DESIGN AND IMPLEMENTATION OF XTM FOR NAVIGATING NEWS DOMAIN

By

AMINU ENESI FEMI

DEPARTMENT OF MATHEMATICS

FACULTY OF SCIENCE

AHMADU BELLO UNIVERSITY, ZARIA, NIGERIA

SEPTEMBER, 2014

DESIGN AND IMPLEMENTATION OF XTM FOR NAVIGATING NEWS DOMAIN

By

Enesi Femi AMINU

B.Sc Computer Science (University of Jos, 2008)

M.Sc./Sci./10104/2011-2012

A THESIS SUBMITTED TO THE SCHOOL OF POST GRADUATE STUDIES

AHMADU BELLO UNIVERSITY, ZARIA

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF A

MASTER OF SCIENCE (M.Sc.) DEGREE IN

COMPUTER SCIENCE

DEPARTMENT OF MATHEMATICS,

FACULTY OF SCIENCE

AHMADU BELLO UNIVERSITY, ZARIA,

NIGERIA

SEPTEMBER, 2014

DECLARATION

I declare that the work in this thesis titled "Design and Implementation of XTM for navigating News Domain" has been written by me in the Department of Mathematics under the supervision of Professor O.S. Adewale and Dr. S.E. Abdullahi.

The information contained from the literature has been duly acknowledged in the text and a list of references provided. No part of this work has been presented for another degree or diploma at any institution.

Aminu, Enesi Femi _____ _____

Name of Student Signature Date

CERTIFICATION

This thesis titled "Design and Implementation of XTM for Navigating News Domain" by Aminu, Enesi Femi (M.Sc./Sci./10140/2011-2012) meets the regulations governing the award of the degree of M.Sc. Computer Science of Ahmadu Bello University (ABU), Zaria, and it is approved for its contribution to knowledge and literary presentation.

…….	……	……
Chairman, Supervisory Committee	Signature	Date
…….	……	……
Member, Supervisory Committee	Signature	Date
…….	……	……
Member, Supervisory Committee (External Examiner)	Signature	Date
…….	……	……
H. O. D. Mathematics	Signature	Date
…….	……	……
Dean, Post Graduate School	Signature	Date

DEDICATION

This thesis is dedicated to God The Almighty, The Architect of the universe for His uncommon favour and mercy upon my life, most especially my academic journey. To HIM be the Glory and Adoration forever.

ACKNOWLEDGEMENT

I want to first acknowledge the mercy of God in my life specially, throughout the course of this programme. Merciful Lord; the joy in my heart is flowing like a river, adoration forever.

I am full of appreciation to my major supervisor Prof. O.S. Adewale who guided me all through the thesis work. His Advice and other contributions have really given the work a new paradigm. I appreciate your patience and love. To my minor supervisor Dr. S.E. Abdullahi, I'm highly indebted to your immense contributions and understanding.

I equally appreciate all my tutors during my course work especially Prof. S.B. Junaidu and Dr A. A. Obiniyi. The list will not complete without gratitude to the Head of my Department, Dr. B. Sani, he is really a father and a friend.

I am highly indebted to my lovely wife. Just like the popular saying; behind every successful Man, there is a Woman. My lovely Angel, thank you for your prayers and unalloyed supports. My little Angel (Daughter) Diniel Elizabeth thanks for your audible voice.

Finally, I cannot end this appreciation without proper salute to my parents, siblings, course mates, friends and family. Mum, you are an epitome of African Woman; which stands for love, cares and all the good tidings of life. I salute you.

ABSTRACT

In this present age, the challenge is no longer the availability of information but the appropriate information. This challenge is attributed to the fact that information is highly unstructured. Also in many areas of knowledge, concepts and relationships become so complex because they are not structured and as such it is difficult to organize and navigate the information clearly and logically. Thus, alternative information-organization techniques are required to be able to effectively and efficiently navigate and retrieve information from systems, web included. By using an innovative technology - XTM for organizing and representing knowledge out of a specific domain, such as news domain, an XTM-based ontology for news domain is developed in this thesis with loads of advantages; for reuse, sharing, and interoperability of knowledge structure. The information becomes structured and the Scope attribute of topics in XTM is used to classify and retrieve information, which will mostly reduce the information querying time. To this end, this thesis, implements a web-based application named NEWS.net which is capable of resolving dead links issue.

TABLE OF CONTENT

DECLARATION iv

CERTIFICATION v

DEDICATION vi

ACKNOWLEDGEMENT vii

ABSTRACT viii

LIST OF FIGURES xii

LIST OF TABLES xiii

ABBREVIATIONS xiv

CHAPTER ONE: INTRODUCTION 1

1.1 Background of the Study 1

1.2 Research Motivations 6

1.3 Research Aim and Objectives 7

1.4 Research Methodology 7

1.5 Contribution of the Study to Knowledge 8

1.6 Structure of the Thesis 8

CHAPTER TWO: LITERATURE REVIEW 9

2.1 Introduction 9

2.2 XML Topic Maps (XTM) 9

2.3 What are XML Topic Maps 10

2.4 The TAO Characteristics of Topic Maps Model 11

2.4.1 Topics 12

2.4.2 Association 13

2.4.3 Occurrences 13

2.5 Topic Identification: Name and BaseName 14

2.6 XTM Attribute: Scope and Namespace 15

2.7 XTM's Published Subject Indicators (PSI) 16

2.8 Producing an XML Topic Maps 17

2.9 XML Topic Maps (XTM) Creation Tools 20

2.10 Ontology Process 23

2.11 Related Study 25

CHAPTER THREE: DESIGN OF XTM FOR NAVIGATING NEWS DOMAIN 29

3.1 Introduction 29

3.2 Basic News Ontology Design 29

3.3 Design Process of XTM 36

3.3.1 Defining the target 37

3.3.2 Build and manage ontology (ontology usage) 37

3.3.3 XTM software application 39

3.3.4 Setting up the XTM concepts 41

3.3.5 Allocation of information resources 42

3.3.6 Establishing Association between Topics 43

3.3.7 Testing the XTM 44

3.4 Visualizing the XTM Document 46

3.5 System Architecture 47

3.5.1 Data tier 48

3.5.2 Server/Maintenance tier 49

3.5.3 Users tier 49

CHAPTER FOUR: IMPLEMENTATION, RESULTS AND DISCUSSION 51

4.1 Introduction 51

4.2 System Requirements 51

4.2.1 Functional requirements 52

4.3 System Implementation 53

4.3.1 Index page 54

4.3.2 Navigating News categories 54

4.3.3 Resolving dead links 56

4.4 Results and Discussions 56

4.5 System Testing 60

4.5.1 Usability testing 60

4.5.2 User acceptance testing 60

4.5.3 Security testing 61

4.5.4 Functional testing 61

4.6 Results of the RSS Model 61

CHAPTER FIVE: CONCLUSION AND RECOMMENDATION 69

5.1 Conclusion **69**
5.2 Recommendations **70**
5.3 Future Work **70**
REFERENCES **72**
APPENDIX A: ONTOLOGY FILE **75**
APPENDIX B: SAMPLE PROGRAM FILE **82**

LIST OF FIGURES

Figure1.1: The Genealogy of XTM (Park and Hunting, 2002). 4
Figure 2.1: Example of Topics and Types 12
Figure 2.2: Sports News Taxonomy 17
Figure 3.1: News Ontology 30
Figure 3.2: Ontopia Application Home Page 39
Figure3.3: Concepts of XTM 42
Figure3.4: Ontology, pointing to different information resources (URL). 43
Figure 3.5: Interface for browsing the ontology. 45
Figure 3.6: News Ontology Vizigator 46
Figure 3.7: Vizigator Properties of (Sports) Topic 47
Figure 3.8: System Architecture 48
Figure 4.1: Index Page of the NEWS.net Application 54
Figure 4.2 (a): Sports News Page 55
Figure 4.2 (b): Sports News sub-category 55
Figure 4.3: Page for online Sports News 56
Figure 4.4: Result of (Dead Links) URLs in the infotech News of NEWS.net System 58
Figure 4.5: Screenshot of Bio informatics XTM-based System's Dead Links Output 59
Figure 4.6: Screenshot of Oil Spill XTM-based System's Dead Links Output 59
Figure 4.7: Results Displayed for two News Items. 64
Figure 4.8: Algorithm that Discuss the Result of the Implementation 65
Figure 4.9a: JSP codes to check Computer Network Connectivity. 66
Figure 4.9b: JSP codes to check Dead links 67
Figure 4.10: Command to startup Server 68

LIST OF TABLES

Table 2.1: XTM Tools 21
Table3.1: Ranked Matrix 34
Table 3.2: Guidelines for the XTM Document 38
Table 3.3: Comparison of some XTM Application Tools 41
Table 4.1a: Comparative Results of Existing Systems and the proposed NEWS.net system 57
Table 4.1b: Comparative Results of Existing Systems and the proposed NEWS.net system in terms of Dead Links 57
Table 4.2: Ranked News Matrix 62

ABBREVIATIONS

ONF:	Ontopia Navigator Framework
XML:	Extensible Markup Language
XTM:	XML Topic Maps
ISO:	International Standard Organization
IEC:	International Electrotechnical Commission
SGML:	Standard Generalized Markup Langauge
HyTM:	HyTime Topic Maps
URIs:	Uniform Resource Identifiers
RDBMS:	Relational Database Management System
RSS:	Ranked Semantic Similarity
GPS:	Global Positioning System
UDDI:	Universal Description, Discovery, and Integration
JSP:	JavaServer Pages
Xlink:	XML Linking Language
PSI:	Published Subject Indicator
URL:	Uniform Resource Locator
HTML:	HyperText Markup Language
SDK:	Software Development Kit
HTTP:	Hypertext Transfer Protocol
RCP:	Rich Client Platform
TMCL:	Topic Maps Constraint Language
API:	Application Programming Interface

OWL:	Web Ontology Language
SQL:	Structured Query Language
RDF:	Resource Description Framework
LTM:	Linear Topic Maps
J2EE:	Java2 Platform Enterprise Edition
XSLT:	Extensible Stylesheet Language Transformations
W3C:	World Wide Web Consortium
RDF/S:	Resource Description Framework/Scheme
TM4L	Topic Maps for eLearning
TM4J	Topic Maps for Java

CHAPTER ONE: INTRODUCTION

1.1 Background of the Study

The availability of information and its related traffic on the internet is always on the rise. For example; book vendors, news agencies and organizations, or business enterprise in general owned by individuals and organizations have been rapidly shifted to the internet, courtesy of web applications. It is often said that in human history the web is possibly the richest information depository; Dichev and Dicheva (2001). Seeking information is a familiar and important human activity. Most often, we are involved in the use of websites and networked information management systems to surf for information. Whatever information users decide to surf for, the tendency that it can be found somewhere in the digital globe is high. But in most cases, surfing for information is not always effective or efficient. When a user surfs for information, the system frequently returns a huge chunk of unrelated results. Hence, in this current age, the challenge is not the availability of the information but getting the relevant information Yi (2008). This challenge is attributed to the fact that information is highly unstructured. Thus, alternative information-organization approaches are required to more effectively and efficiently navigate and retrieve information from systems, web included.

Also as Website grows and turns into a Web portal with a deeply interconnected Website architecture, in which users are provided a gateway to a rich content of links, images, and other types of information, developers are faced with the growing challenges of enforcing link

integrity and navigational order. Tasks that once were simple can turn into laborious and convoluted processes as the information resource base of the site expands.

It is noteworthy to mention that before the emergence of the new technology called Semantic Web; keyword-based search engines such as Google or Yahoo are the major media of using the current Web. It is doubtless that the web would have not recorded the tremendous achievements it had made today, if not for search engines. However, there are shortcomings militating against their use. Surprisingly, in spite of progressive enhancements and maintenances recorded in the technology of search engine, the shortcomings basically remain unchanged. It is very likely that the technology progress could not match the progressive increase in the Web content.

An alternative measure considered is to represent Web content in a format that is machine-understandable and to adopt knowledgeable techniques to utilize full advantage of these representations. We refer to this scheme of transforming the Web as Semantic Web initiative and inventiveness. It is very important to know that the Semantic Web will not be a new universal super information medium that is parallel to the current World Wide Web; instead it will progressively grow out of the existing, Web Antoniou and Harmelen (2008).

It is pertinent to know that the terms ontology and semantic web are closely related and intervolving as ontology is the stronghold and prime element for building up of semantic web content. Ontology is a known term that computer science hijacked from other known disciplines like Philosophy and gave it a precise technical meaning that is quite different from its original use.

For the purpose of this thesis, we will define ontology according to (Gruber, 1993) as an explicit and formal specification of a conceptualization. And according to topic maps terminology (Ontopia, 2010) ontology is a precise description of the kinds of things which are found in the domain covered by the topic map: in other words, the set of topics that are used to define classes of topics, associations, roles, and occurrences. In general, ontology depicts formally a domain of discourse. That is, it defines the terms used to describe and represent an area of knowledge. For example: News media, Medicine, Photography, Shakespeare work, and so on. However, in this thesis the area of knowledge is on News Domain.

The development of ontology will continue to offer a platform for the building of the semantic part of Semantic Web. It is ascertained that we have several markup languages that have been created and using XML a well structured markup language to denote ontologies is not unfamiliar. Today, XML Topic Maps (XTM) is the main topic maps syntax and is supported by nearly all topic map tools for developing ontologies.

Topic Maps is an ISO standard published as ISO/IEC 13250 in 2000 for the representation and interchange of knowledge. The standard defines the basic model and an SGML-based syntax for it, which uses HyTime for linking, and is therefore known as HyTM. When the standard was published it was clear that something more web-optimized was needed, and so an ad-hoc organization known as TopicMaps.Org was formed to create a topic map syntax based on XML and URIs. TopicMaps.Org published its XML Topic Maps (XTM) 1.0 specification in early 2001, and in October of the same year that syntax was accepted into the second edition of ISO 13250 as an annex.

Structurally, the diagram in Figure1.1 presents the developmental history of XTM. XML was formed because a lot of users felt that SGML was too complicated. There was a need to simplify and limit its features to those that are essential for use in a Web context. It is in the same motivation that XTM was designed: to simplify the ISO topic map specification for optimized use on the Web. (However, the development of XML shows that the eliminated complexity is returning in the associated specifications.)

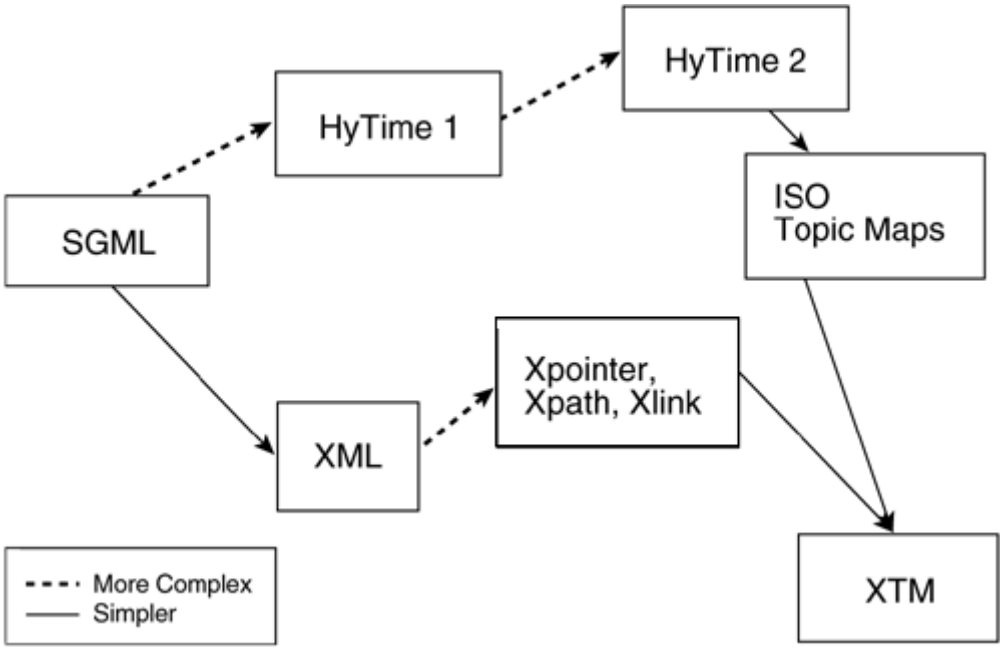


Figure1.1: The Genealogy of XTM (Park and Hunting, 2002).

The purpose of XTM is to convey knowledge about resources through a superimposed layer, or map, of the resources. A topic map captures the subjects of which resources speak, and the relationships between subjects, in a way that is implementation-independent.

In a word, XTM is a kind of data format that can be used to describe the structure of information resources. In addition, it can also quickly locate one or more concepts on a “map”, and represent

the relationship between them. XTM helps by making it possible to relate together, information that comes from different sources through merging and published subjects.

Generally, the heart of every XTM application is known as the topic map engine, which is roughly equivalent to an RDBMS database engine, but designed for XTM. This component knows how to import and export XTM (and other topic map syntaxes), store, update, and query topic maps, and so on. The engine will handle the storage, and any updates would happen through it. For example, applications that implement an XTM-driven portal will sit on top of the engine and use it to access the XTM.

To sum up, topic maps make information findable by giving every concept in the information its own identity and providing multiple redundant navigation paths through the information space. These paths are semantic, and all points on the way are clearly identified with names and types that tell you what they are. This means you always know where you are, which prompted Charles Goldfarb to call topic maps "the GPS of the information universe." Interestingly, Ivan Herman from the World Wide Web Consortium says “that the Semantic Web is in the Uniform Resource Indicators (URIs) and that each resource on the web can be identified through an URI”.

Hence, XTM, known as a meta-data format which describes the structure of information resources, can largely improve information management and delivery. It has great value to study and worth of being researched.

In this thesis, the phrase “topic maps” and XTM are used interchangeably.

