AN ATTRIBUTE-VALUE PAIR (AVP) MODEL FOR CREATING AND EXPLOITING ANNOTATIONS IN ECONOMIC INTELLIGENCE

By

OKUNOYE, Olusoji Babatunde

Matriculation Number: 890808043

A thesis submitted to the School of Postgraduate Studies, University of Lagos, Akoka, Lagos, Nigeria, in partial fulfilment of the requirement for the award of the degree of Doctor of Philosophy (Ph.D.) in Computer Science.

MARCH 2014

Certification

This is to certify that the thesis titled: "AN ATTRIBUTE-VALUE PAIR (AVP) MODEL FOR CREATING AND EXPLOITING ANNOTATIONS IN ECONOMIC INTELLIGENCE" submitted to the School of Postgraduate Studies, University of Lagos for the award of the degree of Doctor of Philosophy (Ph.D.) is a record of original research carried out by OKUNOYE, Olusoji Babatunde in the Department of Computer Sciences.

Name : **UWADIA, Charles Onuwa**
Status : **Professor**
Department : **Department of Computer Sciences, University of Lagos**
Signature :
Date :

Name : **DAVID, Amos**
Status : **Professor**
Department : **Nancy 2 University, Nancy, France**
Signature :
Date :

Name : **OKUNOYE, Olusoji Babatunde**
Status : **Author**
Department : **Department of Computer Sciences, University of Lagos**
Signature :
Date :

DEDICATION

This thesis is dedicated to JEHOVAH, the God Almighty and to all persons that are diligently awaiting the second coming of Jesus Christ.

ACKNOWLEDGEMENT

My heartfelt gratitude goes to my heavenly Father, the Alpha and the Omega, who in His infinity mercies makes this thesis a reality.

Special thanks to my Supervisor Professor C. O. Uwadia for hismotivation, fatherly and intellectual guidance throughout the duration of this work. May the Almighty God continue to bless you.

My profound gratitude goes to my Supervisor at research team SITE, LORIA, Nancy, France, Professor Amos David. I really appreciate your love, encouragement and candid opinion throughout the duration of this work. You accepted me as part of your family. I will never forget those moments we shared together. Your pieces of advice still echo in my ears. Thank you very much.

My gratitude also goes to all members of the research team SITE, LORIA for their encouragement and support most especially to my friend and brother, Dr. Victor Odumuyiwa for making my stay in France a homely one. Big thanks to you. Also to Dr. William Onifade, Dr. (Mrs.) Bolanle Oladejo, Hanene, Rojo and the rest that I could not make mention of their names, I really appreciate you all.

I thank the French Embassy in Nigeria for the funding of part of this research at Equipe Site-LORIA, Nancy, France. Thank you very much.

I specially thank all members of Young Elders' Forum for your prayers and support, God who is not unrighteous to forget your labour of love, will honour you and bless you all.

To my big family at the department of Computer Sciences, university of Lagos, Professor O. Abass, Professor J. O. A. Ayeni, Professor H. O. D. Longe, Dr. E. P. Fasina, Dr. A. P. Adewole, Dr. F. A. Oladeji (my wonderful sister), Dr. B. A. Sawyerr, Dr. O. Oladipupo, Dr. N. A. Azeez,Mr. A. U. Rufai, Mrs C. Banjo, Mrs. C. Ojiako, Ms. R. Isimeto, Mrs. R. Ajetumobi, Mr. L. Ikuvwerha, Miss. T. Ogunniyi I say a big thank you to you all. Also, included in the big family are both the administrative and technical members of the department. I appreciate you all.

My sincere gratitude goes to my darling wife, Opeyemi and my lovely children, David, Damilola and Paul. I really appreciate your love, prayers, understanding, encouragement and endurance which greatly contribute to the success of this research work.

I am grateful to my parents, Mr. And Mrs. S. S. Okunoye for their prayers and support. Thank you very much. I cannot but remember the labour of love of my siblings, I love you all.

I really appreciate the Dean of School of Postgraduate Studies, Prof. L. O. Chukwu for the rolehe played for this achievement. I say thank you.

Finally, I specially thank my external Examiner, Professor Akinwale for his constructive judgment and the internal examiners, Dr. A. P. Adewole and Dr. T. Fashanu for making this pursuit a worthwhile. Thank you very much.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

This study is concerned with the design and development of an annotation model that gives users the power of expressivity as required in Economic Intelligence (EI) context. Most of the existing annotation platforms lack this power of expressivity. The study also discovered a need to develop an annotation model for creating and structuring annotations that will adequately capture the intention of the users (EI actors) in decision making process. In addition, there was a need for a technique that will allow actors to search for information based on the objective of the search.

This study introduced an annotation model called Attribute-Value Pair (AVP) for creating and storing annotations. The study developed a mechanism for exploiting stored annotations based on the context of problem, and used the AVP model to develop a search algorithm that allowed actors to search for information based on the objective of the search. The study used Resource Description Framework (RDF) for the formalism of AVP annotation model. The exploitation phase was implemented using Explore, Query, Analyze and Annotate(EQuA^{2}Te) architecture. A pattern-based algorithm called AVP search was developed that allowed actors to search for information based on the semantic of search objective. The study developed a prototype called Annotation Model and Tools for Economic Intelligence Actors (AMTEA) that used the AVP annotation model for creating, storing and exploiting annotations as well as performing search operations. Two search problems were used as scenarios to evaluate AMTEA system. Results obtained were compared with annotations made on the same set of documents by human agents. The performance evaluation shows that the new AMTEA system detected over 98% of manually made annotations by the human agents. In addition, AMTEA system was able to find new annotations that appear to be relevant. In essence, AMTEA assists human agents to discover new information that might be relevant in decision making process.

CHAPTER ONE

INTRODUCTION

1.1 Background of Study

The term annotation has almost become a household name. Its application cuts across virtually all disciplines. In computer science and information science, annotation could be defined as a note, an explanation, or any other type of external remark that can be attached to a document without necessarily being inserted into the document (Bodain *et al*., 2007). Annotation is value-added information used for contextualizing underlying document. Annotation could be in textual form, voice, ink or graphics. This work focuses on textual form of annotation.

Research work on annotation can be classified as data annotation, ink annotation, and semantic annotation. Data annotation is about annotating structured data stored in information repository. Annotation propagation is a major research challenge in this type of annotation. Some of the works done in this area include the works of Bhatnagar (2007) and Geerts (2006). Bhatnagar defined data annotations that allow users to annotate relational data at five different levels of granularity – database level, relation level, column level, tuple level and cell level. Geerts introduced annotation data model for manipulating and querying both data and annotation using the concept of block of colours to represent an annotated set of values. Li *et al*. (2008) developed view-based annotation propagation scheme called ViP framework.

Research challenges in Ink annotation centres on capturing, recognising, and anchoring free-hand annotations on documents. The works of Wang and Raghupathy (2007), Yang *et al*. (2004), Bargeron*et al*. (2001) are some of the works done o n this type of annotation.

