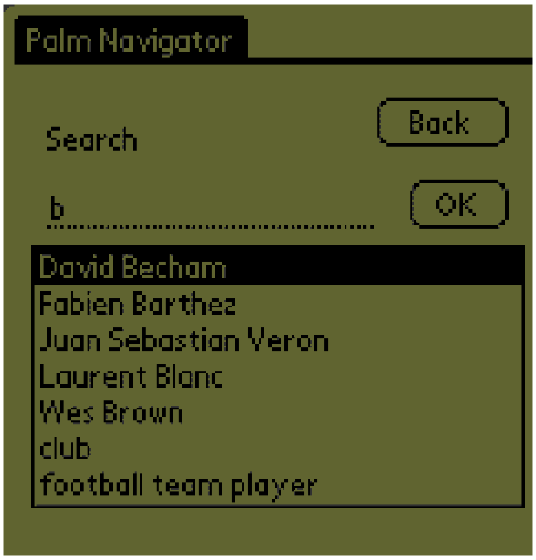
(b) Opening

(c) Starting

(d) List of topics and associations

Figure 9: Palm Screenshots - Part 1

(a) List of roles



(b) String searching

(c) Centered topic (Manchester United 2001-2002)

(d) Centered topic (Fabien Barthez)

Figure 10: Palm Screenshots - Part 2

(a) Occurence (Statistics of Fabien Barthez)

Figure 11: Palm Screenshots - Part 3

5.2 Technical specifications

Both PC's XML parser and Palm navigator were written using Java.

The XML parser uses the Sun's Java APIs for XML processing : JAXP 1.1 [JAX01]. It is implemented as a HotSync Manager's plug-in, using CDKJ 4.0 [CDK01]. Our DTD for a Topic Map on the Palm is a subset of the ISO DTD, which is itself a subset of the DTD used by Mondeca Navigator (A).

Palm Navigator were written using WabaSDK [WAB01] and runs using the Waba virtual machine WabaVM. The IHM was written using Waba Extras v1.30.

The software was tested on a Palm emulator and on a Palm V with 3Mo of RAM, and a processor at 20Mhz. It runs well on the emulator. On the Palm, the number of topics and associations available is too small to have a real use. Moreover, drawing a screen takes too much time.

5.3 Furtherwork

A new version of Palm Navigator is planned for February 2002. The following improvements have been defined for this version :

1. specify screens more similar than those proposed in Mondeca Navigator;
2. rewrite Palm Navigator in C or C++ to improve the number of topics and associations and the speed;
3. improve the XTM Parser.

A XTM DTD for Palm Navigator

```
<!-- XML Topic Map DTD  for Palm Navigator....................... -->
<!-- XML Topic Map (XTM) DTD for Palm Navigator, Version 1.0

     This DTD is a subset of the XTM DTD version 1.0
     XML Topic Map (XTM)
       Copyright 2000 TopicMaps.Org, All Rights Reserved.

        Authors:    Group Mondeca (DESS GL Université Bordeaux I)
        Version:    v1.0
-->
<!-- Use this URI to identify the default XTM namespace:

         "http://www.topicMaps.org/xtm/1.0/"

     Used to identify the XLink namespace:

         "http://www.w3.org/1999/xlink"
-->

<!-- topicMap: Topic Map document element ....................... -->

<!ELEMENT topicMap
   ( topic | association )*
>
<!ATTLIST topicMap
   id              ID        #IMPLIED
   xmlns           CDATA     #FIXED 'http://www.topicmaps.org/xtm/1.0/'
   xmlns:xlink     CDATA     #FIXED 'http://www.w3.org/1999/xlink'
   xml:base        CDATA     #IMPLIED
>

<!-- topic: Topic element ...................................... -->

<!ELEMENT topic
   ( instanceOf*, ( baseName | occurrence )* )
>
<!ATTLIST topic
   id              ID        #REQUIRED
>
```

```
<!-- instanceOf: Points To a Topic Representing a Class .......... -->

<!ELEMENT instanceOf  topicRef >
<!ATTLIST instanceOf EMPTY >

<!-- topicRef: Reference to a Topic Element ...................... -->

<!ELEMENT topicRef  EMPTY >
<!ATTLIST topicRef
   xlink:href      CDATA     #REQUIRED
>

<!-- baseName: Topic or Occurrence Base Name ..................... -->

<!ELEMENT baseName  baseNameString >
<!ATTLIST baseName EMPTY >

<!-- baseNameString: Base Name String Container .................. -->

<!ELEMENT baseNameString  ( #PCDATA ) >
<!ATTLIST baseNameString EMPTY >

<!-- occurrence: Resources Regarded as an Occurrence ............. -->

<!ELEMENT occurrence
   ( instanceOf?, baseName?, resourceData )
>
<!ATTLIST occurrence EMPTY >

<!-- resourceData: Container for Resource String ................. -->

<!ELEMENT resourceData  ( #PCDATA ) >
<!ATTLIST resourceData EMPTY >

<!-- association: Topic Association  ............................. -->

<!ELEMENT association
   ( instanceOf?, member+ )
>
<!ATTLIST association
   id             ID        #REQUIRED
>
```

```
<!-- member: Member in Topic Association ........................ -->

<!ELEMENT member
   ( roleSpec?, topicRef+ )
>
<!ATTLIST member EMPTY >

<!-- roleSpec: Points to a Topic Serving as an Association Role .. -->

<!ELEMENT roleSpec topicRef >
<!ATTLIST roleSpec EMPTY >

<!-- end of DTD XML Topic Map (XTM) for Palm Navigator -->
```

References

[AB01] Pascal Auillans and Olivier Baudon. Graph clustering for very large topic maps. In Pamela Gennusa, editor, *XML 2001*, Berlin, May 2001.

[CDK01] Palm os: Conduit developments kit - windows, 2001. `http://www.palmos.com/dev/tech/tools/cdk/win/`.

[JAX01] Java apis for xml parsing (jaxp), 2001. `http://java.sun.com/xml/jaxp.html`.

[MON01] Mondeca - a better content organization for an accurate information retrieval, 2001. `http://www.mondeca.com`.

[WAB01] @wabasoft, 2001. `http://www.wabasoft.com`.

[XTM01] Xml topic maps (xtm) 1.0, 2001. `http://www.topicmaps.org/xtm/index.html`.