1.2 **Research Motivations**

Semantic web concept and its related technology: XTM is a relatively new field of study (Anyi, 2012) carried out a research survey to investigate what has been written about XTM from year 2000 to 2011. As well as finding out the research and publication trends in XTM. Anyi found out that from 2009 to 2011, it was on knowledge and intelligent based system. Anyi equally pointed out five themes where researches were conducted; in which ontology and semantic web were parts of them.

Anyi further enumerated a good number of future research areas such as knowledge navigation and retrieval improvement, intelligent topic map and knowledge representation modeling. In other words, it means modeling or developing XTM-based ontology that can adapt to other forms of ontology language and extended to other framework or application to resolve issues that may arise.

Interestingly, a few number of research works have been carried out but only restricted to designing an XTM-based ontology on some knowledge areas such as music, languages, meals and so on. However, implementing the ontologies to form XTM-based applications (Ellouze *et al*, 2008; Chen *et al*, 2011) is rarely carried out. There is lack of extensibility in their work. And as such there is no application or framework to maintain the XTM-based ontology, since there is continuous evolution of ontology. The likelihood of dead links over an unmaintained ontology (Grigoras, 2005) is always very high. Dead links occur when URLs of a subject or class (topic in XTM) are wrongly typed by author of ontology or the subject or class no longer exist in the internet at the time ontology is designed. That is, users cannot access the resources available at

the occurrence of a topic not as a result of network connection or computer and network protection of a firewall or proxy and so on.

Having carried out a thorough study of the research issues mentioned above, this research work tries to address these research issues, and to the best of our knowledge there is no work in the literature that addresses how XTM can be maintained to avoid dead links.

1.3 Research Aim and Objectives

The aim of this thesis is to be able to explore XTM and its related tools to design and implement an XTM-based web application for the purpose of more efficient and effective information's navigation in the domain of news. The objectives of the research are to:

i. develop an XTM-based ontology for the domain; and resolves dead links which arise from unmaintained XTM.
ii. implement XTM-based application for the domain.

1.4 Research Methodology

A review of related literature such as XML Topic Maps: Creating and Using Topic Maps for the Web, Using XTM for Navigating UDDI, Semantic Web, Ontological Engineering, and some other related work is carried out.

An XTM-based ontology for news domain is developed. The whole process is divided into some stages and use Notepad++ editors and Ontopia Application Software as XTM's Creation Tool which runs on Apache-Tomcat server.

The system is implemented by interfacing the ontology with Ontopia Navigator Framework (ONF) - a J2EE-based tool and embedded JavaServer Pages (JSP) codes for maintaining the XTM against issues such as dead links.

1.5 Contribution of the Study to Knowledge

The study establishes a pilot NEWS.net; which is an XTM-based web application, capable of maintaining XTM by resolving dead links. This would further enhance XTM for knowledge representation and interchange.

1.6 Structure of the Thesis

The rest of the thesis is organized as follows:

Chapter Two captures the technology: XTM and its concepts in a broader form and reviews some literature that is related to the topic. Chapter Three looks mainly at the processes and stages involved in developing the XTM and the System Architecture. Chapter Four is titled Implementation, Results and Discussion. It captures the requirements for the system to implement and also discusses the results. Chapter Five concludes the research work with conclusion and recommendation.

CHAPTER TWO: LITERATURE REVIEW

2.1 Introduction

This chapter is intended to capture the XML Topic Maps (XTM) technology and its concepts in a broader form and review the related literature of the study. Literature on XTM will be reviewed on general level and its usage, functionality and application.

2.2 XML Topic Maps (XTM)

Topic Map standard began in the 1990's from work on managing data organization (indexes) to computer documentation and was worked on for a good number of years before it finally became an ISO standard in year 2000 and a semantic web standard in 2001 with XTM owing to its XML interchange; Ellouze, *et al* (2009).

Markup languages have been in existence for quite sometimes. For example; SGML and SGML Topic Maps. Then we have XML and XML Topic Maps (named XTM). In early 2000, ISO/IEC 13250 standards, formalized the Topic Map interchanges syntax for the very first time. (This was identified by the acronym HyTM, which stands for HyTime Topic Maps). After sometimes, the parties of HyTM formed TopicMaps.Org, with the founding cochairs Michel Biezunski (of InfoLoom) and Jack Park (of Coolheads Consulting). The TopicMaps.Org had it as goal to bring the Topic Maps standard to the Web as XTM. There are two forms of Topic Map specifications; one is an ISO standard ISO/IEC 13250:2000 which specifies a standard interchange syntax based on SGML and HyTime, and one is a XML specification (XTM) developed by TopicMaps.Org (which defines an interchange syntax based on XML and Xlink) Grigoras (2005).

This thesis deals essentially with the XML specification of Topic Maps, thus when mentioning Topic Maps, it would mean XML Topic Maps (XTM).

The invention of XML was because many users felt that SGML was too difficult. There was a need to abridge and bound its features to those that are necessary for use in a Web context.

XTM was planned and constructed with the same enthusiasm: to abridge the ISO Topic Map specification for optimization benefit on the Web. An XTM maintains a superimposed nature on information resources as portrayed by ISO 13250. A major difference between ISO Topic Maps specification and XTM that would be mentioned is the fact that in the later, addressing is restricted to URIs in XTM while in the former, addressing can be articulated relatively with all kind of notations. This limitation identifies the unique role played by the Web as a worldwide network used as a common, well-defined (in contrast to the chaos web) mechanism on which information can be interchanged, Park and Hunting (2002); Grigoras (2005).

2.3 What are XML Topic Maps

Visiting a place for the very first time without a map could be sometimes very exigent, but when the objective is to locate the place as fast as possible, some kind of maps might be required. Likewise, thanks to the Web, lack of information is no longer an issue, but it is difficult to find quickly, directly accessing the required information. Thus what is desirable is an appropriate and well-suited information structuring. XTM is one of the technologies that proffer a suitable solution to this problem.

XTMs are conceptual structures that can predetermine knowledge and bond the encoded knowledge to appropriate information resources. They are structured around *topics*, that signify

subject of discourse; *associations*, representing relationships between the subjects; and *occurrences*, which connect the subjects to relevant information resources. Topic Maps may be characterized in many forms: in running programs as internal data structures, using Topic Map syntaxes in files, in the databases, or even mentally as a structuring standard in the minds of humans. These are all diverse means of representing the same abstract structure, Garshol and Moore (2008); Grigoras (2005)

XML Topic Maps are described as a language that defines language and as such XTM are not mostly integrated in the information they describe. Topic Maps are kind that exist between object of information from different sources that make feasible that all bits of information applicable to a certain concept, to be connected one with each other.

The intention of an XTM – defined document is to inter-relate information semantically. An XTM permit users to navigate topics that can be found in numerous documents. Topic may be expected to be a simple term, however, it is a link that includes a title and is referencing a point (mostly refer to as anchors) in the documents where they are occurrences of this topic. A Topic Map could be better described functionally as multi – document indexes, glossaries and thesauri; Pepper and Garshol (2007)

2.4 The TAO Characteristics of Topic Maps Model

While Topic Maps have been identified as a capable technology for ontology and semantic web, the next section explains the principal constructs of Topic Maps model; that is, the TAO (Topics, Associations, Occurrences) characteristics of the model.

2.4.1 Topics

The complexity and size of topic maps is determined by the number of topics. Topics are the main building blocks of topic maps, Pack and Hunting (2003).

In its most general view, topics can be anything: a human being, a concept, actually anything regardless of whether it exists or posses any other explicit feature. It forms the foundation for the (XML) topic maps creation. It can equally be seen as a ”multi-headed link, that points to all its occurrences”. This ”link” collects information about a given subject; Librelloto *et al* (2004).

According to (Grigoras, 2005) the appropriate way to describe a Topic is that, it is a 'location', a 'container', 'binding point', basically where we can go to find out 'everything what is knowable' about a 'subject'. A Subject is anything we may want to talk about.

A Topic is a 'proxy' for a subject, object we can store and manipulate in machines. Subjects are things we really cannot put into a computer.

A topic is expected to have a topic type or probably several topic types. Topic type can be described as that of class-instance relationship. Types stand for the classes in which topics belong to, that is, the group of one topic instance. Again, according (Librelloto *et al*, 2004) to Topic types are also topics own their own.

The schematic illustration in Figure 2.1 represents example of topics and topic types.



Figure 2.1: Example of Topics and Types

Minna is a topic of type City, Nigeria a topic of type Country and ChannelTv also a topic of topic type Media Organization.

2.4.2 Association

An association defines as a relationship between one or more topics. For all the classes of topics, associations and the roles involved in associations of Topic Maps are user-definable task, and of course, they are all topics (for instance, they are the subjects of topics that characterize them). Associations are composed of members – for every topic that involve in an association plays a definite member role in the association; Biezunski and Newcomb (2001); Grigoras (2005).

In some cases, it is hard to identify the suitable associations, that is, to precisely signify the relationship topics have with one another. The associations may be in this form: casted by, is a, part of, resided in, edited by, or more explicit like: is a financial organization within, it is indispensable to meet up with the conditions, and so on. A straightforward example for an association can be: the NTA International is located in Lagos (NTA International plays the role of Media and Lagos plays the role of State!). In order to prevent difficulties, association roles must be well stated, or else somebody could construe that Lagos *is located in* NTA International; Grigoras (2005).

2.4.3 Occurrences

In XTM, information resource(s) may be linked with a topic, and vice versa. These resources are termed as occurrences of the topic. Occurrences can be described performing a related function like the page numbers in a back-of-book index, Garshol (2004); Pepper (2000). The resources for

occurrences are either: resource reference (a URI - Uniform resource identifier) or resource data (expressing a piece of information about a subject) Pepper and Moore (2001); Anyi (2012).

The occurrences are, usually, external to the XTM document itself, though it might equally be in it, and they are designed at using, normally, URIs; Pepper (2004); Grigoras (2005).

2.5 Topic Identification: Name and BaseName

Topic normally has form of identification called name, again topic can also have no name or posses numerous names, and each of those names can take several methods. However, the most familiar case is a method of having topic with just a name. Again there is benefit of having many names in that it makes it feasible to illustrate interchangeable networks of knowledge. For example, each plant has a scientific name and at the same time a common corresponding language. Some names have different spellings or aliases like an instance of plant we have just mentioned. Names are not connected together by XML Topic Maps, but rather connect topics that might have several names; Biezunski (2002); Grigoras (2005).

XTM implores the use of scopes, which are topics, to distinguish names, as for instance the scientific name, maybe Botanical, and the corresponding common name, maybe English, as scopes on the names given to a topic.

It has been noted earlier that it is possible for a topic map to be nameless. For instance, an illustration like ‘For more information about the concept Y, navigate to …’ is obviously not exploitable, but in XML Topic Maps terms, this reference to a different location can be rendered as two different occurrences of the same topic; Grigoras (2005).

2.6 XTM Attribute: Scope and Namespace

According to Chen *et al* (2011) Scope identifies the degree of the validity of a topic characteristic assignment. It creates the perspective in which a subject or an occurrence is allocated to a given topic, and the perspective in which topics are connected via associations. To be precise, an affirmation of a topic's characteristic (association, name or occurrence) is legitimate only within a scope. When a topic characteristic declaration fails to indicate a scope, it is valid in the unconstrained scope.

In a broader view, scope is not only restricted to eradicate ambiguity, it is also useful in the area of information navigation, dynamically regulate scope based on the users' profiles during the inquiry procedure, to accomplish the intentions of being more explicit or definite in probing information.

The XTM fragment below defines a scope containing subject "English" using a published subject; TopicMaps (2013):

<scope>

<subjectIndicatorRef xlink:href= "http://www.topicmaps.org/xtm/1.0/language.xtm#en"/>

</scope>

A NameSpace is a class of base names of one or more topics, each of which is distinctive and all of which are the names of their respective topics within a single, common scope; Biezunski and Newcomb (2001).

2.7 XTM's Published Subject Indicators (PSI)

Published Subjects affords a platform by which humans and machines, in communication with machines, can understand when they are referring to same subject matter. They ascertain the uniqueness of subject of communication or interaction. The significance of establishing uniqueness or identity is unquestionable: in the absence of it, there wouldn't be interaction between humans and at the same time, operability between applications would not be there, Pepper and Garshol (2007). Hence, XTM provide us one method of achieving this, refer to as PSI, which is basically a URL or URI and, by shared agreement, ascertain the uniqueness (identity) of some subjects.

The common means of recognizing subjects in the present day networked settings is by addressing, to be more specific, by Universal Resource Location (URL). This would produce high optimization if the subject posses an unequivocal address, for instance, an information resource that is network-accessible. Thus, subject address is an address of a subject that is an information resource. However, there are subjects that are not information resources, the identity of those subjects can be created indirectly. XTM offers a method of handling this through what we called Subject Indicators; which is quite differ from the subject that it indicates. It can be any types: graphics, textual form, even description or names. From the viewpoint of a technical application, if two topics's identities are created by subject indicators that posses identical address, the topics are considered as representing the same subject. Therefore, subject indicator's address is regarded as subject identifier. Subject identifier and subject indicator are two forms of

the same identification platform; the former is for application and the latter for humans; Pepper, (2006).

Indicating a subject in XTM can be achieved in three forms:

- indicating through a <topicRef> element to a <topic> element which shares the same subject,

- indicating through a <subjectIndicatorRef> element to a resource that signifies the subject

- indicating through a <resourceRef> element to a resource that is the subject.

The subject shown by a subject indicator may be either addressable or non-addressable.

2.8 Producing an XML Topic Maps

Considering the simple Sport News taxonomy drawn in Figure 2.2, an XML Topic

Map would be produced, with the intention of presenting the procedural step on how an XTM can be achieved.

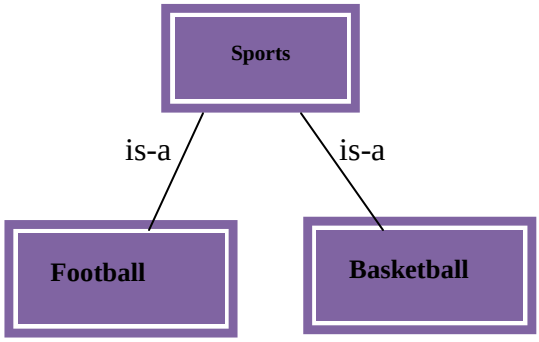


Figure 2.2: Sports News Taxonomy

Some of the salient elements of XTM needed to be introduced in this section are: <topic>, <baseName>, <baseNameString>, <scope>, <occurrence>.

Thus, topic elements with their identification tagged “id” need to be produced based on the Taxonomy on Figure 2.2.: <topic id=”Sports”>, <topic id=”Football”> and

<topic id=”Basketball”>.

In view of the fact that we have identified the topic elements Sports, Football and Basketball, then, we can have their basename and to be contained in the baseNameString tags as follows.

<topic id=”Sports”>

<baseName>

<baseNameString>Sports</baseNameString>

</baseName>

</topic>

<topic id=”Football”>

<baseName>

<baseNameString>Football</baseNameString>

</baseName>

</topic>

<topic id=”Basketball”>

<baseName>

<baseNameString>Basketball</baseNameString>

</baseName>

</topic>

Now we can provide more information or knowledge about any of the elements of the taxonomy. For instance, adding an image of a Footballer.

<topic id=”Football”>

<baseName>

<baseNameString>Football</baseNameString>

</baseName>

<occurrence>

<resourceRef xlink:href=”Rooney.gif”/>

</occurrence>

</topic>

The same procedure is applied to produce the XTM of the other elements.

The *xlink:href* aspect or attribute of the resourceRef element can be considered like the *href* property of a markup languages specifically, an HTML element. It references an information resource, in this case to the graphic file which is the occurrence of a topic called Football. The prefix to the href, xlink: which is like namespace portrays that the attribute follows the XLink W3C specification.

At this stage, more other elements can be injected: <topicRef>, <resourceRef>, <subjectIdentity>, and <subjectIndicatorRef>. We introduced now an identity to the topic “Football” that computers could understand as well as humans, namely the following URI: *http://www.manunited.com.uk/players/striker/fotogallaries/images/ronney.gif*.

Hence, the subject of the topic is formed as follow:

<topic id=”Football”>

<subjectIdentity>

<subjectIndicatorRefxlink:href=“

http://www.manunited.com.uk/players/striker/fotogallaries/images/ronney.gif.”/>

</subjectIdentity>

<baseName>

<baseNameString>Football</baseNameString>

</baseName>

<occurrence>

<resourceRef xlink:href=”ronney.gif”/>

</occurrence>

</topic>

Additional information about the topic can still be created by adding more elements. For instance: <scope> - may be in terms of numbers of languages we can express the topic; <associations>, <member>, <roleSpec> and <instanceOf>.

2.9 XML Topic Maps (XTM) Creation Tools

XTM, a knowledge representation can be created, navigated, and visualized via some tools, even though, some of them are still in their evolving stages. Hence, this section highlights a few synopses of some of the software tools that are available for the knowledge representation and retrieval. The attention of the tools discussed here is only drawn to the open source. Table 2.1 carries some of the tools with their description; Grigoras (2005).