Semantic Annotation (SA) is the process of adding formal semantics such as metadata to web content for the purpose of accessing and managing web content efficiently by both human and machine agents. SA is made popular by the advent of Semantic Web (SW). One of the goals of SW according to World Wide Web Consortium (W3C) is to make web content accessible and processable by both human and machine agents. This could be achieved by converting the web content to web of linked data. SA largely depends on the nature of the web content being

considered – Named Entities or word sense. Named Entities (NEs) in Information Extraction (IE) refer to entities referenced by names such as person, organization, location, date, money, etc. NEs represent instances of classes and are domain specific. Word senses are more generic and are described by using different levels of linguistic description such as morphology (description of the structure of linguistic units such as words and part of speech), syntax (rules for constructing phrases and sentences), semantics(inferred meaning from concepts), and pragmatics (inferred meaning from context). With NEs, SA entails coupling defined entities with formal descriptions.

One of the approaches used for creating metadata is called Tagging. Tagging is a process of attaching a piece of information to a document. There are two types of tagging systems – user-based tagging and system-based tagging. System-based tagging isautomatic extraction of pieces of information called instances from documents based on defined ontology. In user-based tagging, user creates tags based on his/her intention and interpretation of the document being tagged. The creation and management of tags collaboratively among a community of users for the purpose of classifying information are referred to as Folksonomy. Thomas Vander Wal coined the term Folksonomy from folks and taxonomy (Wikipedia, n.d.).

Folksonomy can be regarded as bottom up approach classification as opposed to top-down approach taxonomy in ontology. According to Wikipedia (n.d.), Ontology in computer science and information science formally represent knowledge as a set of concepts within a domain, and the relationships between those concepts. It can be used to reason about the entities within that domain and may be used to describe the domain (Ontology, n.d.).Folksonomy is non-hierarchical ontology that requires no expertise. Users add tags of their choice to web content and make use of the tags to organize, classify or categorize web content.

User-centeredness is one of the major areas in Economic Intelligence (EI). EI is about deducing intelligence by EI actors from available information for the purpose of taking proactive decision. The process of deducing intelligence is called EI process.

1.2 Statement of Problem

Decision-making process in EI context is often based on available information as well as on the person who needs it or interprets it. Actors of interest are Decision Maker (DM), Information Watcher (IW) and Coordinator. The actor's knowledge is elicited by capturing his/her interpretation from the document which the actor has annotated. Capturing the actor's interpretation of a document or document-object requires that such actors express explicitly such interpretation. Manual annotation can be used for creating annotation, but most of the existing manual annotation platforms lack the power of the actor's expressivity as required in EI context. Many of the existing manual annotation platforms have predefined annotation types. The types are used for categorizing user's annotation. The use of any of such platforms by economic actors implies that the actors need to specify one of the annotation types to categorize the annotation they make. In a situation where any of the annotation types specified does not represent the objective of making the annotation, one of the following has to be done; either the actor selects the closest annotation type to represent his/her intention or the annotation schema is extended to accommodate such structural changes. The first option might not be satisfactory to the actor because the actor is constrained to choose from the available options. The second option however, requires that the actor should be an expert and has access right to perform such an operation; this may not be true in most cases. The work of Robert (2007) was on annotation in the context of EI. The work introduced annotation model called Annotation Model for Economic Intelligence (AMIE). The work viewed annotation as a function of user and document in assisting economic actors to interpret retrieved information for solving decisional problems by providing a base for an enhanced information research. The study however, did not consider semantic annotations. Therefore, there is a need for a model tocreate and structure semantic annotations that adequately capture the intention of the actors in decision making process.

Furthermore, an information system often contains pre-selected, sorted, and processed information for a particular use. Thus, information contents are structured according to targeted end use. Given such structured information, users can express information needs more precisely. They can specify values in relation to attributes of the information. However, a user needs to have a minimum knowledge of his/her needs by specifying at least a value of the stored attributes, since information search is still done by the principle of search by content. Problem

arises where the user does not know what he/she is looking for or specifiesan attribute that does not exist. Where a user specifies an attribute that does not exist, there would be a need for structural changes to accommodate new attributes. Such changes might be tedious, time consuming and costly to effect. Therefore, there is a need for a mechanism for exploiting stored information in a situation where the user does not have minimum knowledge of the content of the information system.

In addition, searching for relevant information from large corpus such as World Wide Web (WWW) is still being done by content search. This is as a result of how information is still being represented on the web. A search technique that would allow actors to search for information of interest based on the objective of the search will be very desirable. Therefore, there is a dire need for search algorithm that will enable actors to search for information of interest based on the objective of search.

1.3 Purpose of Study

The aim of this study is to design and develop an Attribute-Value Pair (AVP) model for creating and exploiting annotations in Economic Intelligence. The specific objectives can be summarized as follows:

1. To design and develop an annotation model for creating and storing annotations taking into consideration the semantic context of annotators.
2. To devise a mechanism for exploiting stored annotations based on the context of problems,
3. To develop a search algorithm that will allow actors to search for information of interest based on the objective of the search.

1.4 Research Questions

This study intends to answer the following questions:

1. What annotation model could be developedthat will allow actors to load, annotate and store annotations taking into consideration the semantic context of annotators?

2. How can this study realize the exploitation of the stored annotations and other relevant information that will enhance solving decision problem based on the context of the problem?
3. What search algorithm is required that will enable actors to search for information of interest based on the actors' objective of search?

1.5 Significance of Study

The purpose of this study is to design and develop an annotation model that will improve decision making process. Decision process in the context of EI is based on availability of relevant information as well as interpretation of actors on the available information. This study aims at providing a model that will optimize method of obtaining relevant information as well as capturing the actor's interpretation of document for decision making process.

1.6 Definitions of Operational Terms

AMIE:	Annotation Model in Economic Intelligence
AMTEA:	Annotation Model and Tools for Economic Actors
AVP:	Attribute-Value Pair
DMP:	Decision Maker Problem
EI:	Economic Intelligence
EQuA^{2}te:	Explore, Query, Analyze and Annotate
IE:	Information Extraction
LORIA	Laboratoire Lorrain de Recherche en Informatiques et ses Applications
MEDP	*Modèle pour l'Explicitation d'un Problème Décisionnel* (*Decision Problem Clarification Model)*
NEs:	Named Entities
RDF:	Resource Description Framework
SA:	Semantic Annotation
SIMBAD:	Multimedia Indexing System Based on the Document content Analysis

SITE:	Modélisation et Développement des Systèmes d'Intelligence Economiques
SW:	Semantic Web
W3C:	World Wide Web Consortium
WISP:	Watcher Information Search Problem
XLST:	Extensible Stylesheet Language Transformations
XML:	Extensible Markup Language

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

The central concept of interest in this research work is annotation. In this thesis, annotation is viewed from the context of Economic Intelligence (EI). Hence, there is a need to explain these two principal concepts upon which this thesis is based - Economic Intelligence and Annotation. In this chapter, the concept of Economic Intelligence, as defined in literature is critically examined; its systemic components are identified and discussed; interaction among the components and how it affects the goal of EI is also reviewed. In addition, the concept of annotation is discussed. The existing annotation models are examined. Also, the relevance of annotation in achieving the overall objective of EI process is also explained. The aim of this Chapter is to deduce the conceptual framework for the design of annotation model in the context of EI.