Table 2.1: XTM Tools

Topic Map Tool	**Description**	**available at**
Ontopia	Ontopia is an open source application development platform for XTM. It is the most comprehensive suite of Topic Maps tools available, and consists of a number of interrelated products that address every aspect of the Topic Maps project life cycle, including topic map creation (both automated and manual), validation, maintenance, storage, querying and end-user delivery. The core component is the Engine, an easy-to-use SDK for Java developers.	http://www.ontopia.net/download/freedownload.html
Wandora	A general purpose information extraction, management and publishing application based on Topic Maps and Java. It has graphical user interface, huge collection of information extraction, import and export options, embedded HTTP server with several output modules and open plug-in architecture. It suits well for constructing and maintaining vocabularies and ontologies.	http://www.wandora.org/www/
Onotoa	*Onotoa*, a schema editor for Topic Maps, is an Eclipse RCP application, which is written in Java. Its whole design is based on the current TMCL draft, thus knowledge of TMCL would be helpful. To start *Onotoa* you need the newest Java Runtime Environment, which you can download here: http://java.sun.com".	http://onotoa.topicmapslab.de/tutorial
Perl XTM	Enables to read and write XML Topic Maps and to manipulate and query Topic Maps via an API.	http://sourceforge.net/projects/perlxtm/
Topic Map Designer	Tool for creating and viewing Topic Maps. (old tool)	http://www.topicmap-design.com/
TMTab	A tab-widget plug-in for Protégé (a free, open source ontology editor and knowledge-base framework) which enables the use of Protégé as	http://www.techquila.com/tmtab.html

	a topic map editing tool. The current plug-ins functionality is limited, because it supports only the export of XTM files from an ontology developed in Protégé, which is based on the "topic map ontology" distributed with the plug-in. Doesn't support the import of already existing ontologies. Above all, TMTab is not currently under active development. The last version of Protégé that it is known to work with is Protégé 2.1.2.	
TM4J	Topic map processing engine implemented entirely in Java and providing a pure Java API; among other facilities, supports the import and export of XTM syntax.	http://tm4j.org/
SemanText	A prototype application developed to prove how the topic map standard can be used to represent semantic networks.	http://www.semantext.com

From table 2.1, take for instance, Ontopia one of the most popular tools for XTM ontology is a collection of Java-based tools that consists of the following integral components; Dunaway and Wu (2013).

i. Ontopoly: used to construct and manage an XTM-based ontologies.

ii. Omnigator: it can be used to merge topic maps, to query XTM with the aid of tolog and essentially to browse XTM.

iii. Vizigator: it can be used within either Omnigator or Ontopoly to visualize XTM which enhance navigation easily.

Besides, there are some of the tools or the ontology editors that are designed for specialized domains such as education, Dicheva and Dichev (2006).

2.10 Ontology Process

The strengths of Semantic Web design and development heavily relies on the term ‘Ontology’. Thus, it is highly pertinent to comprehend the relevance of the term and the process of developing it; Rubbani (2007).

In semantic web, Ontology process can be viewed as a process of organizing data by way of creating subjects and relations. Subjects characterize concepts in a general domain but in XTM, it represents Topics. Hence, this process designs a data model that includes all vocabularies of subjects (topics) and properties. In a domain built by XTM, the concepts or subjects of the domain are referred to as Topics which help in organization and classification of data. While relations which are commonly refered to as Associations help to establish the relationship that exist between the topics and their related properties in the domain. The related properties which are described as attributes are used when topics create its instances.

Ontology can equally include some additional information about the topics. For instance, occurrences, information resources, and scope for the topics: Antoniou and Harmelen (2004).

When we create or configure ontology in some applications, the instances of the (subjects) topics automatically create their own also.

A typical example is shown for developing an ontology using OWL for Journalist that casts news and equally maintains the news based on different classes and geographical locations. Different classes of the news as Sports, Education, Economy, Science, Infotech, Entertainment are all distinct concepts or topics of the domain. And the distinctive locations of the news as Africa, America, Nigeria, and Kaduna are also distinct topics. Relationship (Association) can easily be

deduced from hierarchy of Topics. For instance, any news regarding Ahmadu Bello University or Universities generally will automatically be about topic called Education; Rubbani (2007).

When data needed to be knowledgeably stored or published in a structural pattern and also where retrieval or searches need to be carried out with the aid of relations, then the choice of Ontology becomes inevitable: a tool of necessity for the objective. With this type of tool for knowledge representation and retrieval, the machine can understand and process the required data.

However, the process for building ontology is similar to that used to build an object-oriented conceptual model or an entity-relation database diagram. The primary difference is that ontologies seek semantic expressiveness unconstrained by the various purposes the other models must fulfill. The following is a sequence of basic steps for building ontology; Park and Hunting (2002).

i. Define a universe of discourse, which simply means that portion of the world you intend to model. Such as: media and journalism, medical research, recording artists, the works of Shakespeare, etc. Start this universe-of-discourse-defining process by listing the concepts you want to include in that universe, that is, the meaningful objects of that corner of the world, as far as you understand them. XTM refers to these concepts or objects as Topics.

ii. List how the various things relate to each other. For example, Jack edits the news. Note that such things as edits are relations between two concepts. While doing this, you may discover things that you inadvertently left out in the first step. Well, add them in.

iii. Now build the ontology based on your analysis in steps 1–2 by using a knowledge representation engine (i.e., an ontology management tool such as Ontopia) to create classes for the concepts and relations.

In addition, some listed key items to consider when developing ontology are:

i. What the ontology is intended to be about (in general, plus assumptions)
ii. What you want to state in the ontology (in more detail)
iii. Concepts that should be included in the ontology (conceptualize the detail)
iv. Whether there are any ontologies in any set of available ontologies that may contain terms that can be used to develop the ontology
v. What modifications you may want to make to the ontology over time

If you would like to build ontology yourself, we suggest that you do so by using an ontology management tool. For example, download Ontopia, Protégé or any open source ontology editor available to the public for free.

2.11 Related Study

In the article of Dunaway and Wu (2013), the authors presented a coherent and practical understanding of creating a very large XTM-based ontology by means of combining the strengths of Ontopia and Wandora which are two different topic maps creation tools with the aims of achieving a very large scale domain. They considered the existence of some interoperability between the tools in order to realize the objective of multiple topic maps tools

integration. The whole essence of integration is that the strengths of one tool can complement the other's weaknesses.

But they admitted that it was time-consuming experience to understand both tools and to integrate them. However, we argue that since both XTM creation tools used in this article are suite of java tools, the researchers would have designed a program that can complement or solve the weakness of a tool preferably ontopia, and thereby integrate the program to it for efficiency in time consuming.

According to the thesis work of Ayin (2012) which sought to examine the extent of work that have been done about using XML Topic Maps starting from the range of year 2001 – 2011, and equally investigated the research and publication tendencies in Topic Maps. "The study was based on quantitative methodology, which was bibliometric analysis". The findings exposed the attention of Topic Maps from year 2001 to 2004 solely on data integration and interoperability, from 2005 to 2008 was on information theory, while from 2009 to 2011, its focus and applications largely lies on knowledge and intelligent based system.

In addition, Ayin recognized five subject areas in the research field of XTM, these includes semantic web, ontology, information architecture, retrieval and navigation, repository and content management.

Ayin concluded that there is a high tendency that the future research fields of interest on Topic Maps could be distributed knowledge management based on extended topic maps, intelligent topic map, knowledge navigation and retrieval improvement, collaborative e-learning system, knowledge service system, and knowledge representation modeling amongst others.

In this thesis work, we designed and implemented an XTM-based ontology for a news domain which is a typical study of the future research area recommended in Ayin's thesis.

The article of Chen *et al* (2011), presented a topic map based for modeling teaching system. Just like any other systems, teaching system is associated with numerous vocabularies (concepts) and very convoluted association of these concepts. As such, it is most tasking and time consuming to evaluate and systematize all the needed information in a knowledgeable and accessible manner. In order to solve the problem, a full advantage of XTM was applied to build a model to demonstrate the modeling technique from the angle of teaching and illustrate the use of Scope property of a topic.

The rationale behind the choice of XTM is the Scope feature that the topic of XTM posses which has the capability of eliminating ambiguity and also provide information navigation sources and thereby mitigate the time rate in information retrieval.

Scope is a very unique feature of topic (concept) in XTM which would have been better demonstrated if implemented in that domain of teaching system to really portray the beauty of XTM technology in building ontology and semantic web.

Considering also the paper presentation of Husáková and Olševičová (2008), they described the process of (XML) Topic Maps's application creation for the purpose of Observatory and Planetarium in Hradec Králové, Czech Republic. Two tools were used, TM4L editor for developing the ontology, and Ontopia Navigator Framework for creating the web presentation.

The pilot web presentation contains information and links to knowledge resources via the Topic Maps document (ontology) which helped to access web pages related to astronomy and to ensure

better navigation. However, they admitted that in a later future, effort should be focused on procedures for maintaining the Topic Maps for consistence utilization of the application, and visualization of the Topic Maps structure should be presented on the website. Thus, we argue that the first deficiency would have been addressed by further utilizing the ONF to embed what we called remedy codes.

According to Grigoras (2005) the weakness associated with UDDI registry of web services is as a result of lack of semantic. An exact keyword for searching item and data need to be provided for the UDDI interface. Grigoras explained a situation where consumer searches for a keyword car but the producer registered it as automobile, definitely the result could be very difficult to get. Grigoras addressed the problem with the aid of XTM technology as a good representative of semantic web.

Specifically, the aim of Grigoras's work was presenting XTM as a powerful agent for knowledge organization and representation of a particular domain, in Grigoras own case, bioinformatics. However, in the conclusion, Grigoras categorically stated the issue of dead links in the work and that lack of XTM maintenance is the cause since the creation tool of XTM is still evolving. Therefore, we are proposing to explore and exploit the ONF to provide remedy.

CHAPTER THREE: DESIGN OF XTM FOR NAVIGATING NEWS DOMAIN

3.1 Introduction

The first stage of the thesis was an involved investigative reading of articles of different journals, theses, and books in respect to building ontology and semantic web technologies. The main technology learnt and applied is XTM. In this chapter, we applied the technology to design an XTM-based ontology otherwise known as XTM Document for the News Domain. It also contains the system architecture.

3.2 Basic News Ontology Design

It is paramount in this section of the work to understand our case study and its requirement data. Designing ontology is similar to database design. The difference is, in (relational) database design, tables are required that is composition, of rows and columns with keys (primary and foreign) and data must be added to these tables strictly according to their design specifications. In ontology design (according to XTM specifications) we make topics, occurrence and relationships, called associations, individually and then add these associations to the topics. An association (part-of) might be present between more than two topics. As soon as ontology is being built then instances (of these topics otherwise called classes) can be instantiated that store in a model. The model could be in database or file.

The Figure 3.1 depicts the schematic design of our case study that we refer to as “News Ontology”. The design consists of “nodes” and “edges” which correspond to “topics” and “association” respectively in XTM.

The nodes (for example, Information Technology news, Education news, Business news) at the first level of the design represent the subclasses news of the super class (topic type) - NewsDomain. A subclass node or topic can equally be a super class. Sports News and ForeignScene News are typical examples. A topic can have relationship (that is, association) with one another.

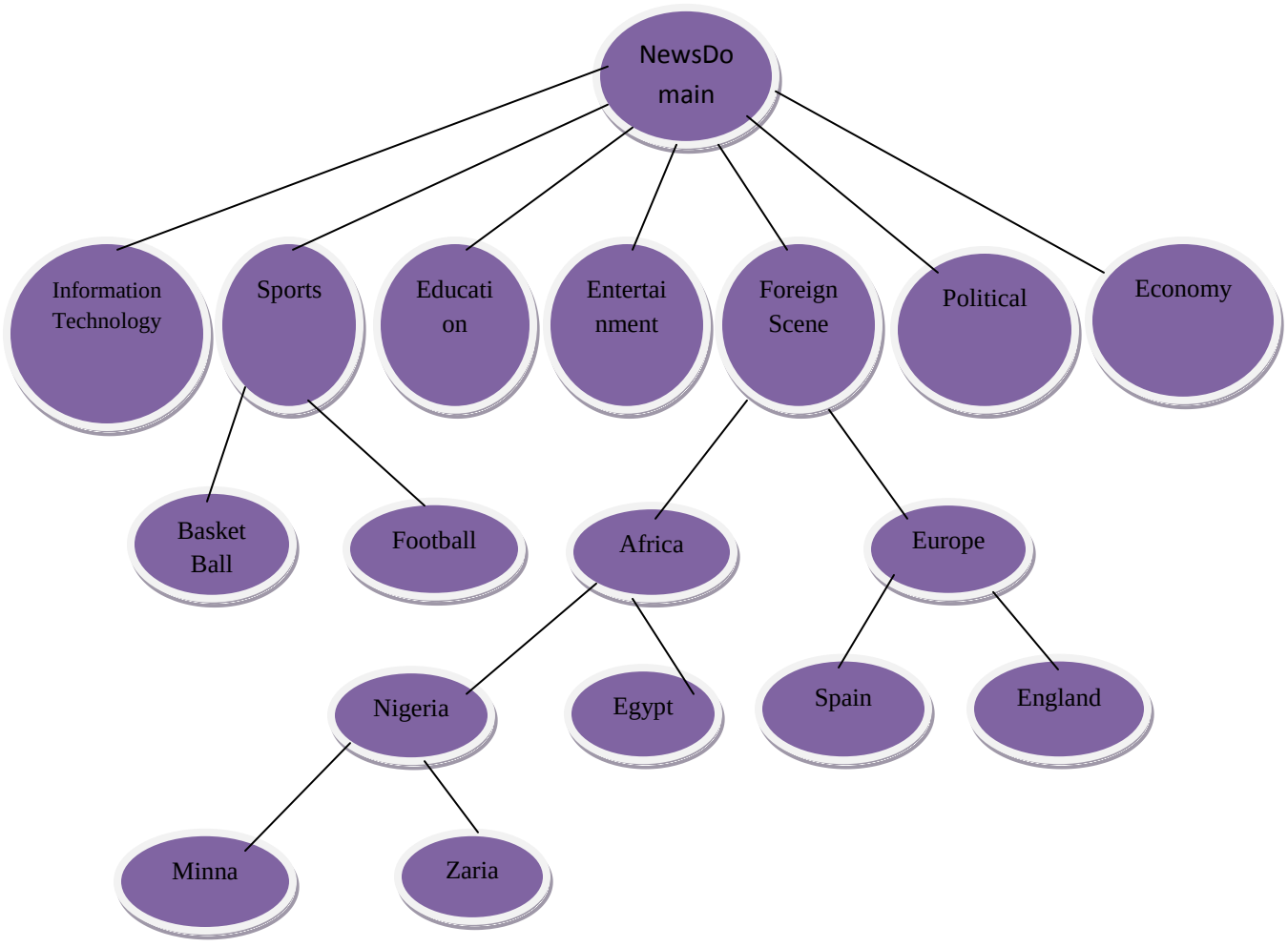


Figure 3.1: News Ontology

Figure 3.1 is modeled as follows: An NO-model (News-Ontology-model) structure is a 3-tuple structure. That is, NO-model: = (R; N; T) where

R is the root topic of NO-model

N is the set of Non Terminal topics of NO-model.

T is the set of Terminal topics of NO-model.

Root NO-model (RNO-model) structure is defined as RNO := ({*ROOT*}, Ø) . ROOT is the root topic, each other topic must subclass ROOT (it may do so indirectly). Each level of NO-model must include RNO and thus gain visibility to the root topic. This concept of the NO-model is related to object-oriented languages approaches. For instance, in Java every class extends java.lang.Object class.

An NO-model represents a self-contained unit of structured information that may be reused. Generally, each element or node in NO-model is refers to as topic (t) weather R, N or T. Thus, the ontology contains a set of n topics.

Assuming the proposed NEWS.net is an online news media that is coming into the market of news. After studying the markets, an innovative technology (XTM) for knowledgeable representation of data that organizes news data in an efficient manner is introduced in order to gain edges over competitors. For instance, when giving only topic name like Africa during news data creation, the system should be able to retrieve its instances as Egypt or Nigeria. Again, instead of the user, the system should be able to identify links that are dead. To this end, a semantic similarity model to measures XTM-based ontologies that behave in the manner of the proposed system described in this section is examined.

Semantic similarity models are specific types of semantic measures: a mathematical techniques used to measure the strength of the semantic relationship between concepts (topics) or instances, via a numerical description obtained according to the comparison of information formally or implicitly supporting their meaning or describing their nature. Practically, semantic similarity can be measured for instance by defining topological similarity, by using ontologies to define a distance between topics; IJntema *et al* (2010).

The semantic similarity measures of the XTM-based News ontology is based on *is a, part of, casted by, created by* relationships and checking the semantic relatedness by way of measuring the similarity between news article being read and user profile created by articles read.

The user profile is represented as the following set:

$$U = \{t^u_{1,}\ t^u_{2,}\ t^u_{3,}\ \dots, t^u_{y}\} \text{ where } t^u_i \in NO.$$

The user profile consists of y topics identified by NO in the news previously read by the user.

A news article can also be formulated as a set of q topics that appears in the article:

$$N = \{t^n_{1,}\ t^n_{2,}\ t^n_{3,} \dots, t^n_q\} \text{ where } t^n_{j,} \in NO.$$

Semantic similarity can be measured by Ranked, Binary Cosine, Jaccard or SemRel models. However, the results of Binary Cosine, Jaccard and SemRel are not always accurate and reliable. In fact Binary Cosine and Jaccard do not consider the numbers of occurrences of a topic in an article. Also, they do not consider the topics related to the ones found in an article. Therefore, Ranked Semantic Similarity is adopted. When you read news about topic t_1 which is related to topic t_2 and topic t_3 you enhance not only your knowledge in topic t_1 but also in the other two

topics. Every topic is assigned a value called rank. For instance, when a user reads about SportingsNews, same user might also be concerned in others, like BSkyBNews, but also in news is about Football. Both are considered to be in direct relation to the topic SportingsNews. Thus we increase the rank for SportingsNews, BSkyBNews, and Football. Unrelated topics, that is, topics that are not directly connected to the current topic, also need to be addressed. This means, if a user profile consists of topics $t_{1,}^{u}$ and $t_{2,}^{u}$ and the next news the user reads, contains topic $t_{3,}^{u}$ which is directly related to $t_{1,}^{u}$ but not related to $t_{2,}^{u}$ we increase the rank of t_1, and decrease the rank of t_2. By decreasing the rank for such a topic we make the user profile adaptive to the user's main interest.