2.2 The Concept of Economic Intelligence

Discovery of relevant information for solving decision problems in timely manner is pivotal to making right decision. Decision-making process is often based on available strategic information. Such information however, has to be sought, collected, processed with a view of eliciting knowledge relevant to the problem as well as representing both the collected information and elicited knowledge in a form that will facilitate information reuse. Decision based on information alone may not be sufficient if the preference of the decision maker is not taken into consideration. While the information might be available, with all signals visually transparent, inability to properly and timely understand such signals as possible indicators could be very costly. Information of significance to one person may not be to another person depending on individual interest. Hence, availability of strategic information is as important as the person who needs it or interprets it. Therefore, in the context of Economic Intelligence (EI), decision making does not depend on the available information only but also on the interpretation given by the concerned users called economic intelligence actors. EI therefore, spans two major research domains – Strategic Information System and User Modelling.

Economic Intelligence is different from related concepts such as Business Intelligence (BI), Competitive Intelligence (CI), Territorial Intelligence (TI), Technology Watch (TW), and Technical Intelligence to mention but a few. The focus of BI is on using business processes and tools to process internal information of an organization for decision making purposes. CI concept is mostly used in United States of America. It is used for analyzing competitors-environment and stakeholders for decision making purposes. In TI, the study is on the impact of socio-cultural beliefs as they affect decision making. The other related concepts however, consider information sources majorly, the external information that support decision making. The specificity of each approach makes it different from EI which is all encompassing. Therefore, there is need to have a closer look at EI as a concept.

2.3 Economic Intelligence and its Systemic Components

There are many definitions of EI depending on the point of view. Prominent among the set of definitions of EI are those given by Martre (1994) and Revelli (1998), imbibed and cited by the research team SITE in LORIA. According to Martre, EI is defined as "a set of coordinated actions of search, processing and distribution for exploitation of useful information for economic actors. These actions are carried out legally with all the necessary protection for the safeguard of the company's patrimony, and with the best quality, delay and cost". Revelli, defined the EI as "the process of collection, processing and distribution of information with the goal of reducing uncertainty in taking strategic decisions". From the definition of Martre, EI can be viewed as a *system* that delivers well articulated and value-added information to a company that needs it at reasonable cost and at a minimal time. Revelli, however, viewed EI as a *process* of gathering and delivering such information. Combining the two definitions, it can be deduced that the goal of EI is to produce actionable knowledge necessary for taking strategic decisions. From systemic view of EI, the components in this specific decision environmentare human actor, information, decision problem, analysis tools and medium of communication which interact to produce the desired goal. The definition of EI inMenendez et al. (2002) is more encompassing. The definition combines the systemic as well as the process views of EI. It explains EI as follows:

"Economic Intelligence, concerns the set of concepts, methods and tools which unify all the coordinated actions of research, acquisition, treatment, storage and diffusion of information, relevant to individual or clustered enterprises and organizations in the framework of a strategy"

EI involves transformation of collected data from various identified sources first into information, then into knowledge and knowledge into intelligence through EI process and various available tools. It is to be noted here that the term Intelligence means actionable knowledge (Odumuyiwa and David, 2008). The responsibility of Intelligence creation is solely that of human. Menendez (2002) put it in this way: "…the human factor is pivotal in the process of creating Intelligence in any kind of company or organization".

Having discussed the concept of EI in the preceding section, the next section examines two of the identified systemic components of EI – human actor and information. The choice of these two components for discussion is centred on the goal of this thesis. David (2010) put it in this way, "the three elements constituting the basic elements of all projects in EI are user, information, and the use of information."

2.3.1 Human Actor

Human actor refers to various users involved in the creation of intelligence in the context of EI. Such users are referred to as economic intelligence actors. According to research team SITE, actors of importance are mainly Decision Maker (DM), Information Coordinator (IC) and the Information Watcher (IW).

Decision maker: Decision maker is a person in an organization saddled with responsibility of initiating decision problem and taking decision as it affects the organization in terms of stake, risk or threat (David and Thiery, 2001). Decision maker works in collaboration with other economic intelligence actors to resolve decision problem. In Kislin (2007) and Odumuyiwa (2010), the functions of a decision maker include ability to identify weak signal in the environment as it affects the organization and devise strategies of improving organizational performance; operating one of the identified strategies of choice depending on the characteristics

of the organization; allocating financial and human resources to specific actions arising from the decision making; and taking responsibility for consequences resulting from decision taking.

Information Watcher: Information watcher (or simply watcher) is the person responsible for translating decision problem into information search problem (Kislin, 2007). Watcher identifies, collects, analyses information and identifies necessary indicators in the processed information to make it more lucid for decision maker to make decision. One of the major roles a watcher plays in the process is in adding value to the processed information based on interpretation given to calculated indicators.

Coordinator: According to Knauf(2007) and Odumuyiwa (2010), the Coordinator, also known as infomediary in (Knauf, 2007) refers to an actor that coordinates the process of economic intelligence among other involving actors. The actor (coordinator) serves as intermediary between the decision maker and information sources, has a good knowledge of the environment (Goria, 2006), and communicates regularly with the decision maker on how the environment may affect decision process. He/she is considered a specialist who consistently monitors changes in information contents that are of relevance to decision process and reports same to decision maker.

The relative importance of human resources to available information in decision-making process spurred researches in this domain. The work of Bouaka (2004) centred on modellingdecision maker and decision problem. Kislin (2007) focused on Information watcher and Information search problem. The work of Knauf (2007) was on Infomediary, a nomenclature that was later changed to Coordinator. These actors complement each other's activities in resolving a decision problem.

2.3.2 The Concept of Information

Information is pivotal to the growth of any organization, be it social, political, economic or scientific. Organizations need to obtain relevant information to improve their efficiency in order to compete favourably in the ever increasing digital economy. However, with the advent and rapid growth of the Internet, there exist vast sources of information on the web. The problem is not on the availability of such information but in identifying and devising an efficient way of

retrieving relevant information from sources in timely manner, thus contributing to solving the decision problem. This research study believes that a user's query expressed as Attribute-Value Pair will add semantic context to such query and this will greatly enhance identification of relevant information from large corpus such as World Wide Web. The proposed AVP search algorithm in this thesis will assist user to retrieve contextualized information.

Various attempts have been made to define the concept of information in literature. Information is a generic term that has variety of connotations in different research fields. This study agrees with Buckland(1991) who gave tripartite definitions of information concept as information-as-thing, information-as-process, and information-as-knowledge, depending on the context of use:

- **Information-as-thing**: This term refers to the information objects obtained, perceived or read e.g. in a document, that has the quality of imparting knowledge or communicating information to a reader. It tends to answer *what* information is.
- **Information-as-process:** According to Oxford English Dictionary (1989) Information-as-process is defined as "the act of informing…; communication of the knowledge or 'news' of some fact or occurrence; the action of telling or fact of being told of something". Here, information is regarded as an act of changing the level of knowledge of a person being informed. It tends to answer *how* information is being communicated to the receiver. This process could be a set of actions or reactions, as the case may be, leading to the specified goal.

  However, the definition failed to address the role of the receiver, the person being informed. His/her role is vital because a decision problem depends not only on the person who formulates it but the interpretation of information depends on who needs it. Robert (2007) addressed this problem in his definition of information as a process. Robert defined Information-as-process as an act of man giving a specific meaning to the data, its processing activities and the communication of that interpretation. Only when his/her specific interpretation is communicated and understood by the receiver that we can say that there is information.
- **Information-as-knowledge:** This focuses on the semantic meaning of the information being communicated with the aim of reducing uncertainty. It answers *why* of information.

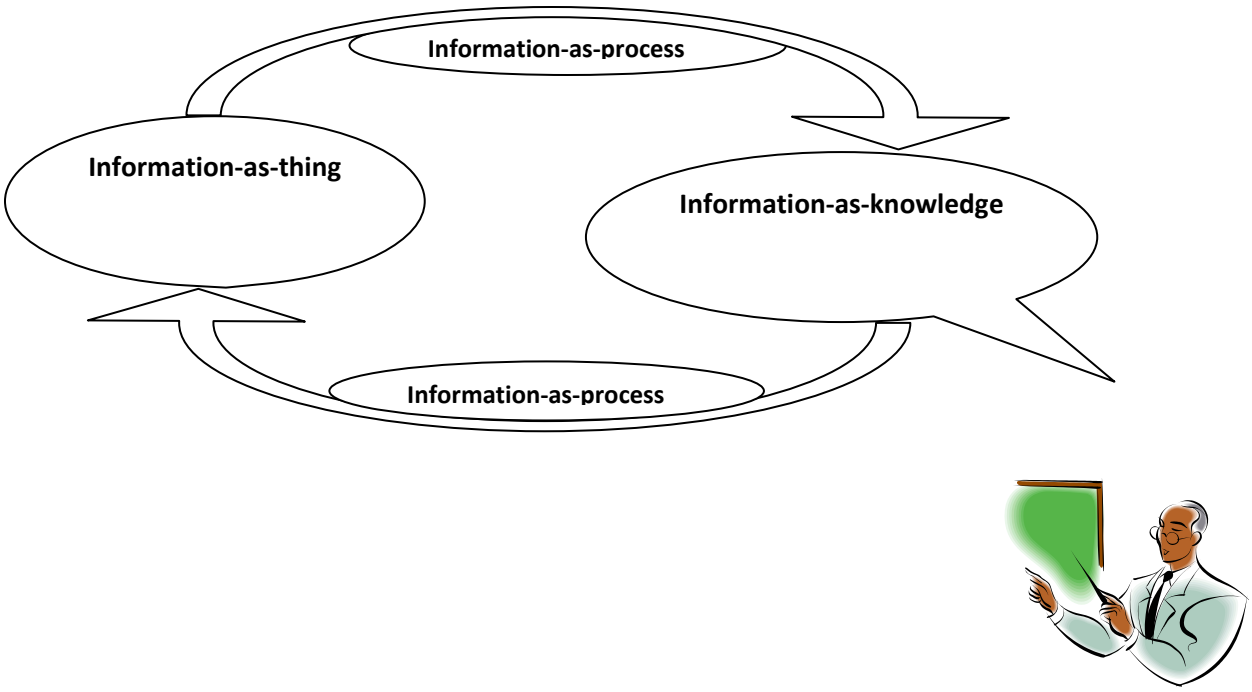

Figure 2.1: The Inter-Relationship among the Tripartite Meanings of Information

It can be inferred from these tripartite definitions of information that the transformation of information-as-thing into information-as-knowledge is through information-as-process. This is illustrated in Figure 2.1. The process however, is recursive. What is regarded as knowledge at an instance could be regarded as information object in another instance, depending on the context of use and /or the level of knowledge of the receiver.

2.3.3 Sources of Information

There are various sources of information – books, World Wide Web, databases, journals, encyclopaedia, to mention but a few. The sources can, however, be classified as formal and informal sources. While the formal sources refer to information that has been documented and published, informal sources refer to verbal information passed from one person to the other. Formal information can be classified into two major sub-classes. These are open sources and information from Information bases (David, 2010) sources. With the open sources, information is freely, directly and lawfully accessible with little or no security. Examples of information in this category include published information on the web. Restrictive source, however, allows only authorized persons to have access to its usage.

Web has undergone series of changes, from Web 1.0 to Web 2.0. Web 1.0 is a non-participatory website containing read-only and static pages. The depositor determines the information contents based on his/her initiative and perception of the domain. The users, on the other hand, use the available information to meet their needs. There is a possibility of a gap existing between the goal of the user and the goal of the depositor (or author) as a result of lack of correspondence between the information contents and the user needs (David, 2010). The need to proffer solution to this problem as well as advancement in technologies led to the development of a more robust, dynamic and participatory web called Web 2.0 where web contents are jointly authored and used by different users. Social networking sites, folksonomies, blogs, wikis are some of the examples of activities possible on Web 2.0. While a community of users can co-author information, representation of such information, for easier information retrieval, becomes an important challenge.

The Information repositories contain information that has been pre-selected, and sorted. Thus, the information contents are structured according to the use identified by the author. Given such structured information, users can express information needs more precisely. They can specify values in relation to the attributes of the information. However, users could be faced with some challenges. For instance, it is required of the user to have a minimum knowledge of his/her needs by specifying at least some values of the stored attributes, since information search is still done by the principle of search by content. Problem arises where the user does not know what he/she is looking for. Another challenge is structural changes to the system that may arise as a result of evolving needs, since the designer may not be able to anticipate all possible attributes a user may need. Where a user specifies an attribute that does not exist, there would be a need for structural changes to accommodate new attributes (Afolabi, 2007). Such changes might be tedious, time consuming and costly to effect.

In the next section, EI process for creating intelligence is discussed, the relevant EI actors involved in each stage are stated as well as issues arising from such process.

2.4 Economic Intelligence Process

EI process is an activity that is made up of a set of actions for solving decision problems. The focus is on obtaining relevant but legal information that could be useful to economic actors to formulate and represent precisely, decision problems in such a way that would optimize the solution to the problem of interest. According to the research team SITE, LORIA, EI process is decompressed into eight activities from the usual ten activities in Bouaka and David (2004). The activities are interwoven and the process iterative. Each activity is as important as the other. The Economic Intelligence process activities and the major actors are depicted in Table 2.1.

Table 2.1: EI process and major players

Phase	EI Activity	Major player(s)
A	Identification of a decision problem to solve	Decision Maker
B	Transformation of decision-problem into information search problem	Economic Watcher
C	Identification of relevant information sources	Economic Watcher
D	Validation of the information sources.	Decision Maker & Economic Watcher
E	Collection and validation of information	Economic Watcher
F	Processing the collected information for calculation of indicators	Economic Watcher
G	Interpretation of indicators	Decision Maker
H	Decision making for the resolution of the problem	Decision Maker

The iterative process of Table 2.1 is shown in Figure 2.2. David (2008) identified five issues based on the EI process. These are reproduced below:

1. Issues relating to the decision-making process that cover phases a, g and h, whose players are mainly policy makers. Here, the interest is on understanding of the problem by both the decision maker and the person responsible for searching for information to solve the decision problem.