The set of related keywords to topic t_i is defined as:

$$r(t_i) = \{ \ t_{1,}^{i} \ \ t_{2,}^{i} \ \ t_{3,}^{i} \dots, t_{p}^{i} \ \} \qquad \dots 3.1$$

R is defined as the union of all related topics to the topics in the user profile:

$$R = \bigcup_{t_i^u \in U} r(t_i^u) \qquad \dots 3.2$$

And finally U_X is described as the set of all topics and corresponding related topics, it is referred to as the extended user profile:

$$U_X = U \cup R. \qquad \dots 3.3$$

The use of extended user profile is significant because it will raise the interest of user in certain topics that are not in the user profile, but are related to the topics in the user profile. To compute the final ranks for each topic, we organize the topics in a matrix. This is done because we have to assign a rank to each topic in the extended user profile for each topic the user has read about. Reading about topic t_1 increases its value with 1.0. If topic t_2 is directly related to topic t_1, then its value is increased with 0.5. If there exist a topic, say topic t_3, in the extended profile which is not

equal to topic t_1 nor related to topic t_1, its value is decreased with 0.1. These constants were determined by experimenting with values ranging from 0 to 1 with a step size of 0.1. Thus, the matrix's columns contain the items from the extended user profile (U_X) and the rows contain the items from the user profile (U).

Table3.1 shows a rank matrix, where $t_i^e \in U_X$ and $t_i^u \in U$. Adding the values of the cells in a column of the matrix, and repeating this process for each column, results in a vector with the final ranks for each topic, in the extended user profile.

Table3.1: Ranked Matrix

	t_1^e	t_2^e	t_3^e	…	t_x^e
t_1^u	r_{11}	r_{12}	r_{13}	…	r_{1x}
t_2^u	r_{21}	r_{22}	r_{23}	…	r_{2x}
t_3^u	r_{31}	r_{32}	r_{33}	…	r_{3x}
.	.	.	.	…	.
t_y^u	r_{y1}	r_{y2}	r_{y3}	…	r_{yx}

The user might have read one or more news about a topic. Rationally, the user is assumed to be more interested in topics that are found in some news. The number of news the user has read about topic t_i^u, is referred to as the weight w_i^u.

$$W = \{ w_1^u, w_2^u, w_3^u \ldots; w_y^u \}. \quad \ldots 3.4$$

Thus, the value for each cell in the matrix of Table3.1 is computed using:

$$r_{i,j} = w_i^u * \begin{cases} +1.0 & if \ t_j^e = t_i^u \\ +0.5 & if \ t_j^e \neq t_i^u, \ t_i^e \in r(t_i^u) \\ -0.1 & otherwise \end{cases} \quad \ldots 3.5$$

The final rank for each topic from the extended user profile, can be calculated by adding the values of the corresponding column in the matrix:

$$\text{Rank}\ (\ t_j^e\ ) = \sum_{i=1}^{y} r_{i,j} \qquad \ldots 3.6$$

The additions are kept in a vector V_U. Every topic in the extended user profile now has a rank. We need to ensure that the range of the ranks is [0,1] before we can compare the user profile with an unread news item. The normalization is done as follows:

$$V_U[t_i^e] = \frac{t_i^e - \min\ (t_i^u\ )}{\max(t_i^u) - \min\ (t_i^u)} \qquad \ldots 3.7$$

$t_i^e \in V_U$ and $t_i^u \in V_U$. With this normalization the extended user profile can be compared to a new item that needs to be classified. The new item consists of a set of topics, designated as I:

$$I = \{i_1, i_2, i_3 \ldots, i_h\} \qquad \ldots 3.8$$

The vector containing the ranks for this item is described as V_I:

$$V_I = (b_1, b_2, b_3, \ldots, b_c), \qquad \ldots 3.9$$

$$\text{Generally, } b_i = \begin{cases} rank(t_i^e) & if\ t_i^e \in I \\ 0 & if\ t_i^e\ !\in I \end{cases} \qquad \ldots 3.10$$

Each topic from the extended user profile that appears in the item is assigned the same rank as the one in V_U. The remaining topics are assigned zero. Topics appearing in the item but not in the user profile are ignored.

In this research work it is assumed that all topics found in a news item are equally important. To compare the news item with the user profile we proposed to calculate the level to which the item fits the profile by dividing the sum of the ranks of topics in the item by the sum of the ranks of the topics in the user profile

$$\text{Similarity } (V_I, V_U) = \frac{\sum_{b_i \in V_I} b_i}{\sum_{t_i^u \in V_U} t_i^u} \quad \ldots 3.11$$

Having experimenting with values ranges from 0 – 1 with a step size of 0.1, the news item with the highest similarity measure fits best with the user profile.

3.3 Design Process of XTM

Generally, in Software Development Life Cycle, whenever a system has to be developed it is pertinent to judiciously follow some well-defined procedures to attain the goal of the system. Consequently, when a new ontology is set to be developed, series of salient questions in respect to the tools, methodologies, languages among others arise. For instance, (Corcho *et al.*, 2002) which methods to adopt in developing the ontologies, either from scratch or reusing existing ontologies available at ontology server? Are the ontologies stored at the databases or files? And host of other questions. However, there is no known standard of modeling domain but some alternatives viable ways exist. For example, such known alternative was designed at the Stanford University (Noy and McGuiness, 2000). It expresses a possible means to develop ontologies with the aid of software tool called Protégé.

Hence, in commencing the development of XTM-based ontology, there are step by step methodologies that need to be followed. The methodology consists of the following seven steps.

i. defining the target.
ii. build and manage ontologies
iii. XTM Software application
iv. setting up the XTM concepts
v. allocation of information resources
vi. establishing association between topics

vii. testing the XTM

The following subsections engage these steps in a more practical detailed and explained what measures and questions have been looked into during each step.

3.3.1 Defining the target

Defining the target or what we could be termed as define a universe of discourse, simply means that portion of the world we intend to design. We would start this universe-of-discourse-defining process by listing the concepts we want to include in that universe, that is, the meaningful objects of that corner of the world, as far as we understand them. As it has been captured, one of the objectives of the thesis is to develop an XTM-based ontology in the area of News Domain.

Thus, the main consumers of this ontology may be journalists and general public in search of earlier unknown resources on the internet.

3.3.2 Build and manage ontology (ontology usage)

An inevitable question at this step that always come to mind and must not be ignored is weather an existing ontology should be reuse for the domain of discourse or a new one should be completely created. But Ellouze *et al.* (2008) in their work admitted that presently XTM are usually manually developed by domain experts.

It is often argue among the semantic web community that the development of a new knowledge representation (ontology) should not be taken too lightly or miscalculated in that it is quite time-consuming and most often requires the implication of domain specialists. Ontologies act as representation of knowledge out of a specific domain and thus envisage to be reused.

However, in a situation where the author of the proposed ontology wants to test his shrewdness or expertise, then creating a new one becomes inevitable. Presumably ready to pay the price of time consumption and the cost of consulting the domain specialists. Thus, the later argument is exactly the case of this study under the summarize guidelines and requirements in Table 3.1 drawn with the aid of domain experts.

Table 3.2: Guidelines for the XTM Document

User's requirements	**Priority**	**Comments**
Capturing the ontology of news terms	High	Highly essential because it is a news based ontology
Implementation of external occurrences	High	Users are often work with hyperlinks
Defining internal occurrences	Medium	Internal definitions are not necessary
Implementation of associations among topics to enable navigation	High	It is expected to be one of the main benefit of users
Topics and Topics Types implementation	High	Highly essential because its form the relationship
Defining topic names in different languages	Low	It is limited to English Language only

Thus, we have resolved to develop and manage a new ontology with the aid of Ontopia application which is shown in Figure 3.2

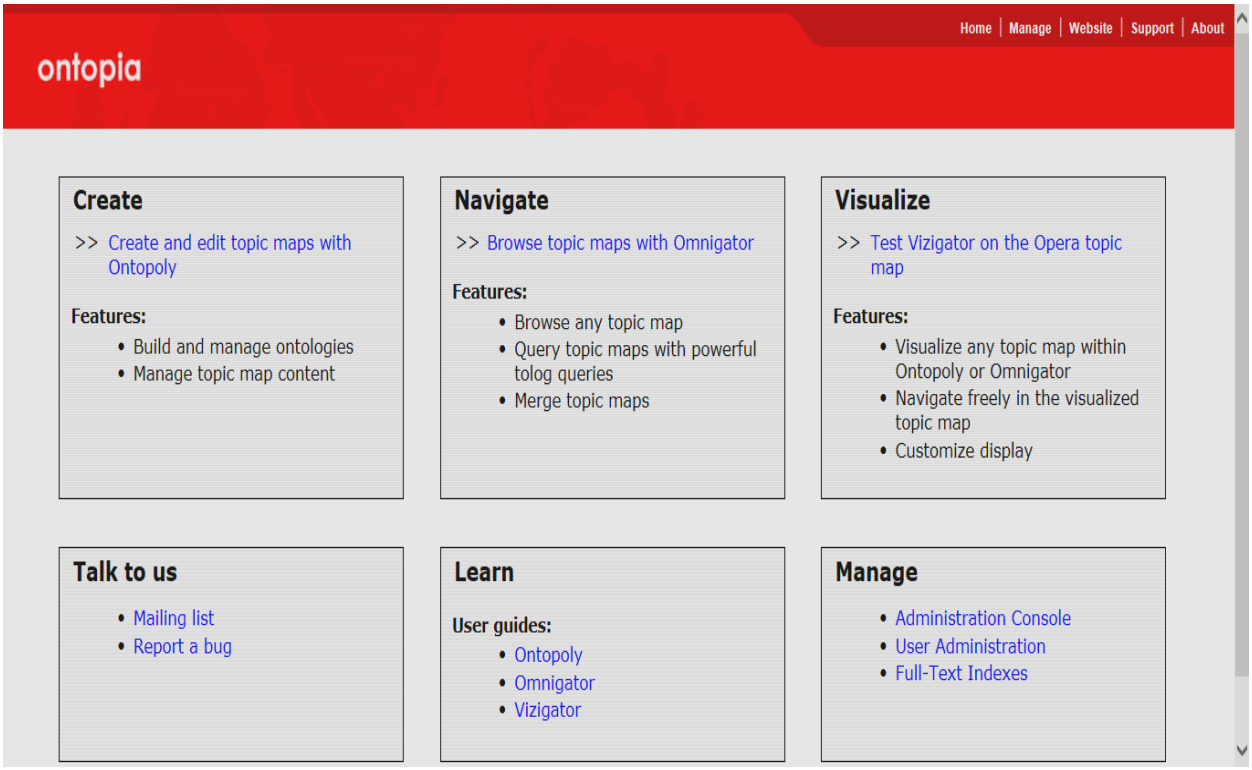


Figure 3.2: Ontopia Application Home Page

3.3.3 XTM software application

The third step in this methodology is very significant, in that it is at this step we decide in what application the knowledge base has to be developed. The XTM would be created manually in a simple text editor (notepad++) and the ontopia's editor itself to create a new topic map called, "news.xtm". We define the following requirements for the XTM's developing application.

i. Powerful and flexible graphical visualization and navigating tools for the XTM document content are required, where each topic is represented as a node in the graph like semantic network.

ii. Periodical automated testing of functioning web links, stored as external occurrences in XTM document, is important.

iii. An application that is pleased with tolog. Tolog is a query language for XTM which is loosely based on Prolog programming language and some features of Structured Query Language (SQL).

iv. Simple installation and maintenance of any platform. Although, windows version is used on this thesis, low costs on future enhancing the application.

There are many applications to design the XTM-based ontology, but the very one adopted is called Ontopia, version 5.1.3 (2011), a knowledge suite of java tools application that consists Ontopoly, Omnigator and Vizigator.

Ontoploy is a built-in ontopia's editor, while Omnigator is the browser for XTM-based ontology and Vizigator is a tool that visualizes the ontology developed.

The choice of Ontopia is motivated by its ability to fulfill most of the requirements earlier mentioned in this section. Also Built on top of it is the Navigator and Web Editor Frameworks that allow web developers to create powerful applications.

Other strengths associated with the tools: it has the capacity to save ontologies in RDF and later export it to XTM. It equally allows ontology to be written in Linear Topic Maps (LTM) and capable to export it to XTM format. Table 3.2 summarizes its comparison with others.

Table 3.3: Comparison of some XTM Application Tools

Requirements	Ontopia	TMTab	Wandora	Onotoa
Periodical upgrade	Yes	No	Yes	Yes
Web application frameworks	Yes	No	Yes	No
Ability to save ontologies of other technology(RDF,LTM)	Yes	No	No	No
Visualization	Yes	Yes	Yes	Yes
Perceptiveness	Yes	No	Yes	No
Low Price	Free	Free	Free	Yes

3.3.4 Setting up the XTM concepts

It was noted earlier that Topic, Association, and Occurrence are the three major concepts or characteristics of building XTM among others. The first concept considered in this step of the methodology is Topic. It was earlier stated in Figure 3.1 that the nodes are referred to as Topics in XTM. As a result, it makes it easier for us to set up our topics (football) and topic type (sports) appropriately.