2. Issues relating to the process of gathering data, which cover phases b, c, d, e and f where the players are the watchmen, the staff of the organization and designers of information systems. Here, the interest is on skills to search for relevant information and experiences in the information search.
3. Issues relating to the adequacy of information covering phases b, c, d and e, whose players are mostly watchmen; here, the interest is on the relevance of information in solving the decisional problem.
4. Issues relating to the protection of information heritage covering all phases where the players are policy makers, watchmen, the staff of the organization and designers of information systems; Here, the interest is on knowledge precautions to be taken in order to safeguard the material objects of the organization.
5. Issues relating to the use of information as a weapon of influence either positively or negatively. The players are mainly policy makers but also watchmen. This point is important because a decision often has an impact on the external environment of the organization. It is necessary to have an expert who could determine issues, or facilitate the implementation of the decisions taken.

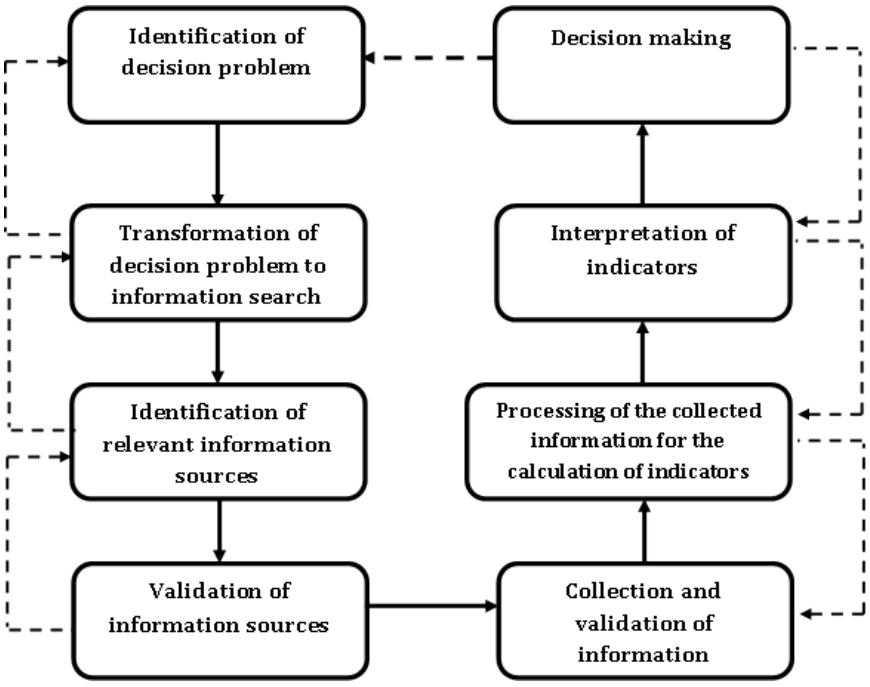

Figure 2.2: Economic Intelligence process

From Table 2.1, it can be seen that information is pivotal in EI process. Information that is adjudged to be relevant to a decision problem is collected, processed and distributed to relevant actors. Knowledge is inferred when context is added to the processed information (Pohl, 2000).

Research team SITE sub-divided Economic Intelligence process into four phases. These are: Selection, Mapping, Analysis, and Interpretation. Figure 2.3 shows EI architecture and existing models according to research team SITE.

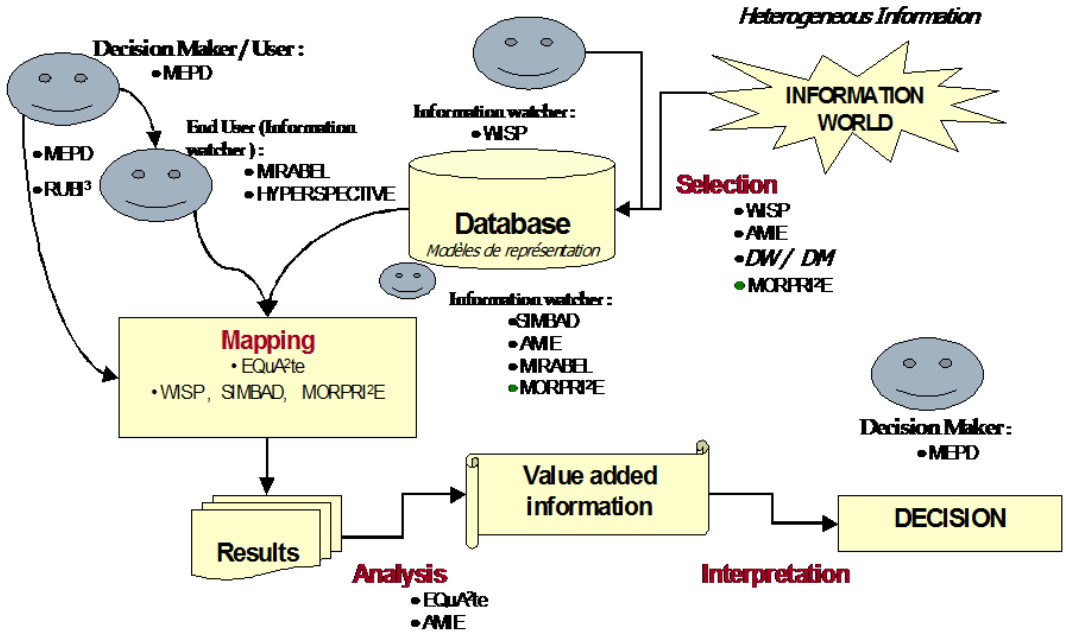

Figure 2.3: EI Architecture and Designed Models

2.5 Relevance of Information in EI

In the previous section, it was established that information is important in EI process. Equally important is the person who needs the information or interprets it. However, user interpretation depends on two important things: the knowledge of decision maker on the decision problem and his/her willingness to fully share the same with the other participating actors such as Information watchers. Information that meets the need of a user could be referred to as relevant information

(David, 2010). Such information is expected to satisfy the expectation of users in terms of the goal for solving decision problem. Information that is adjudged to be relevant to solving decision problems is primarily the prerogative of the user, (decision maker in this case). In EI, information relevance is judged based on the contribution of such information to resolving decision problems. Three indicators (see Figure 2.4) used to measure information relevance in EI are: knowledge of the decision problem, knowledge of information search problem derived from decision problem (David, 2010) and user (Information watcher) interpretation.