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<topicMap xmlns="http://www.topicmaps.org/xtm/1.0/"
      Topic id=”sports”   xmlns:xlink="http://www.w3.org/1999/xlink">
<topic id="football">
<baseName>
<baseNameString>Football News</baseNameString>
```

</baseName>

</topic>

</topicMap>

Besides a topic having type, it is also expected to have form of identification called id and a base name enclosed in a base name string tags. Figure 3.3 shows some concepts of the XTM built.

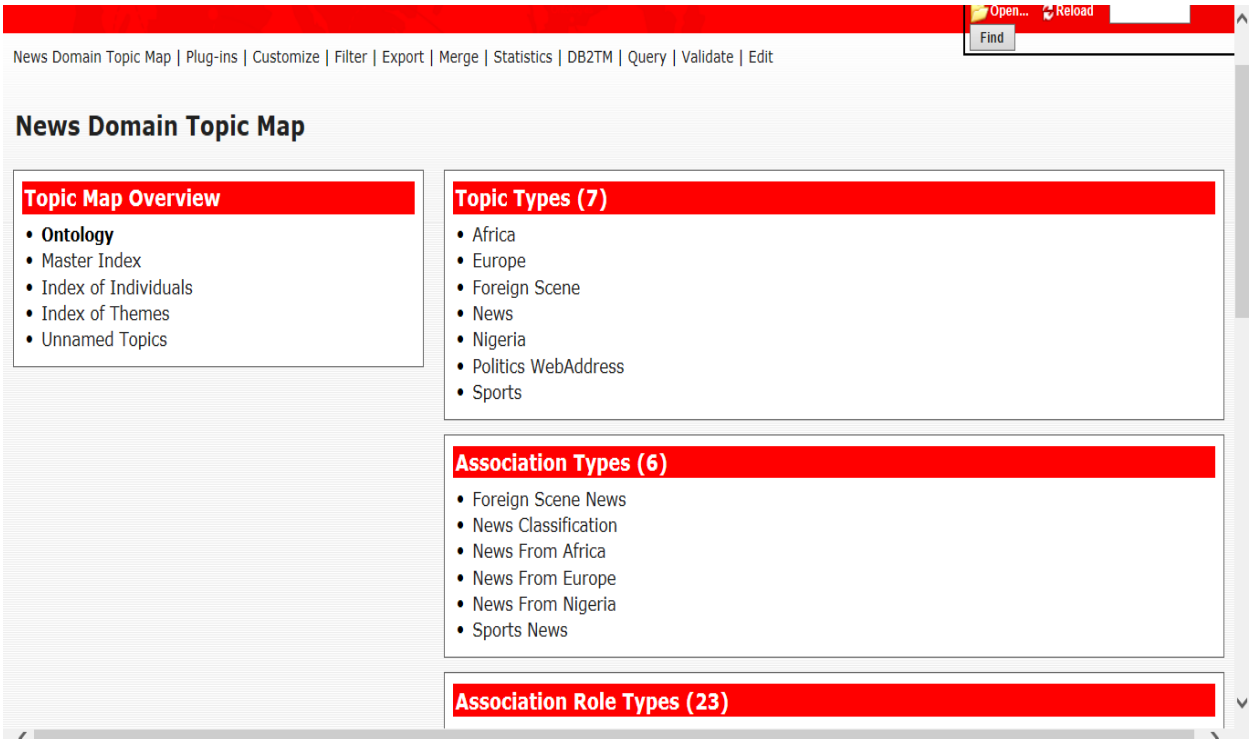


Figure3.3: Concepts of XTM

3.3.5 Allocation of information resources

While setting out id, name, even scope for topics, information resources in most cases are clearly defined and assigned for the topics. Generally, in XTM, such information resource is an addressable entry for instance; URI enclosed in a subjectIdentity and subjectIndicatorRef tags. It is attributed to the XTM concept called Occurrence.

Every Topic is expected to have a unique and corresponding resources; resources in terms of URL or published subjects indicators (PSI) for accessing any particular choice of news available in the domain.

Figure 3.4 a prototype that demonstrates what forms of resources can ideally be assigned to a topic of an ontology defined in XTM. The space-like shape represents ontology with the small circles and lines in between representing topics and relationship respectively. The rectangular shapes denote occurrences (internal or external) of ontology being URLs or PSIs and the arrows representing topic linking up with or accessing the occurrences.

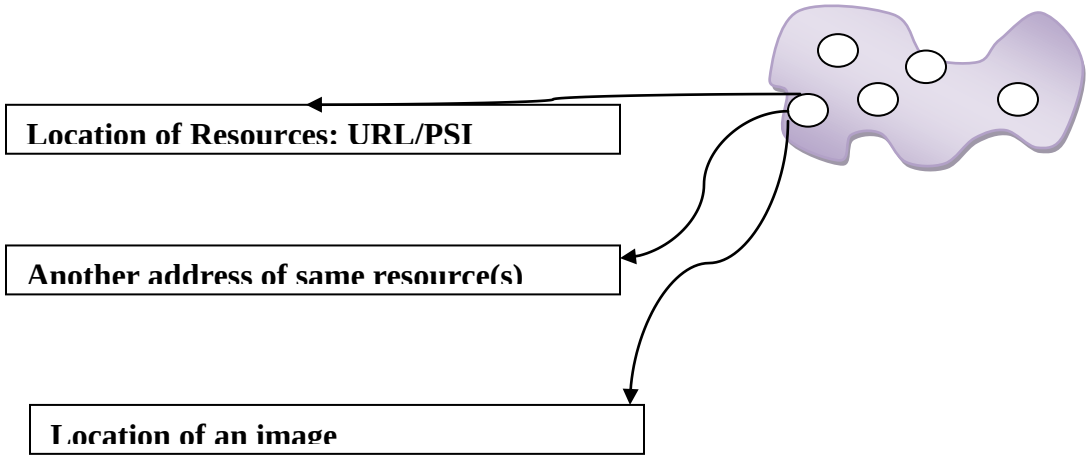


Figure3.4: Ontology, pointing to different information resources (URL).

3.3.6 Establishing Association between Topics

It is of no doubt that in using XTM to build ontology, relationship exist between topics. Especially, those topics developed at figure 3.1. Thus, this step sought to establish the relationship. And it is called “association” that exists between topics in topic maps. Therefore, each edge in the ontology design of the figure3.1 is either a “part-of” or “is-a” relation. For

instance, BusinessNews is part of the NewsDomain or better still NewsDomain is a News platform.

3.3.7 Testing the XTM

The XTM document needs to be tested to ensure the workability of the News Domain Ontology. That is, to prove that the information and values contain in our ontology are reliable and error-free to our target, we exploit the Omnigator of Ontopia Knowledge Suite to load the document (XTM). The moment it discover error(s), the document would not load until the necessary actions to debug take place.

At this step, we can now conclude that we have successfully developed an ontology using XTM, which is still subject to build progressively. In other words, we have designed a structured document that is set to browse using Omnigator. For example, Figure 3.5 shows that topic-Entertainment is clicked and all the appropriate information regarding the topic are shown. The same effect is applicable to any other topics of the domain.

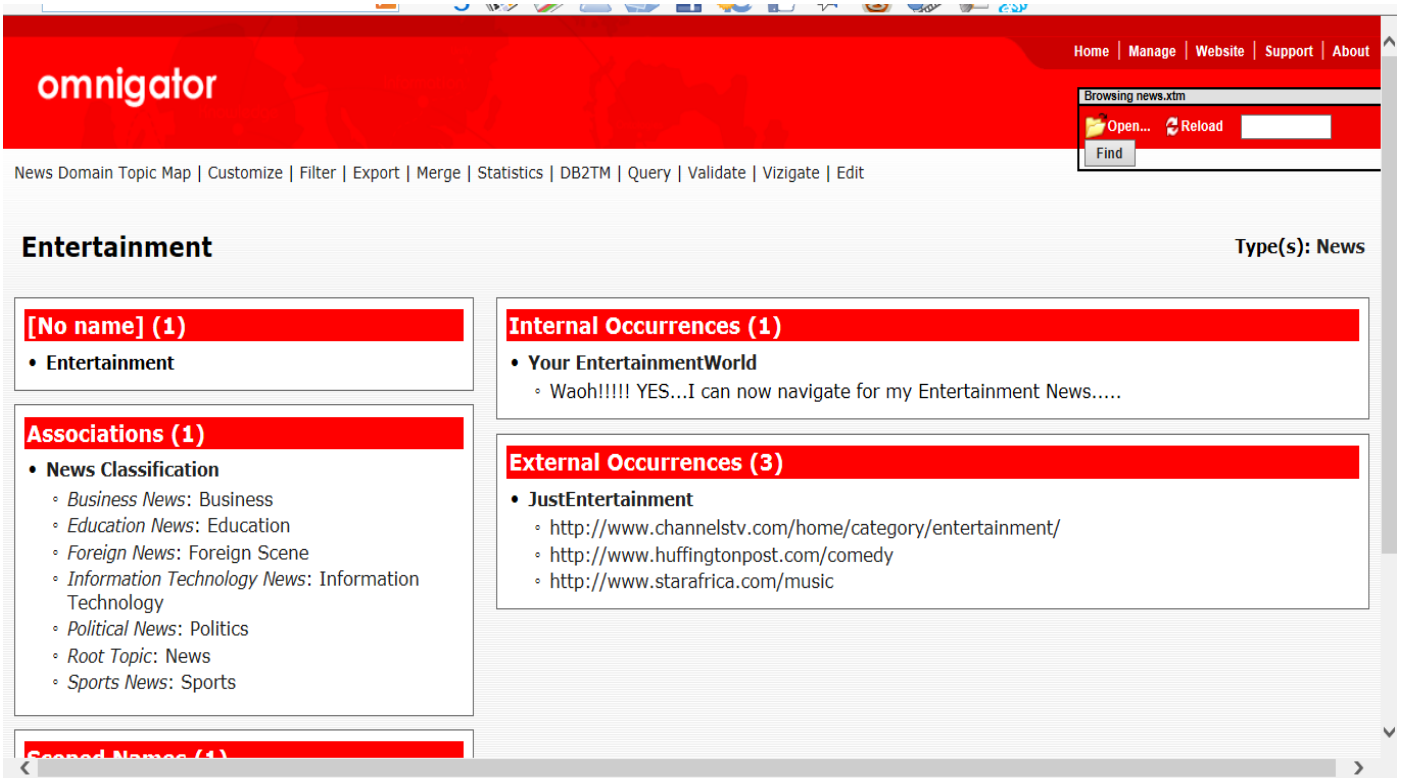


Figure 3.5: Interface for browsing the ontology.

From the right side of the figure we have some links as information resources to the subject of the topic which is tabbed as internal occurrence and external occurrences, in this case, there are four numbers of these resources. Internal occurrence can be statement about the topic, images, or any related files of the topic. External occurrence is basically external links. And on the left side, we can navigate through the XTM by clicking any of the news categories in the association tab Obviously, we confirm that the topic-Entertainment is involved in the "is-a" and "part-of" associations to the root topic.

The figure equally depicts that only topics arranged at the initial level of depth can be accessed directly. The XTM document is not presented as a whole. Should a user decided to access a topic

placed on a lower level in the graph, the accessibility would be achieved via clicking from the initial, then the intermediate(s), and lastly to the desired level (topic).

3.4 Visualizing the XTM Document

Graphical visualization of the document is quite appealing because XTM is often viewed or judged as network of topics (Park and Hunting, 2002)".

Vizigator an OKS supports this features, which visualizes any XTM within Ontopoly or Omnigator, navigate freely in the visualize XTM document and customize display. With the aid of this facility, users should be able to have a broad view on the document, thereby making the choice and accessibility of the desire level of the graph's depth possible as represented in Figure 3.6.

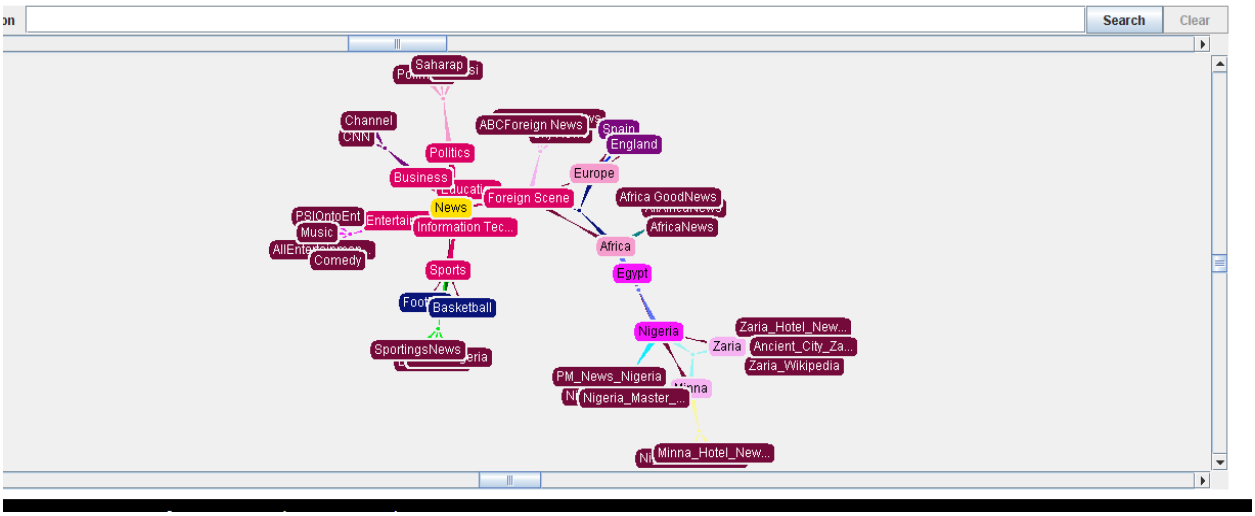


Figure 3.6: News Ontology Vizigator

At the top left corner of Figure 3.6 we can fill in the box the desired number of levels to be displayed (in our case - 2). Small red boxes with numbers indicate further depths in the representation. In the same figure the two relationships/associations is distinguished through different colours; pink edges stand for the "part-of" and purple edges stand for the "is-a"

associations. The start topic is always displayed in yellow and represents the topic where the user usually starts his navigation. The thick purple topics represent occurrence topics for user profile. But the Vizigator offers additionally an alternative way to find topics, the search box. All or part of a topic name has to be filled in the search box and the hit will blink in green for instance, the topic-sports as shown in the figure.

Now, in case we decide to get the detail characteristics of a topic (in this case - sports) all we have to do is to right click on it and the result will be shown as in figure 3.7.

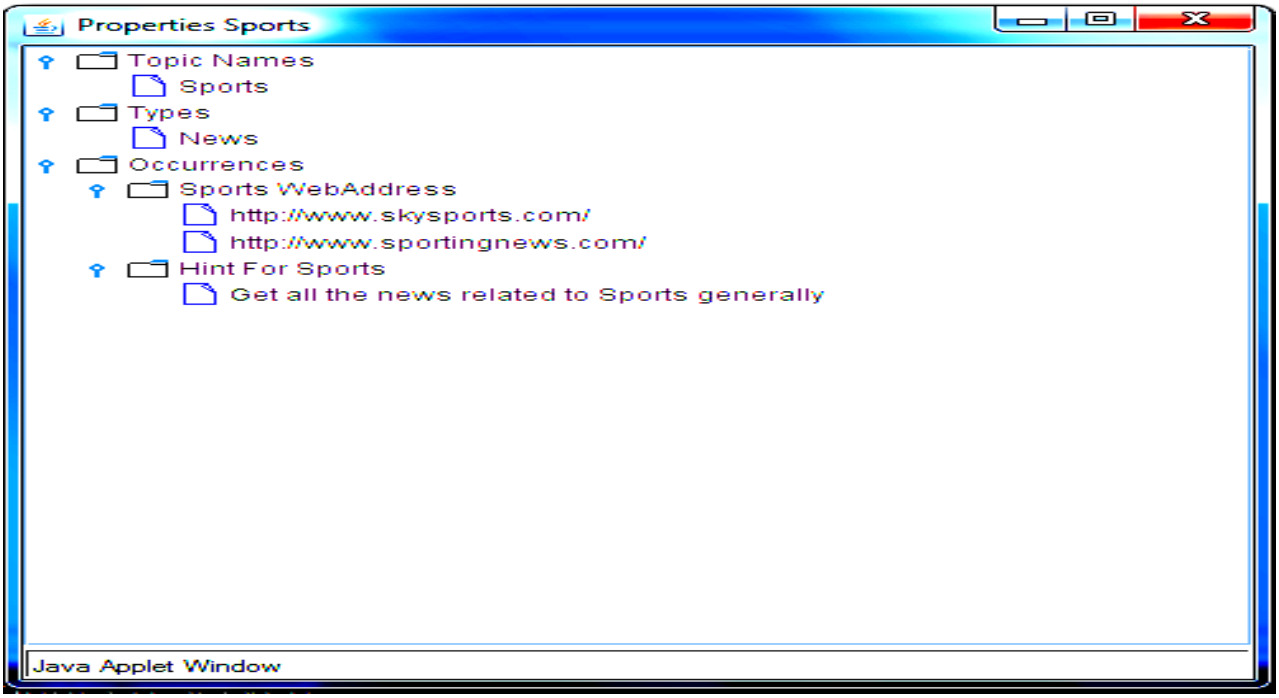


Figure 3.7: Vizigator Properties of (Sports) Topic

3.5 System Architecture

The architectural design of the system which is described as 3-tier (Data tier, Server/Maintenance tier and Users Tier) architecture proposed for the implementation of NEWS.net is largely on Ontopia tool and other related applications.

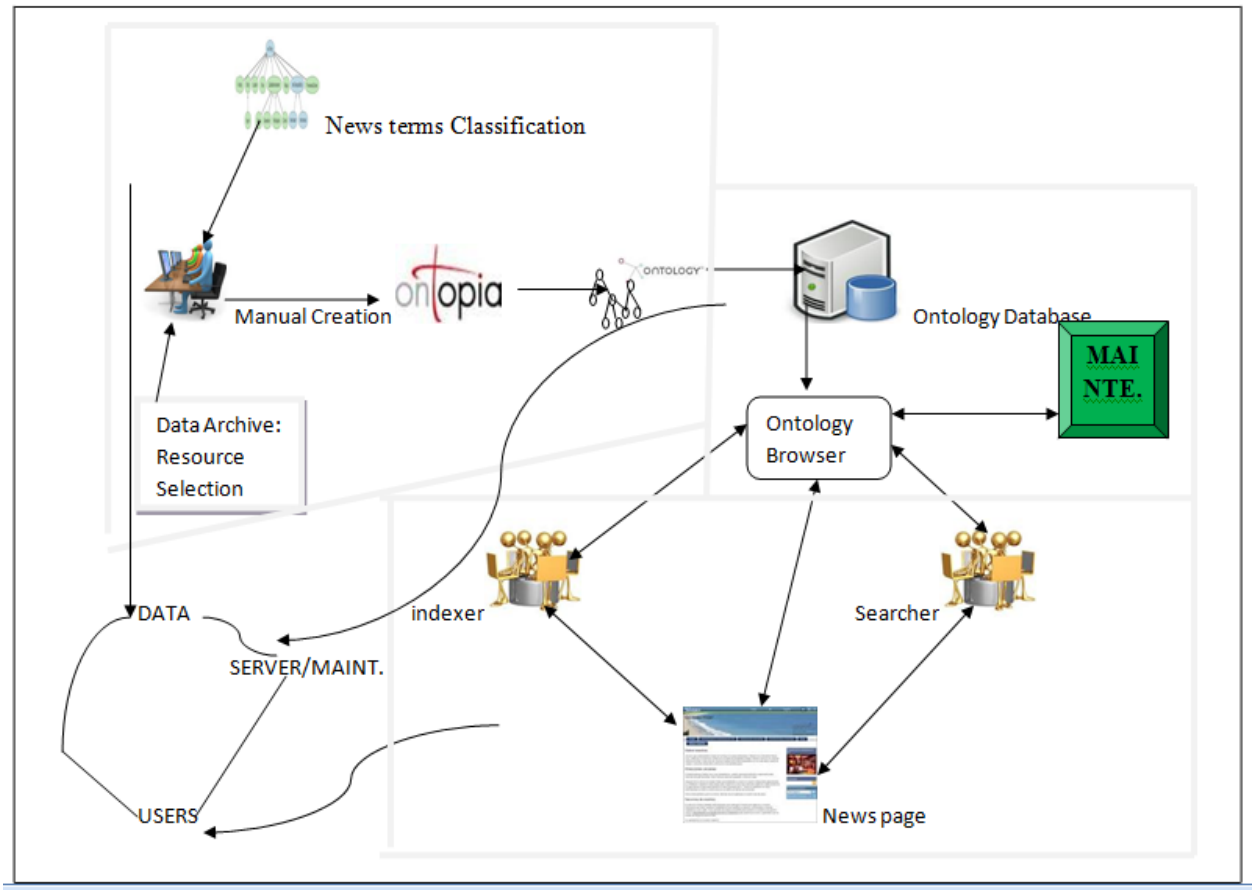


Figure 3.8: System Architecture

3.5.1 Data tier

Data tier consists of developing XTM model manually or what could be described as building an XTM-based ontology. This tier is accomplished by converting the news ontology model in figure3.1 into XTM using the Ontopia-5.1.3 application. Resource selection in accordance with the terms of the news from topic maps and google form part of the data used. The XTM-based ontology is stored at the topicmaps (tomcat) server of the ontopia which acts as database.

The tier considered and analyzes all the required data for developing the XTM. The step by step procedures of designing this tier have been extensively described in section3.3 of chapter3 of this thesis.

3.5.2 Server/Maintenance tier

The Ontopia software is a very robust (J2EE) application. Its runs on Tomcat Web Server and the server equally acts as database for the XTM developed in the first tier. The server is like a container where you retrieve every stored data accordingly "garbage in, garbage out".

The XTM-driven ontology is browsed by a built-in application of Ontopia called Omnigator which enable users to browse for news item of their choice.

XTM at this point needs to be maintained to resolves topic(s) that does not has semantic similarity with the topics that form user profile or in order words, dead links among selection of resources. The maintenance is done by a platform called Ontopia Navigator Framework (ONF). Importantly with this framework, the ontology can be browsed via any client-side web browsers (Internet Explorer, Firefox, Google Chrome etc).

Therefore, the proposed NEWS.net would be implemented by way of interfacing the framework with the XTM and becomes a user web application in the third tier of the architecture. It is important to note that web pages written using the navigator framework are written as JSP, and JSP is a Java technology that lets you write web pages as HTML with embedded Java code.

3.5.3 Users tier

The Users formed the third (final) tier of the architecture. The NEWS.net will be accessed and navigate by users at this phase of architecture. Users are split into two: indexers and information

searchers. These provide a good audience for assessment, representing various levels of knowledge and different needs and ways for accessing the application.

The application will first check user's computer network connectivity status – weather is connected or not and finally reports any dead link to user.

CHAPTER FOUR: IMPLEMENTATION, RESULTS AND DISCUSSION

4.1 Introduction

This chapter would be looking into the system implementation, the results of the technology developed and maintained for navigating the domain in question. Therefore, we will implement a system by using the ontology we have developed through the ONF technology. However, before implementation, we will first consider the system's hardware/software and functional requirements.

4.2 System Requirements

For the implementation of the proposed system, we need to analyze and discuss the (hardware and software) requirements at this stage. Thus, the minimum hardware requirements for the implementation of the system are as follows:

i. A Pentium 4 equivalent CPU 1.8 GHz processor with 512MB memory

ii. A hard drive with 40GB capacity

iii. A graphic adaptor with screen resolution of 800 x 600 pixels 32 bit quality and 15.4" monitor

iv. A CD/DVD Rewritable drive

v. 10/100 Mbps Ethernet LAN card

vi. Internet ready with 56k modem for dial-up access

vii. Wireless LAN with 802.11b/g wireless network card

viii. Laser jet printer and line printer for draft copies

The minimum software requirements for the design and implementation of the systems are equally as follows:

i. Java 2 Platform, Software Development Kit (SDK) version 1.5

ii. Windows XP Professional

iii. Apache (Tomcat) web server version 1.3

iv. XTM's tools: Ontopia

v. Notepad editor

vi. Web programming languages skills: HTML, XML, JSP

4.2.1 Functional requirements

News.net, an instance of our case study is a branded new enterprise news supplier who is coming into news market and requires a system that could assist to arrange and classify their news data and information in an effective and efficient manner. The branded new enterprise of course, demands to apply such (XTM) technology in order to gain edge over it competitors. They require a system that could accomplish some main functional requirements regarding their news data and information, and more importantly the system should be simple to absorb new requirements. Most of their requirements are about publishing news URL on web. The key functional requirements are captured as follows:

i. It should be possible to navigate news about particular category. For instance, Foreign scene, Sports, etc or even more specific category like in foreign scene subcategories might be Africa, Europe, Nigeria etc.

ii. For any given news, the system should be able to retrieve all the linked news, and news regarding the given news's subcategory.

iii. The system should be very flexible to accept new category of data as news items.

iv. For any given categories of news, the system should be able to report automatically the status of its resources.

4.3 System Implementation

The goal of this section is to implement a user's friendly system that would interact with the XTM-based ontology and fulfill nearly all the functional requirements. A framework called Ontopia Navigator Framework (ONF) makes the implementation possible. ONF is based on the J2EE (Java 2, Enterprise Edition) platform, which runs in any compliant J2EE application server, (for example, Apache-Tomcat) supports JSP (JavaServer Pages) 1.1 or higher. The web pages written using the navigator framework are written as JSP. JSP is a Java technology that lets you write web pages as HTML with embedded Java code.

The framework essentially consists of a set of tag libraries (tolog and template tag libraries) that make it possible to write JSP pages. This makes the pages easy to read and maintain; the tags being something like an XSLT for topic maps.

The framework is an open source and of course its design features that make it extensible to the application (ontopia) which we used to design the ontology make it a good choice. Above all, the

framework's support for Java and its features enhances XTM maintenance which is in line with the objective of the research work. That is, resolving dead inks.

4.3.1 Index page

Figure 4.1 present the index page of the News.net which is the starting page of the news domain. Users can access and navigate any categories of the news available. For example, Information Technology News, Education News, Sports News and so on.



Figure 4.1: Index Page of the NEWS.net Application

4.3.2 Navigating News categories

When user decided to access and navigates any categories of the news, the displayed result regarding to the news is directly from the XTM-based ontology which is made possible by the framework. Essentially, the result consist of the information resources (that is, both internal and external occurrences), the basename and sub categories of the news if there exist (that is,

association). Hence, the user can now access the available information resources specifically, the URLs for the requested news.

Figure 4.2a present the result of accessing Sport News and 4.2b further present navigating the Sports News sub categories (Football).



Figure 4.2 (a): Sports News Page

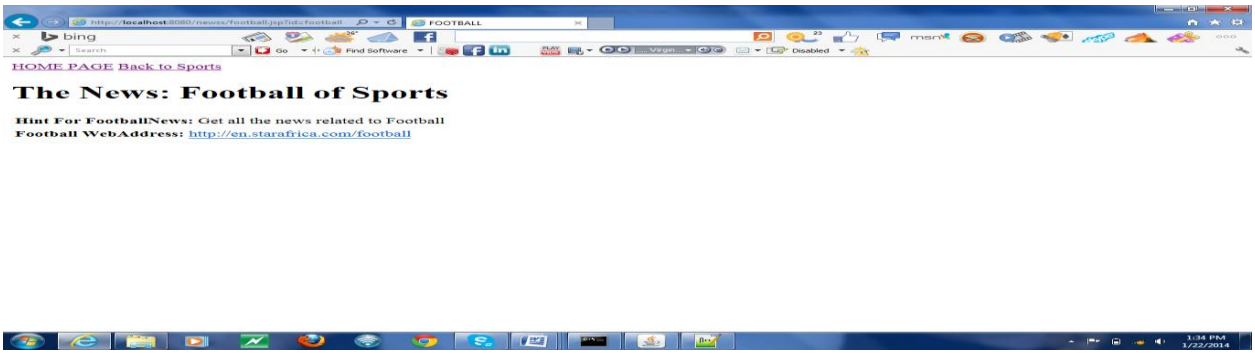


Figure 4.2 (b): Sports News sub-category

Figure 4.3 shows the user accessing the detail news of the sports via the first URL on the page.



Figure 4.3: Page for online Sports News

4.3.3 Resolving dead links

Dead links occur when URLs of a topic is wrongly typed by author of ontology or not longer exist in the internet at the time ontology is designed. In most existing XTM applications if not all, the system cannot report precisely that URLs or links are dead.