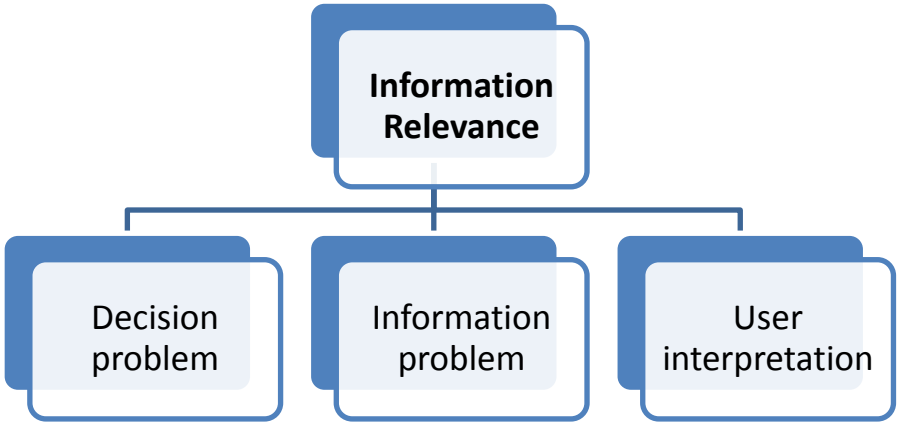

Figure 2.4: Aspect of Information Relevance in Economic Intelligence

2.5.1 Information Relevance with regards to Decision Problem

A decision problem often starts when a decision maker perceives signals of concern or receives information requiring decision making. This is referred to as identification of decision problem phase in EI process. In David(2010), decision maker is responsible for this phase. Decision maker therefore initiates a decision problem which is contained in a document called initial demand. The information contained in the document is a materialized knowledge of the decision maker on the decision problem to be solved. Decision maker works in collaboration with other economic intelligence actors to resolve a decision problem – particularly the watcher. Resolving a decision problem collaboratively tends to offer better and more feasible solution than when resolution is based on personal perspective only. Collaborative problem resolution according to Windle & Warren (n.d.) tends to proffer better solutions as it gives the opportunity to consider many options based on divergent perceptions resulting from individual expertise and interest that could meet the objective of resolution. However, taking collaborative approach entails that an

individual does not stick to his/her position as the only acceptable solution as this can lead to disagreement and dissatisfaction among members. An individual's position may largely be influenced by his/her level of understanding and his/her perception about the problem. In Windle & Warren (n.d.), perception is defined as "our interpretation of our world and our experiences impacted by our values, beliefs, fears, and desires." The perception of red cloth to a person in Africa may symbolize danger while to a Chinese, it is a symbol of love. Individuals may share their perceptions of the decision problem through annotation in a collaborative environment.

Bouaka (2004) proposed a model called MEDP (*Modèle pour l'Explicitation d'un Problème Décisionnel*) for specifying decision problems. The model is aimed at specifying decision problem to enhance better understanding of decision problems among the concerned actors. The model specifies decision problem in terms of information on the decision maker, the stake involved, and information on the internal and external environment of the concerned organization.

According to research team SITE, a stake is a set of:

- *object*- environmental object detected or proposed by decision maker,
- *signal* – the meaning the decision maker gives to the detected object, and
- *Hypothesis* – the possible result or outcomes associated with each signal.

What constitutes the *object* may be inferred from the decision problem. However, the signal(s) and the hypotheses are based on the overall objectives of solving the decision problem. This phase is of utmost importance in that a well defined and well articulated decision problem will greatly assist in collection of information relevant for solving such problem. The research question to be asked is "how can an annotation be used as knowledge elicitation mechanism for proper definition of decision problems?"

2.5.2 Information Relevance with regards to Information Search Problem

Transforming explicitly defined decision problem into information search problem constitutes the second phase of EI process as stated in section 2.4. Information search problem specifies information needed. This is achieved by retrieving relevant information from verified information sources. This function (retrieving relevant information) is performed by information

watcher (IW) in EI context. Based on the decision problem to be resolved, we identify relevant information sources as well as perform the retrieval functions. Here, their experience plays a crucial role.

Kislin and Bouaka (2002) proposed WISP (Watcher Information Search Problem) model for transforming decision problem into information search problem. A closer look at model WISP reveals the definition of stake as part of the model but with different notion. Stake is defined over four tuples: the type of stake, the goal of stake, the purpose of stake, and hypotheses. These tuples are well defined and acceptable according to the research team SITE. IWs could instantiate these attributes through the process of annotation.

2.5.3 Information Relevance based on the Actor's Interpretation of Document

When information sought for from identified sources is retrieved and organized based on need, such information is analyzed and interpreted by the concerned economic intelligence actors with the aim of reducing uncertainty in taking decision. Information Analysis (IA) is a phase in EI process. The phase centres on providing a relevant and valuable knowledge from synthesis of collected information. According to Menendez et al. (2002), IA consists of two major steps – information validation and transformation of information into knowledge. While information validation validates collected information with respect to its relevance and usefulness to the decision problem being resolved, information transformation involves adding an actor's interpretation to information based on the meaning adduced from such information. Documenting an actor's interpretation of information sources in the form of annotation-object based on the actor's experience and/or level of skill could prove very valuable. Such annotation patterns of the actor overtime could reveal interesting patterns that could assist in discovering new knowledge about the actor. Similar decision problems may not necessitate similar solution. The solution largely depends on decision maker's interpretation, time and context.

Information collected from different sources could be used to populate the information base of the concerned organization. Such information, however, has to be structured according to the schema of the information base. Annotation-process could be used to add other attributes that are

not provided for in the design of the information base schema. The purpose of processing information in the context of economic intelligence is to identify possible signals that could be relevant to solving decision problem. Therefore, it is necessary to capture what an actor designated as relevant information to solving decision problem; capture the actors' interpretations as value-added information to such information as well as the exploitations of the captured information. Annotation could be used to perform the above outlined tasks.

From the foregoing discussion, annotation is relevant in the context of EI. It is therefore, pertinent to discuss the concept of annotation.

2.6 Annotation as a Concept

The definition and use of annotation are closely linked with the objectives, context of use and available information. Several definitions of annotation exist in literature. Annotation is defined in Bodain et al. (2007) as a note, an explanation, or any other type of external remark that can be attached to a document without necessarily being inserted into the document. From the definition, annotation is viewed as separate document containing extra information to the existing information (object). Its purpose is for interpreting the underlying document (document interpretation). Brusilovsky (1997) also defined annotation as "*any object (annotation) that is associated with another object (document) by some relationship"*. The definition does not only consider annotation as object but also as an action involving *anchoring* the object with the concerned document. Of particular interest is the definition of annotation given by MacMullen (2005). He gave tripartite definitions of annotation in similitude to Buckland's definition of information (Buckland, 1991); that is, Annotation-as-thing, Annotation-as-process, and Annotation-as-knowledge. He defines each part of the concept as follows:

Annotation-as-thing: *An annotation is an intentional and topical value-adding note linked to an extant information object.* Apart from viewing annotation as object, the definition also states the purpose of annotation as value-added information.

Annotation-as-process: *annotation is a process that has the function of creating or modifying an information object called an annotation.*

Annotation-as-knowledge: *Annotation is the intellectual component of an annotation, distinct from its physical manifestation*. Annotation as source of knowledge is very interesting especially when such knowledge is available to reduce uncertainty surrounding the decision making process. In Menendez et al. (2002), knowledge is the existence of pattern relation and its implications resulting from processed information that is realized and understandable by the user of such patterns. The word "user" could be human agent or not.