Thus, developing application with XTM and taking the advantage of the ONF, the system can out-rightly report any incident of dead links without allowing the user to do much of the work by continuous searching which ends with no expected results; and the issue of time wastage is thereby addressed.

4.4 Results and Discussions

Unlike existing XTM-based systems, the proposed NEWS.net has the capability to identify dead links associated with XTM-based ontology. To this end, Table 4.1 presents and discusses the results of some existing XTM-based systems in comparison with the proposed system.

Table 4.1a: Comparative Results of Existing Systems and the proposed NEWS.net system

	Bioinformatic XTM-based System	Oil spill XTM-based System	Proposed System-NEWS.net
Maintenance	**No**	**No**	**Yes**

Table 4.1b: Comparative Results of Existing Systems and the proposed NEWS.net system in terms of Dead Links

	Output without using Maintenance			**Output using Maintenance**		
	Dead links	Total Links		Dead links	Connecting Links	Total Links
Bioinformatic XTM-based System	0	56		54	2	56
Oil spill XTM-based System	0	10		2	8	10
Proposed NEWS.net (topic-infotech)	---	---		2	3	5

From Table4.1a, the existing XTM-based system just as noted during the review of related study, does not include any functionality (maintenance program) for XTM developed. Since the XTM are manually created and information resources that users access cannot be said to be stable at all time thus, the possibility of dead links is always very high. And that is the essence of maintenance program-to identifies dead links. For example; in Table4.1b, the existing Bioinformatic XTM-based system has fifty six (56) links and none is identify to be dead until

user discovers it while accessing the system. However, same system is executed with the maintenance module of the proposed NEWS.net system, it was discovered that only two (2) links are readily available for use while fifty four (54) are dead.

To this end, the proposed system has greatly mitigated the time taken by user who wants to navigate and access information regarding bioinformatics. The same mechanism was also applied to the other existing (oil spill) system in Table4.1b. Since the proposed system has the maintenance module incorporated, the output without using maintenance would of course automatically be nil while that of output with maintenance would definitely generate results-as in the case of a topic of the system called infotech which has a total of five (5) links but instantly identified user that two (2) are dead, thereby left with three (3) other links to use. This saves the user's time from accessing the two dead links.

Figure 4.4 shows the result of navigating Infotech News of the proposed application. The system instantly reports the dead links. Figure 4.5 represents the dead links output of Bio informatics XTM-based system and Figure 4.6 equally represents the dead links output of Oil Spill XTM-based system.

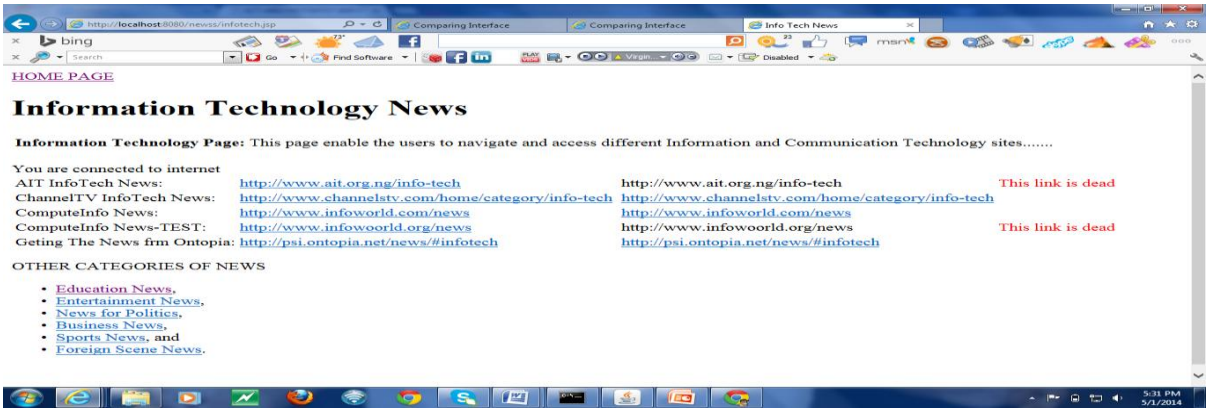


Figure 4.4: Result of (Dead Links) URLs in the infotech News of NEWS.net System

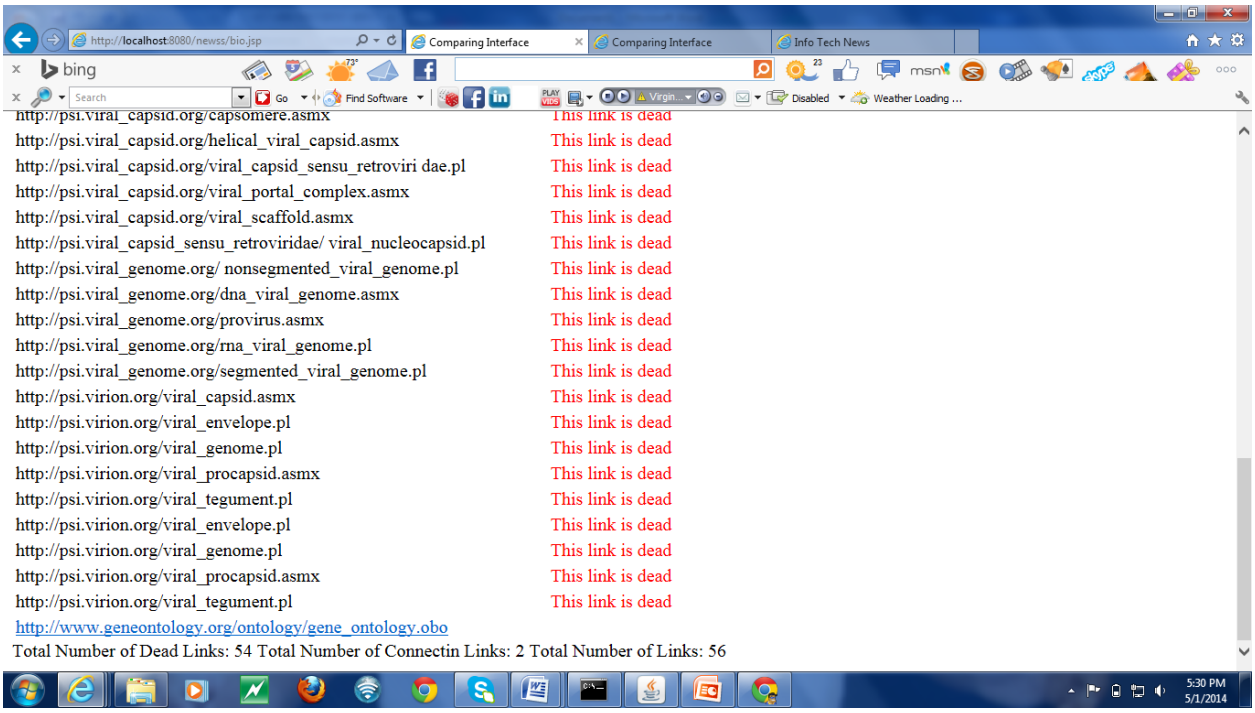


Figure 4.5: Screenshot of Bio informatics XTM-based System's Dead Links Output

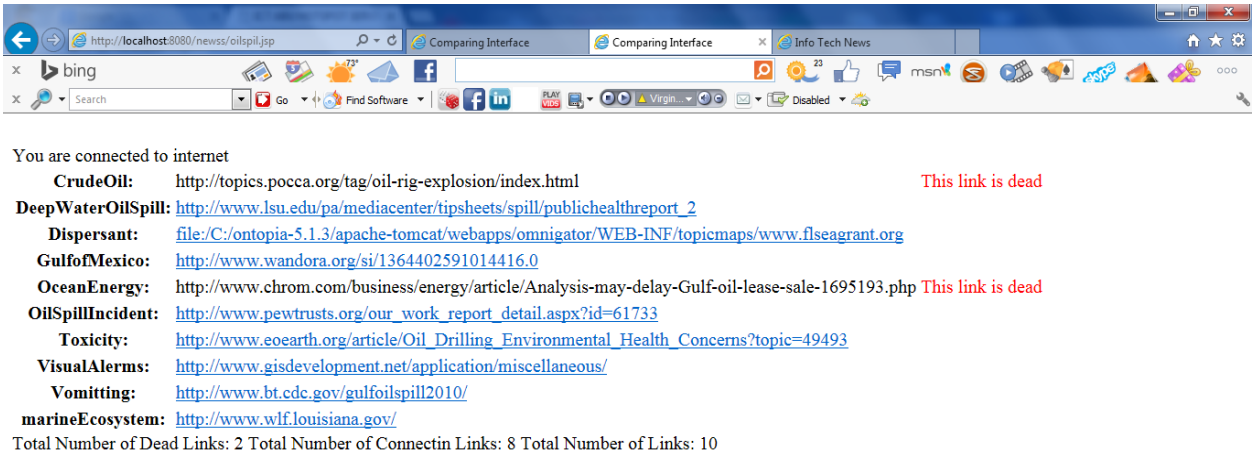


Figure 4.6: Screenshot of Oil Spill XTM-based System's Dead Links Output

4.5 System Testing

In every system development, web application inclusive; testing is very essential. The process of performing a variety of tests on a system to explore functionality or to identify problem is called system testing. System Testing is generally required before and after a system is put in place. There are different testing methodologies to test web applications according to Palani (2011). But for this research work, only four testing methods have been conducted to ensure that the News.net is working as expected. They are:

4.5.1 Usability testing

This stage of testing which was carried out in this thesis involves, ensure proper navigation between web pages; having a site map; use appropriate colour combination and best practices for instance, colour coding conventions specify that to stop a process, red is used and green to start one; amongst others.

4.5.2 User acceptance testing

The objective of this testing is ensure the application meets the functional requirements or so to say, expectations of the user. A prime tip of this testing is to also ensure browsers compatibility. For this research work, we have carried out the testing in three different browsers which include: Firefox, Google Chrome, and Internet Explorer.

Alpha and Beta testing are the two types of user acceptance testing: alpha testing simply mean a test carried out by developers in a development environment while, beta testing is carried out by end users in a deployment or client environment. For this research work, the former has been carried out severally and successfully.

4.5.3 Security testing

Security testing can be static or dynamic. For this research work, it only covers dynamic testing. Dynamic testing involves running the application to know if the response is as expected for the associated request. Should a system cannot locate URL in a server, because the URL is not existing and the system is connected to internet, checking the computer network connection should not be an option to the browser. Thus, dynamic testing is very similar to black box testing.

4.5.4 Functional testing

Functional testing in web application ranges from testing weather links are working to checking weather changes made by developers are reflected in the database or the XTM document (ontology). Some of the functional tests for web applications include: database/ontology testing, configuration testing etc. In this research work, configuration test was conducted to ensure the JSP codes in the framework (ONF) can directly access the XTM-based ontology.

4.6 Results of the RSS Model

In this section we present and discuss the results of the Ranked Semantic Similarity (RSS) Measures we examined on some topics of the XTM-based news ontology; and the algorithm in figure 4.7 presents and discuss the results of the system implemented.

The model uses values ranges from 0 to 1 with a step size of 0.1, and cut-off value of 0.5 for the set of new news item to experiment.

The User Profile is U = {Sports, ForeignScene, Business}

The weights W (number of news read) for the corresponding user profile topics are:

W = (4, 3, 2).

The sets of related topics for each topic in the profile are as follows:

r(Sports) = {SportingsNews, BSkyB},

r(ForeignScene) = { africaNews},

r(Business) = {africaNews}.

The set of related topics to the user profile topics is as follows (Equation 3.2):

R = r(Sports) ∪ r(ForeignScene) ∪ r(Business)

= {SportingsNews, BSkyB, africaNews}

The extended user profile is (Equation 3.3):

U_X = {Sports, ForeignScene, Business, SportingsNews, BSkyB, africaNews}

The rank matrix which is called Ranked News is shown in Table 4.2 following the model of Equation 3.5.

Table 4.2: Ranked News Matrix

	Sports	ForeignScene	Business	SportingsNews	BSkyB	africaNews
Sports	4.0	-0.4	-0.4	2.0	2.0	-0.4
ForeignScene	-0.3	3.0	-0.3	-0.3	-0.3	1.5
Business	-0.2	-0.2	2.0	-0.2	-0.2	1.0
Rank	**3.5**	**2.4**	**1.3**	**1.5**	**1.5**	**2.1**

The normalized rank vector V_U is (Equation 3.7):

V_U = (1.0, 0.5, 0.0, 0.091, 0.091, 0.364).

The following set of two news items comprises of both related and unrelated topics of the XTM-based news ontology were examined with the model.

I_1 = {SportingNews, africaNews, gongolaNews}

I_2 = {Sports, africaNews}

The vector representations of these two items following the model b_i of Equation 3.10 are:

V_{I1} = (0.091, 0.364, 0.0)

V_{I2} = (1.0, 0.364)

Thus, the ranked semantic similarity of the two news items to the extended user profile are computed as follows (Equation 3.11):

$$\text{MaSemSi}I_1 = \frac{0.091+0.364+0.0}{1.0+0.5+0.0+0.091+0.091+0.364} = 0.222 \cong 0.2$$

$$\text{MaSemSi}I_2 = \frac{1.0+0.364}{1.0+0.5+0.0+0.091+0.091+0.364} = 0.667 \cong 0.7$$

Considering the cut-off value of 0.5, we can only consider the second news item (I_2). That is, I_2 can be conveniently compared with the extended user profile. This is graphically represented in Figure 4.7.

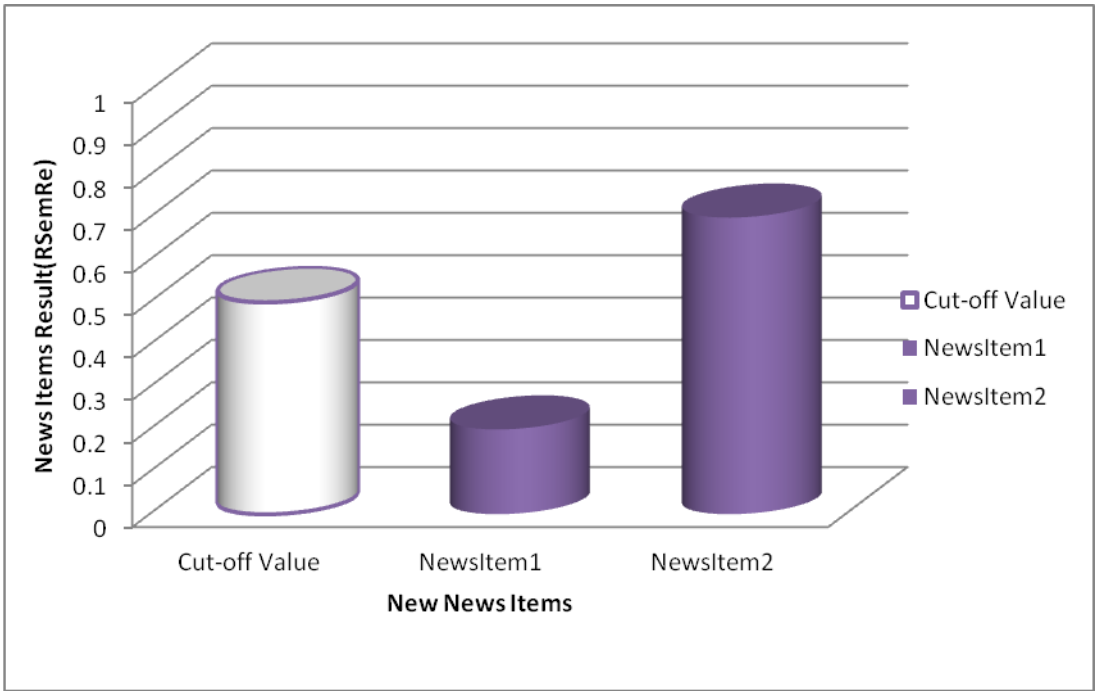


Figure 4.7: Results Displayed for two News Items.

Clearly, since the model's user interestingness for cut-off value is set for 0.5, the proposed system would only present interesting and reliable news to users (for example: NewsItem2) and eliminate uninteresting news (as in the case of NewsItem1) thereby, proffering solution to the issue of high recall, low precision related to XTM. Thus, the results of the model can be described to be similar in functionality to the maintenance module of the proposed system.

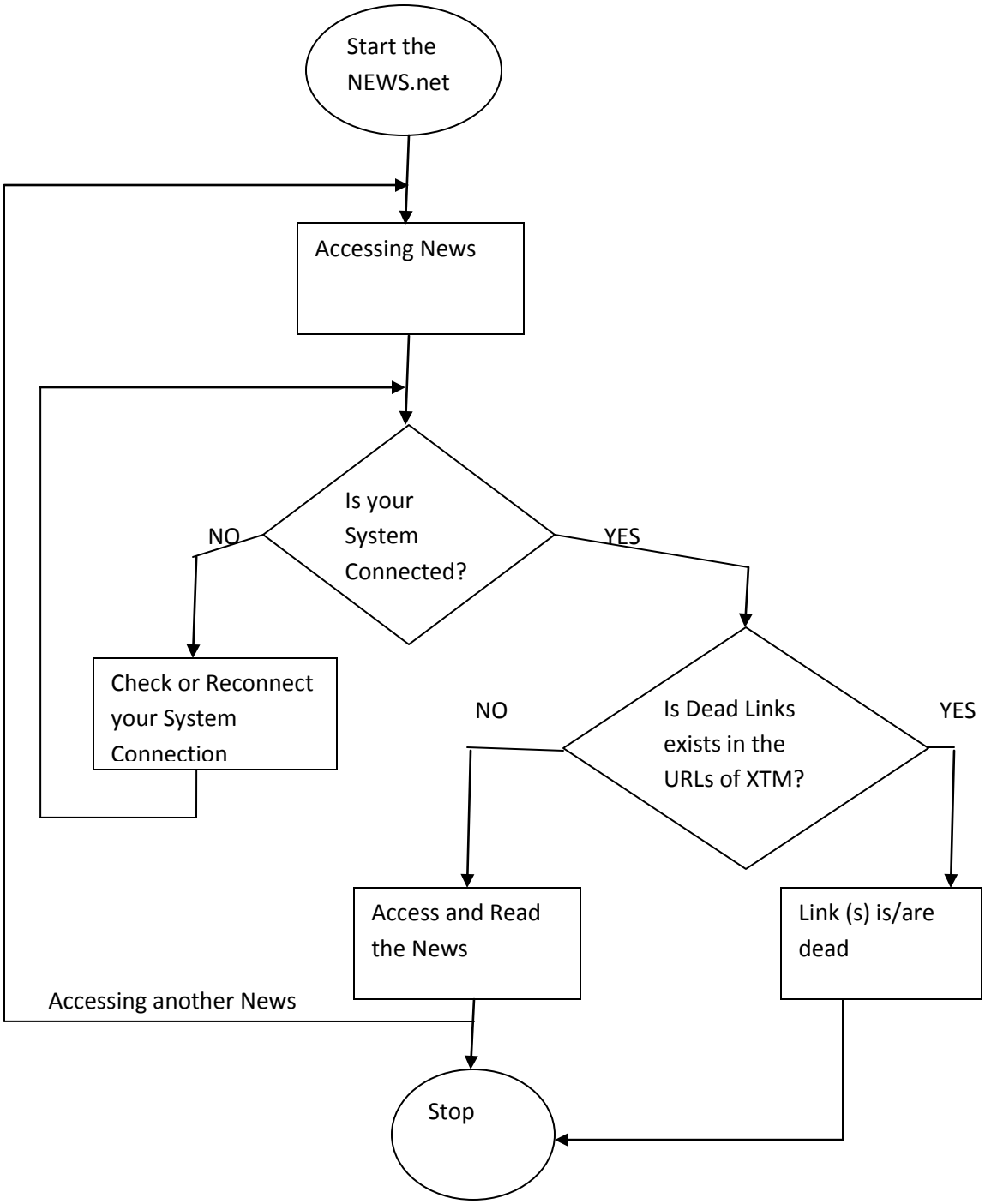


Figure 4.8: Algorithm that Discuss the Result of the Implementation

Before user access the NEWS.net: http://localhost:8080/newss/ via any browser, it is expected that the server (Apache-tomcat) is on. Immediately user selected news to access, the application

first establishes the user's computer network connection status to check and display if the user's computer is connected to internet or not. If the system is connected to internet, then the application through the ONF further check and report if there is/are URL(s) that is/are dead.

The server which acts as the database display the result of the selected news on the browser which includes all its URLs to read the news from XTM-based ontology and highlighted the links that are dead, if any. Hence, user would in turn navigate any of the news URLs that are not dead to read the news.

It is noteworthy to mention that the application's strength to check the network connectivity of the user's system and detect any dead links is attributed to the ONF we extended to the XTM document. To be more precise, the ONF in this context is the JSP codes we embedded in the news categories .jsp files which is shown in the snapshot at Figure 4.9a and 4.9b.

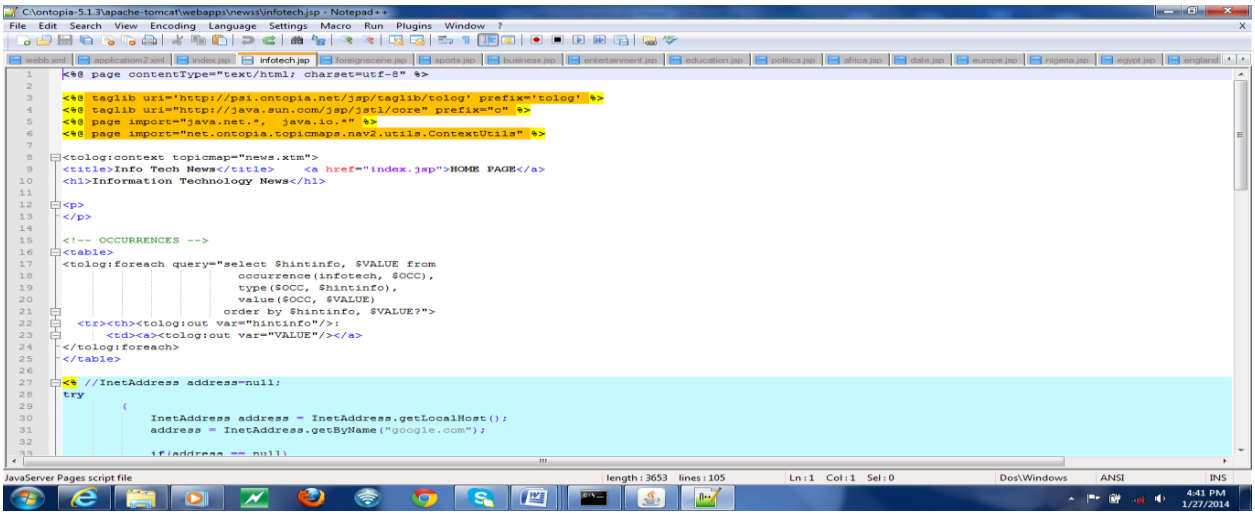


Figure 4.9a: JSP codes to check Computer Network Connectivity.

The yellow colour lines of code are the set of the framework tag library and that of java to import classes. The codes on the grey green colour are to check if the user's system is connected to internet.

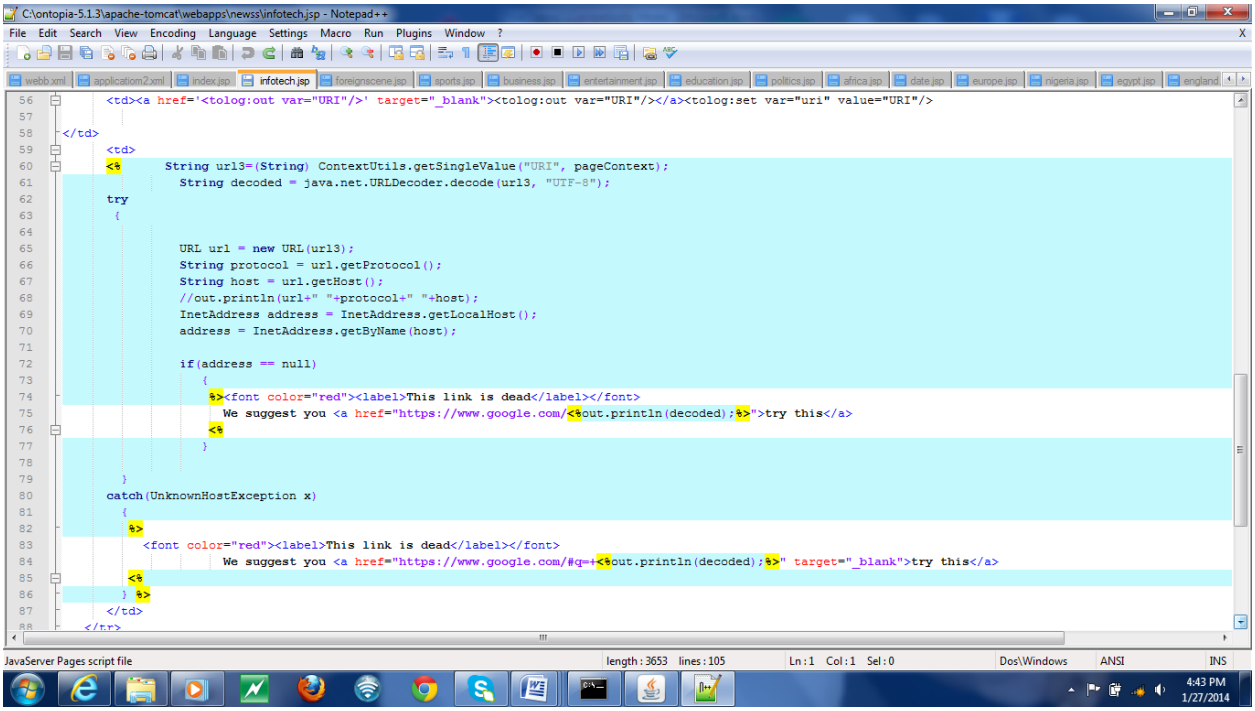


Figure 4.9b: JSP codes to check Dead links

The JSP codes that check if dead links exist in the XTM-based ontology are highlighted in green grey colour.

Figure 4.10 is the command to startup the Tomcat Server. At the command prompt, we first change the system's directory to the location of the server that is, C:/ ontopia5.1.3/apache-tomcat/bin> then follow by startup.bat command. The .bat extension is for windows-based systems.

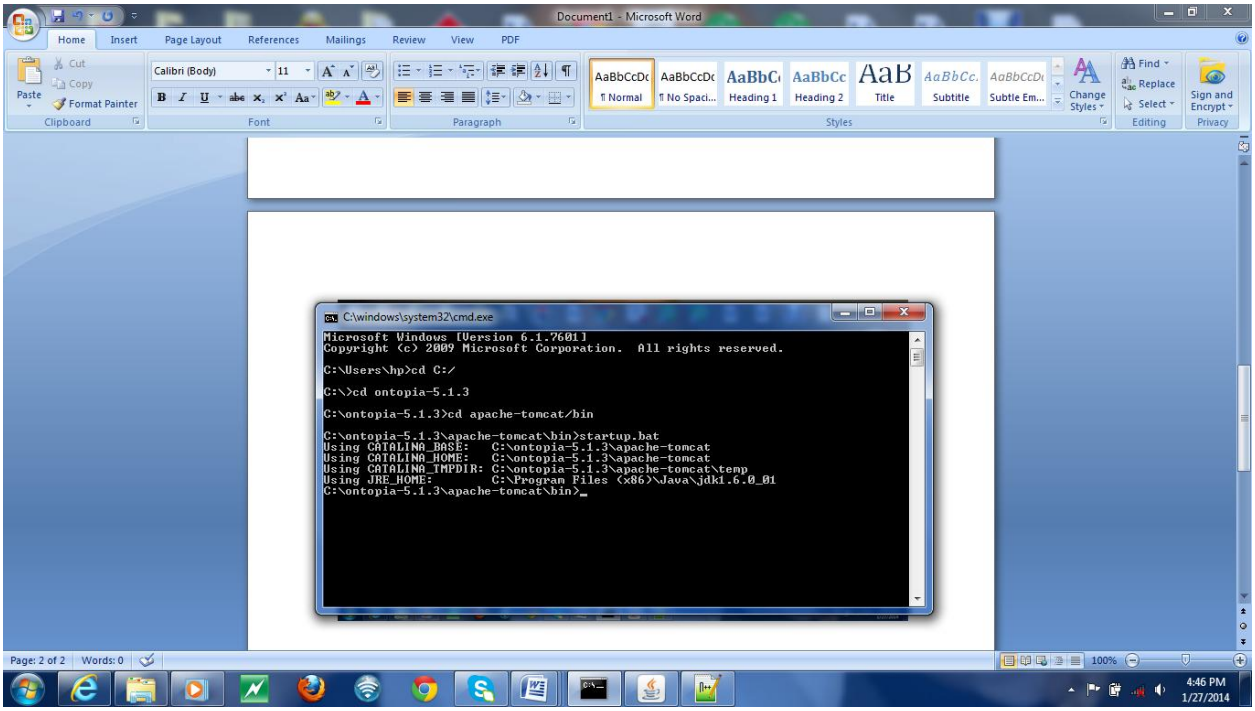


Figure 4.10: Command to startup Server

CHAPTER FIVE: CONCLUSION AND RECOMMENDATION

5.1 Conclusion

This research work focuses on how to navigate and access news domain by developing and maintaining XTM. News application contains a large volume of information and complex resources. To this end, XTM provides a good model for structuring information and its resources. Also it is considered as a standard for the representation and interchange of knowledge. For instance, a news type has its own information which includes information resources or their corresponding sub-types of the news. Since the news items have been structured with the aid of XTM, users would no longer spend time to access an application that will retrieve a chunk of irrelevant information. Thus, in the proposed XTM-based application, news items are structured along side with their relationship and their related information resources.

Considering the objective of this research work, which is to resolve dead links that may arise from unmaintained XTM which we developed and implemented on other existing systems, we were able to present an XTM-based web application stored as file in the tomcat server which acts as database. Dead links were resolved by interfacing the XTM-based ontology with the J2EE-based framework (ONF). Besides, before resolving dead links, the application can verify if user's computer is connected to internet.

Hence, this research work has shown that using XTM (on ontopia application software) and tolog-XSLT with JSP (which form the set of ONF library) as the source code or site map of a

web application, offers a reliable and convenient platform of equally maintaining web applications with large and complex information resources.

5.2 Recommendations

The significance of NEWS.net to resolves dead links which may arise from unmaintained XTM cannot be overstressed, thus it is recommended for XTM, LTM or RDF developers in any area of knowledge and to a great extent, the proposed system can identify dead links by itself.

Since the NEWS.net is an XTM-based application and XTM is a topic map defined in XML syntax of W3C, then XTM supports structuring organization's documents or even make an organization's application talk to other applications. Also, in line with the versatility feature of the Ontopia Software Application, XTM document can be exported to RDF or LTM format and vice versa thereby, making it easier for RDF/S developers to use.

5.3 Future Work

Just like in any other research work and with no exceptional to this thesis; we encountered challenges which we describe in this section as possible areas of improvement and research. The graphical visualization and navigation of the XTM structure (like Vizigator) can be presented on the NEWS.net.

XTM's sensitive element (e.g., occurrence) is largely created by domain expert and consumes a lot of efforts. Future research should focus on automatically creating and maintaining such element of the knowledge base by interfacing the knowledge base to support a semantic search

engine. Besides the improvement of the knowledge base, the ranked semantic similarity model can be improved as well.

REFERENCES

Antoniou, G. and Harmelen, F.V. (2008). A Semantic Web Primer. The MIT Press Cambridge, Massachusetts London, England.

Anyi, K.W.U. (2012). Topic Maps: A Bibliometric Study. Unpublished M.Sc. Thesis. Oslo and Akershus University College of Applied Science.

Biezunski, M. and Newcomb, S. R., (2001). XML Topic Maps: Finding Aids for the Web. In: Leu, P. (Ed) *Siemens Corporate Research*. IEEE Multimedia, PP. 104 – 108.

Chen, X., Hou, X., and Li, N. (2011). Teaching System Modeling Based on Topic Maps. In: Lin, S. and Hoang, X. (Eds) *CSEE*. Springer-Verlag Berlin, Heidelberg. PP. 197–204.

Corcho, O., Fernandez-Lopez, M., and Gomez-Perez, A. (2002). Methodologies, tools and languages for building ontologies. Where is there meeting point? *Elsevier Science* 46(2003): 41 – 64.

Dicheva, D. and Dichev, C. (2001). Deriving Context Specific Information on the Web. *World Conference on the WWW and Internet (WebNet'2001),* Orlando, Florida, pp. 296-301.

Dicheva, D. and Dichev, C. (2006). TM4L: Creating and browsing educational topic maps. *British Journal of Educational Technology*. 37(3) 391 - 404

Dunaway, D.J., and Wu, Y. (2013). Creating a large topic maps by integrating Wandora and Ontopia. *Library Hi Tech/Emerald* 31(1) 64-75

Ellouze, N., Ahmed, M.B. and Metais, E. (2008). Overview of Topic Maps Construction Approaches. In: IEEE Computer Society of the 22nd International Conference, on

Advanced Information Networking and Applications – Workshops, pp 1642 – 1647.

Ellouze, N., Lammari, N., Metais, E. and Ahmed, M.B. (2009). Usage-Oriented Topic Maps Building Approach. In: Sartori, F., Sicilia, M.A., and Manouselis, N. (Eds) Springer-Verlag Berlin Heidelberg, PP. 13 – 23.

Garshol L. M. and Moore G. (2008). Topic Maps – Data Model, Retrieved March, 30, 2012, from http://www.isotopicmaps.org/sam/sam-model/

Garshol, L M. (2004). Metadata? Thesauri? Taxonomies? Topic maps! Making sense of it all. *Journal of Information Science*, *30*(4), 378–391.

Gregoras, D. (2005). Using XTM for navigating UDDI. Unpublished M.Sc Thesis. Department of Mathematics and Computer Science University of Osnabrück, Germany and Department of Computer Science, University of Twente, The Netherlands.

Gruber, T.R. (1993). A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*, 5(2): 199-220.

Husáková, M. and Olševičová, K. (2008). Creating Web Presentation for Observatory and Planetarium with Topic Maps. In: Maicher, L.; Garshol, L. M. (Eds) Subject-centric computing. Topic Maps Research and Applications, PP. 237 – 246. Fourth International Conference held at Liezpig, Germany, 16 – 17 October.

Ijntema, W., Goossen, F., Frasincar, F., and Hogenboom, F., (2010). Ontology-Based News Recommendation. In: Proceedings of the ACM, EDBT held at Lausanne, Switzerland, 22–26 March.

Librelloto, G.R., Ramalho, J.C. and Henriques, P.R., (2004) TM Builder: An Ontology Builder based on XML Topic Maps. *CLEI Electronic Journal*, 7(2).

Noy, N., and McGuiness D. (2000). Ontology Development 101: A Guide to Creating Your First

Ontology. Retrieved April, 22, 2013, from http://protege.stanford.edu/publications/ontology_development/ontology101.pdf

Ontopia, (2010). Retrieved September, 27, 2013, from http://www.ontopia.net/omnigator/docs/navigator/userguide.html

Palani, G.S., (2011). Summary of web application testing methodologies and tools. Retrieved July 22, 2013, from developerWorks IBM database.

Park, J. and Hunting, S. (2002). XML Topic Maps: Creating and Using Topic Maps for the Web. Addison Wesley, Boston, U.S.A, PP. 79 – 95.

Pepper, S. (2006). The case for Published Subjects. Retrieved September, 27, 2013, from http://www.ontopia.net/topicmaps/materials/The_Case_for_Published_Subjects.pdf

Pepper, S. (2000). The TAO of Topic Maps. Finding the Way in the Age of Infogut. Retrieved March, 5, 2012 from http://www.ontopia.net/topicmaps/materials/tao.pdf

Pepper, S. and Moore, G. (2001). XML topic maps (XTM) 1.0. Retrieved July, 14, 2012, from http://www.topicmaps.org/xtm/1.0/

Pepper S. and Garshol L. M. (2007). The XML Papers - Lessons on Applying Topic Maps, www.ontopia.net/topicmaps/materials/xmlconf.html,

Rubbani, H.H. (2007). Semantic Web Solutions. Unpublished M.Sc Thesis. IT University of Copenhagen, Denmark.

Yi, M. (2008). Information Organization and Retrieval Using a Topic Maps-Based Ontology: Result of a Task-Based Evaluation. *Journal of the American Society for Information Science and Technology* 15(12): 1898 – 1911.

APPENDIX A: ONTOLOGY FILE

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<topicMap id="newsdom"
        xmlns="http://www.topicmaps.org/xtm/1.0/"
        xmlns:xlink="http://www.w3.org/1999/xlink">
<!--Reification of Topic Maps itself to what what we called News Domain Topic Maps-->
<topic id="newsdom-topic">
  <subjectIdentity>
    <subjectIndicatorRef xlink:href="#newsdom"/>
  </subjectIdentity>
  <baseName>
    <baseNameString>News Domain Topic Map</baseNameString>
  </baseName>
</topic>
<!-- An individual topic called Information Technology-->
<topic id="infotech">
    <instanceOf>
    <topicRef xlink:href="#news"/>
    </instanceOf>
    <baseName>
      <baseNameString>Information Technology</baseNameString>
    </baseName>
            <occurrence>
      <instanceOf>
                <topicRef xlink:href="#hintinfo"/>
              </instanceOf>
                        <resourceData>This page enable the users to navigate and access different
                                Information and Communication Technology sites.....
                        </resourceData>
  </occurrence>
  <occurrence>
      <instanceOf>
        <topicRef xlink:href="#webinfo"/>
      </instanceOf>
      <resourceRef xlink:href="http://www.channelstv.com/home/category/info-tech"/>
  </occurrence>
  <occurrence>
      <instanceOf>
        <topicRef xlink:href="#webinfo2"/>
      </instanceOf>
      <resourceRef xlink:href="http://www.ait.org.ng/info-tech"/>
  </occurrence>
  <occurrence>
      <instanceOf>
        <topicRef xlink:href="#webinfo3"/>
      </instanceOf>
      <resourceRef xlink:href="http://www.infoworld.com/news"/>
  </occurrence>
  <occurrence>
      <instanceOf>
        <topicRef xlink:href="#webinfo4"/>
      </instanceOf>
      <resourceRef xlink:href="http://www.infowoorld.org/news"/>
```

```
 </occurrence>
 <occurrence>
    <instanceOf>
      <topicRef xlink:href="#webinfo5"/>
    </instanceOf>
    <resourceRef xlink:href="http://psi.ontopia.net/news/#infotech"/>
 </occurrence>
</topic>
<topic id="hintinfo">
  <baseName>
          <baseNameString>Information Technology Page</baseNameString>
        </baseName>
</topic>
<topic id="webinfo">
  <baseName>
          <baseNameString>ChannelTV InfoTech News</baseNameString>
        </baseName>
</topic>
<topic id="webinfo2">
  <baseName> <baseNameString>AIT InfoTech News</baseNameString>
        </baseName>
</topic>
<topic id="webinfo3">
  <baseName>
          <baseNameString>ComputeInfo News</baseNameString></baseName>
</topic>
<topic id="webinfo4">
  <baseName>
          <baseNameString>ComputeInfo News-TEST</baseNameString>
        </baseName>
</topic>
<topic id="webinfo5">
  <baseName>
          <baseNameString>Geting The News frm Ontopia</baseNameString>
        </baseName>
</topic>
<topic id="news">
  <baseName>
    <baseNameString>News</baseNameString>
  </baseName>
</topic>
<!--TOPIC EDUCATION OF TYPE NEWS-->
<topic id="education">
  <instanceOf>
  <topicRef xlink:href="#news"/>
  </instanceOf>
  <baseName>
    <baseNameString>Education</baseNameString>
  </baseName>
        <occurrence>
    <instanceOf>
          <topicRef xlink:href="#hintedu"/>
        </instanceOf>
              <resourceData>This page enable the users to navigate and access different
              educational sites. For example: Univerties sites, the Nigeria's Universities
```

```
                regulatory commission and some other vital sites as information resources.....
                </resourceData>
  </occurrence>
  <occurrence>
     <instanceOf>
       <topicRef xlink:href="#websitedu1"/>
     </instanceOf>
     <resourceRef xlink:href="http://www.abu.edu.ng"/>
  </occurrence>
  <occurrence>
     <instanceOf>
       <topicRef xlink:href="#websitedu2"/>
     </instanceOf>
     <resourceRef xlink:href="http://www.nucc.edu.ng"/>
  </occurrence>
  <occurrence>
     <instanceOf>
       <topicRef xlink:href="#websitedu2"/>
     </instanceOf>
     <resourceRef xlink:href="http://www.nbte.gov.ng"/>
  </occurrence>
  <occurrence>
     <instanceOf>
       <topicRef xlink:href="#websitedu4"/>
     </instanceOf>
     <resourceRef xlink:href="http://www.scholarship-position.com"/>
  </occurrence>
  <occurrence>
     <instanceOf>
       <topicRef xlink:href="#websitedu5"/>
     </instanceOf>
     <resourceRef xlink:href="http://www.universityofebira.edu.ng"/>
  </occurrence>
  <occurrence>
     <instanceOf>
       <topicRef xlink:href="#websitedu6"/>
     </instanceOf>
     <resourceRef xlink:href="http://psi.ontopia.net/news/#education"/>
  </occurrence>
</topic>
<topic id="hintedu">
   <baseName>
            <baseNameString>Brief Description</baseNameString>
          </baseName>
</topic>
<topic id="websitedu1">
   <baseName>
            <baseNameString>ABU Portal</baseNameString>
          </baseName>
</topic>
<topic id="websitedu2">
   <baseName>
            <baseNameString>Regulatory Bodies</baseNameString>
          </baseName>
</topic>
```

```
<topic id="websitedu4">
   <baseName>
            <baseNameString>Scholarship Links</baseNameString>
          </baseName>
</topic>
<topic id="websitedu5">
   <baseName>
            <baseNameString>Portal of Ebira University</baseNameString>
          </baseName>
</topic>
<topic id="websitedu6">
   <baseName>
            <baseNameString>Geting the EducationNews from Ontopia</baseNameString>
          </baseName>
</topic>
<!--TOPIC ENTERTAINMENT OF TYPE NEWS-->
<topic id="entertainment">
  <instanceOf>
  <topicRef xlink:href="#news"/>
  </instanceOf>
  <baseName>
    <baseNameString>Entertainment</baseNameString>
 </baseName>
   <occurrence>
     <instanceOf>
              <topicRef xlink:href="#hintenta"/>
            </instanceOf>
                    <resourceData>Waoh!!!!!ENTERTAINMENT!!!Life is too short, and we are living in a  fast paced world which is too demanding hence it is very difficult for us to stay calm, cool and friendly all the time while dealing with our.....
                    </resourceData>
 </occurrence>
 <occurrence>
     <instanceOf>
       <topicRef xlink:href="#websiteenta1"/>
     </instanceOf>
     <resourceRef xlink:href="http://www.channelstv.com/home/category/entertainment/"/>
 </occurrence>
 <occurrence>
     <instanceOf>
       <topicRef xlink:href="#websiteenta2"/>
     </instanceOf>
     <resourceRef xlink:href="http://www.starafrica.com/music"/>
 </occurrence>
 <occurrence>
     <instanceOf>
       <topicRef xlink:href="#websiteenta3"/>
     </instanceOf>
     <resourceRef xlink:href="http://www.huffingtonpost.com/comedy"/>
 </occurrence>
 <occurrence>
     <instanceOf>
       <topicRef xlink:href="#websiteenta4"/>
     </instanceOf>
     <resourceRef xlink:href="http://psi.ontopia.net/news/#politics"/>
```

```
  </occurrence>
</topic>
<topic id="hintenta">
   <baseName>
            <baseNameString>Your EntertainmentWorld</baseNameString>
          </baseName>
</topic>
<topic id="websiteenta1">
   <baseName>
            <baseNameString>AllEntertainment Link</baseNameString>
          </baseName>
</topic>
<topic id="websiteenta2">
   <baseName>
            <baseNameString>MusicEntertainment Link</baseNameString>
          </baseName>
</topic>
<topic id="websiteenta3">
   <baseName>
            <baseNameString>ComedyEntertainment Link</baseNameString>
          </baseName>
</topic>
<topic id="websiteenta4">
   <baseName>
            <baseNameString>Better....Semantically</baseNameString>
          </baseName>
</topic>
<association>
  <instanceOf>
    <topicRef xlink:href="#enter-news"/>
  </instanceOf>
  <member>
    <roleSpec><topicRef xlink:href="#enter-one"/></roleSpec>
    <topicRef xlink:href="#allenter"/>
  </member>
  <member>
    <roleSpec><topicRef xlink:href="#enter-two"/></roleSpec>
    <topicRef xlink:href="#ontoenter"/>
  </member>
  <member>
    <roleSpec><topicRef xlink:href="#enter-three"/></roleSpec>
    <topicRef xlink:href="#comedy"/>
  </member>
  <member>
    <roleSpec><topicRef xlink:href="#enter-four"/></roleSpec>
    <topicRef xlink:href="#music"/>
  </member>
  <member>
    <roleSpec><topicRef xlink:href="#enterNewsCastedBy"/></roleSpec>
    <topicRef xlink:href="#entertainment"/>
  </member>
</association>
<topic id="enter-news">
  <baseName>
     <baseNameString>EntertainmentNews Casted By</baseNameString>
```

```
  </baseName>
  <baseName>
    <scope><topicRef xlink:href="#allenter"/></scope>
    <baseNameString>casted by </baseNameString>
  </baseName>
  <baseName>
    <scope><topicRef xlink:href="#ontoenter"/></scope>
    <baseNameString>casted by</baseNameString>
  </baseName>
  <baseName>
    <scope><topicRef xlink:href="#comedy"/></scope>
    <baseNameString>casted by</baseNameString>
  </baseName>
  <baseName>
    <scope><topicRef xlink:href="#music"/></scope>
    <baseNameString>casted by</baseNameString>
  </baseName>
</topic>
<topic id="enter-one">
  <baseName>
    <baseNameString>Entertainment One</baseNameString>
  </baseName>
</topic>
<topic id="enter-two">
  <baseName>
    <baseNameString>Entertainment Two</baseNameString> </baseName>
</topic>
<topic id="enter-three">
  <baseName>
    <baseNameString>Entertainment Three</baseNameString>
  </baseName>
</topic>
<topic id="enter-four">
  <baseName>
    <baseNameString>Entertainment Four</baseNameString>
  </baseName>
</topic>
<topic id="enterNewsCastedBy">
  <baseName>
    <baseNameString>EntertainmentNews CastedBy</baseNameString>
  </baseName>
</topic>
<topic id="allenter">
  <baseName>
    <baseNameString>AllEntertainment</baseNameString>
  </baseName>
</topic>
<topic id="ontoenter">
  <baseName>
    <baseNameString>PSIOntoEnt</baseNameString>
  </baseName>
</topic>
<topic id="comedy">
  <baseName>
    <baseNameString>Comedy</baseNameString>
```

```
  </baseName>
</topic>
<topic id="music">
  <baseName>
    <baseNameString>Music</baseNameString>
  </baseName>
</topic>
```