It can be deduced from the foregoing definitions that annotation is an object created on existing document; it is also the process of creating annotation; the purpose of annotation creation could be for document interpretation, value-adding, document comprehension, or serves as source of knowledge. The work of Robert(2007) dealt extensively with annotation-as-thing and annotation-as-process in the context of EI. Annotation-as-thing and annotation-as-process refer respectively to annotation as an object and as an action in the study. The aim of this thesis is to look at the semantic annotation in relation to decision problem.

2.7 Semantic Annotation

The Web has been dubbed as the largest source of Information. However, the bulk of information representation still support search by content. One of the goals of semantic Web according to World Wide Web Consortium (W3C) is to make Web content accessible and processable by both human and machine agents by converting the web content to web of linked data. It implies that web contents are coupled with additional information that interprets such web content. Such added information is regarded as metadata - data of data content. The metadata are linked with knowledge repository using Resource Description Framework (RDF) technology, for example, to give formal semantics to the web content. The used metadata connotes the interpretation of web content or its fragments. While with the human agents, information interpretation can be subjective and mostly depends on some factors such as intellectual capacity, experience, cultural background and environment. Information interpretation to machine agents however, is expected not to be ambiguous. To achieve this, it implies that such information representation must be of the form that is recognisable by machine agents. Semantic Annotation (SA) has been a major method used in adding metadata to Web contents.As pointed out in the section 1.1, one of the natures of web content being considered is

Named Entities (NEs). In Sekine (2004), the history and future of NEs were discussed extensively. With NEs, Semantic Annotation entails coupling defined entities with formal descriptions. Furthermore, it was mentioned in section 1.1 that with user-based tagging, user creates tags based on his/her intention and interpretation of the document being tagged.Tag Cloud is a visual representation of the list of user-generated tags in a website arranged alphabetically with the font size or font colour indicating the relative importance of one tag to the other. It also refers to word content distribution in a website for describing such website. Three types of tag cloud were identified (Tag cloud, n.d.) depending on applicability. In each type, the attributes of text chosen such as font size, font colour and weight determines the representation. A chosen attribute represents the frequency a tag is applied to an item; frequency of items a tag has being applied; and the use of tag for categorization method with a chosen attribute representing number of subcategories. Other variations of tag cloud are listed in Martin and Mark (2005).

2.8 Annotation Models

Several annotation models have been proposed. We shall briefly examine some of them, and technology deployed to implement the developed annotation tools with a view to determining what constitutes an annotation process.

2.8.1 A Common Annotation Framework (CAF)

A Common Annotation Framework (CAF) proposed by Bargeronet al. (2001) and based on a "principled logical model" was designed out of the need to support all-inclusive varieties of annotation activities that are simple, and extensible. CAF was based on abstraction defined in XLink technology. The model uses XPointer technology to position data for anchoring annotations. A CAF annotation object consists of the anchors, the annotation relationships among the resources they specify and the metadata properties describing characteristics of the annotation relationship. Based on the model, Bargeronet al. (2001) came up with an Annotation Markup Language Schema (AML-Core Schema) that encapsulates the structure, functionality, and interrelationships described in the model. Attributes defined in the XLink were used in

designing the schema. Based on this model, a document or any part of it is regarded as a resource structured and is described by RDF/XML technology. Annotation created is also seen as a resource.

Figure 2.5 shows an example of an annotation object. Anchoring system is however not part of the core framework. Keyword Robust Anchoring proposed in Brushet al. (2002) was used in implementing the model in WebAnn.

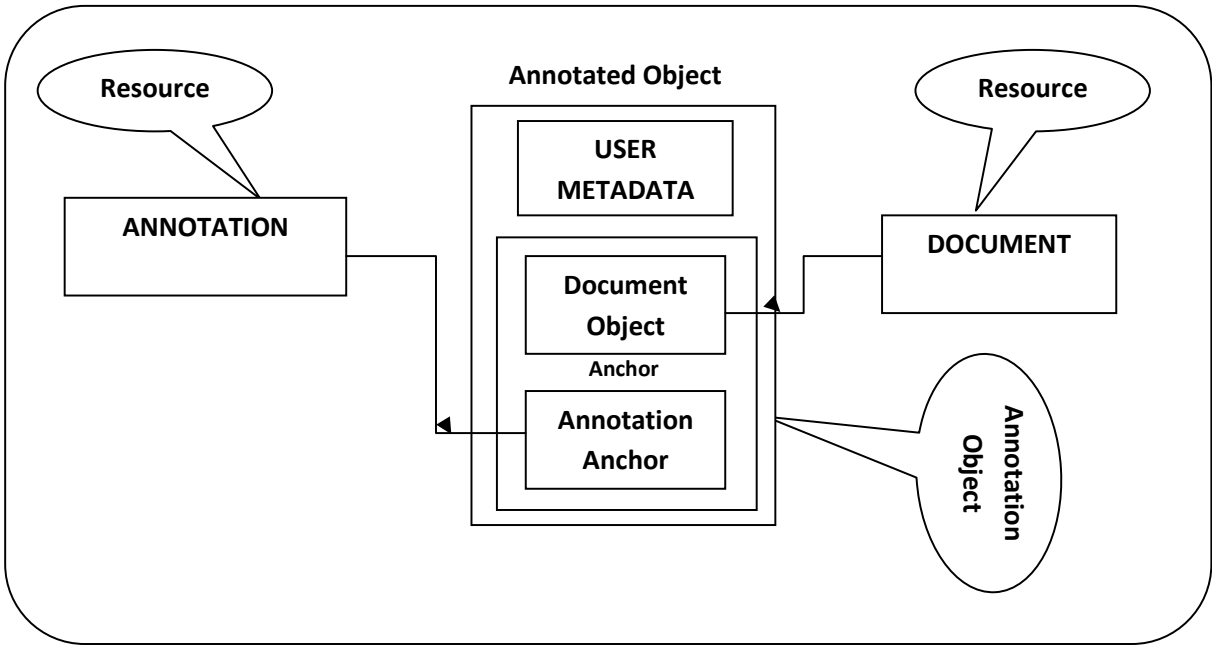

Figure 2.5: Annotated object

Source: (Brush et. al., 2002)

2.8.2 Personalized Annotation Management (PAM)

The structure of Personalized Annotation Management in Yang et al. (2004) contains two models: content model and annotation model that formally describe e-documents in HTML format and annotation, respectively. They established relationship between annotations made by the annotator and selected object in the e-document by anchor. Clustering mechanism was developed to cluster annotation description and e-documents separately. Using the term-

document matrix format, they established inter-relation between annotations and inter-relation among e-document. Users could therefore, search for related annotations and e-documents using the term ‘cluster’.

XSL technology was used to invoke and incorporate annotation into e-document during browsing without modification to the original copy ***(XSL-based Anchoring mechanism).*** Figure 2.6 shows the adapted form of the structure and Figure 2.7 indicates the two models that the structure described.