APPENDIX B: SAMPLE PROGRAM FILE

```
<%@ taglib uri='http://psi.ontopia.net/jsp/taglib/tolog' prefix='tolog' %>
<%@ taglib uri="http://java.sun.com/jsp/jstl/core" prefix="c" %>
<%@ page import="java.net.*,  java.io.*" %>
<%@ page import="net.ontopia.topicmaps.nav2.utils.ContextUtils" %>
<td>
        <%      String url3=(String) ContextUtils.getSingleValue("URI", pageContext);
        String decoded = java.net.URLDecoder.decode(url3, "UTF-8");
    try
     {
                 URL url = new URL(url3);
                String protocol = url.getProtocol();
                 String host = url.getHost();
                 //out.println(url+" "+protocol+" "+host);
                InetAddress address = InetAddress.getLocalHost();
                address = InetAddress.getByName(host);
     if(address == null)
        {
          %><font color="red"><label>This link is dead</label></font>
         We suggest you <a href="https://www.swoogle.com/<%out.println(decoded);%>">try this</a>
          <%
          }
     }
    catch(UnknownHostException x)
     {
      %>
```

```
            <font color="red"><label>This link is dead</label></font>
              We suggest you <a href="https://www.swoogle.com/#q=+<%out.println(decoded);%>"
                    target="_blank">try this</a>
        <%
        } %>
</td>
```