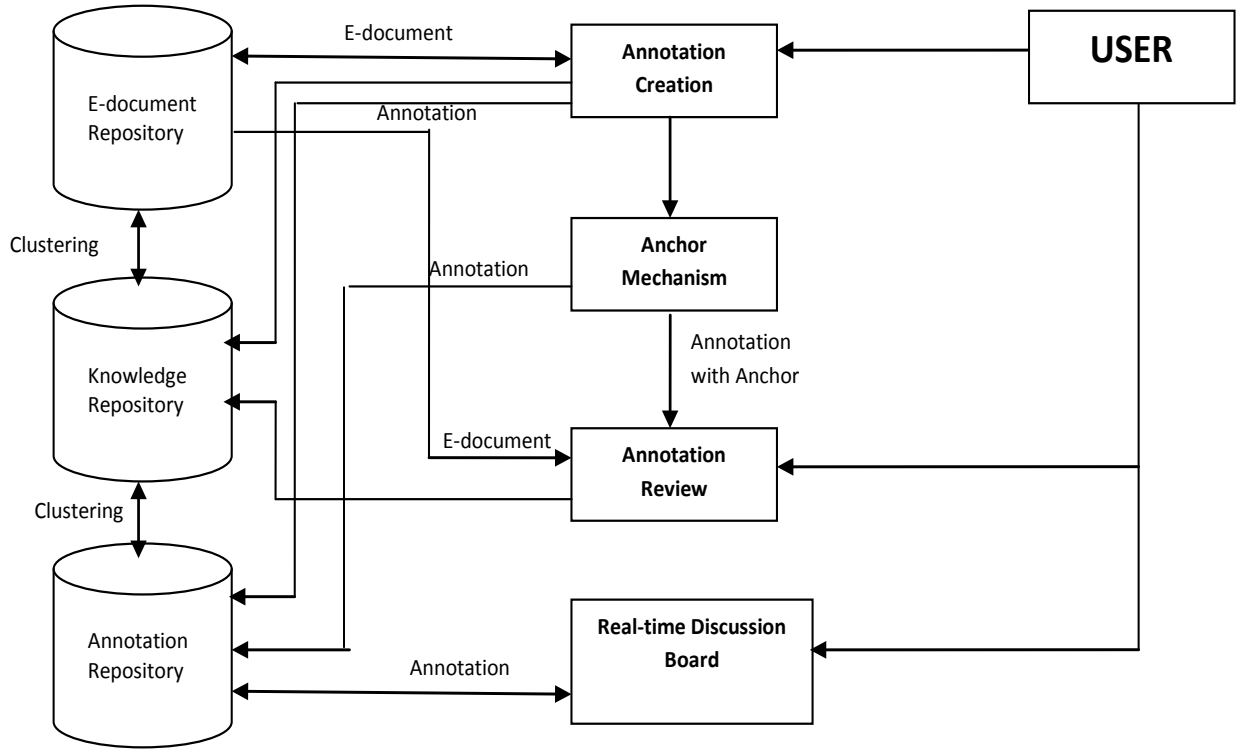

Figure 2.6: Adapted PAM Framework

Source: (Yang et. al., 2004)

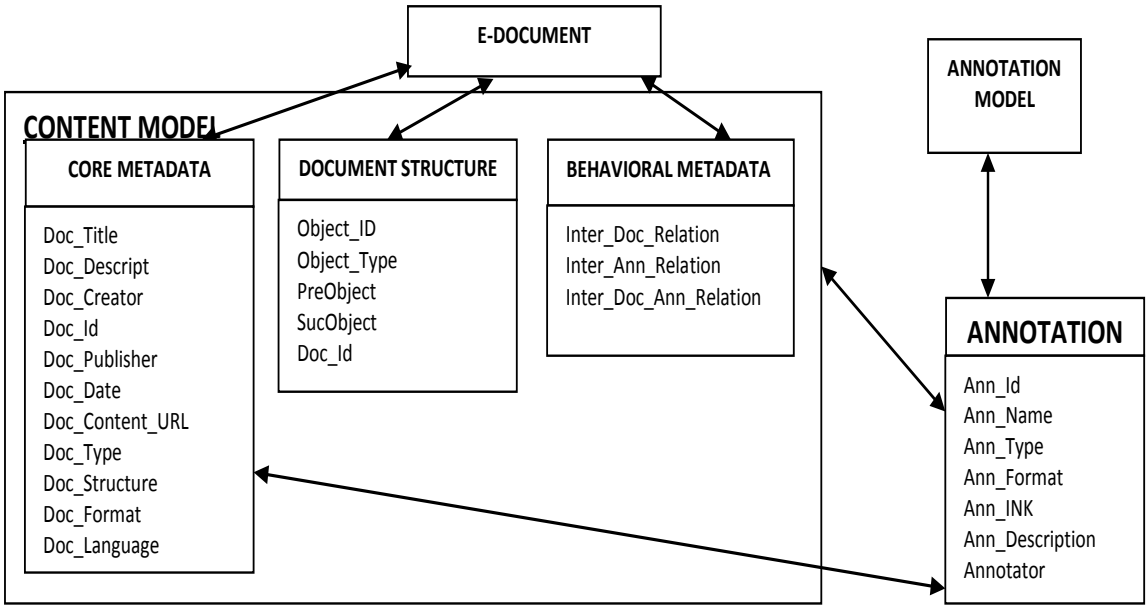

Figure 2.7: Content Model and Annotation model of PAM

Source: (Yang et. al., 2004)

The anchoring algorithm used is stated below:

- Retrieve structural metadata of the concerned e-document
- Identify the desired document object by highlighting
- Compute anchor position which consists of a start and a stop position of the highlighted object
- Handwrite annotation using annotation pad
- Recognise the handwriting note as digitalized text format
- Store the handwriting note and recognising the associated text information with anchor position

The major concern here is associating annotation created to its annotated object. It however did not include the metadata of the personality of the annotator. The knowledge on the annotator is very important in order to know how annotator's personality has influenced the added information. For example, the cultural background of the annotator may have effect on his/her interpretation of colour.

2.8.3 Web Indicator by Sharing Personal Annotations framework

The model by Sannomiya et al. (2001) shown in the Figure 2.8 was aimed at retrieving information from a community of web users through personal annotation. In information retrieval, the model allows users to make personal annotation on retrieved web pages and ascertain the usefulness of such pages if they meet their information need. The model also allows users to browse other users' annotations and make annotation on other users' annotations. The framework defined three behaviours: Personal Annotation (PA), Merged Annotation (MA) produced as a result of searching of others' annotations based on rules (such as *range of search*, *merge condition* and *display condition*) defined by the users, and the feedback from information given. Each PA is stored as separate web page as a formatted XML document, (using an XPath data model for its structure) either on the PA Deport or user's local disk. Temporary XPath data model is used for representing the history of the XML documents (to take care of changes). XML schema used to define PA's schema and validated using XML Schema Validator. PA manager saves users' PA, displays their own PA using XSLT after identification, and restricts unauthorized users from accessing other PAs. Query Manager searches and retrieves relevant PAs from the PA Depot and/or local disk based on the user's rules; restructure them (using XPath), merge them together (using XLink extended) and store in the MA database. It then calls the Display Manager to sort and convert it into an XHTML or HTML document for display to the user (using XSLT). Users' notification is realized through dialog box system and not through e-mails

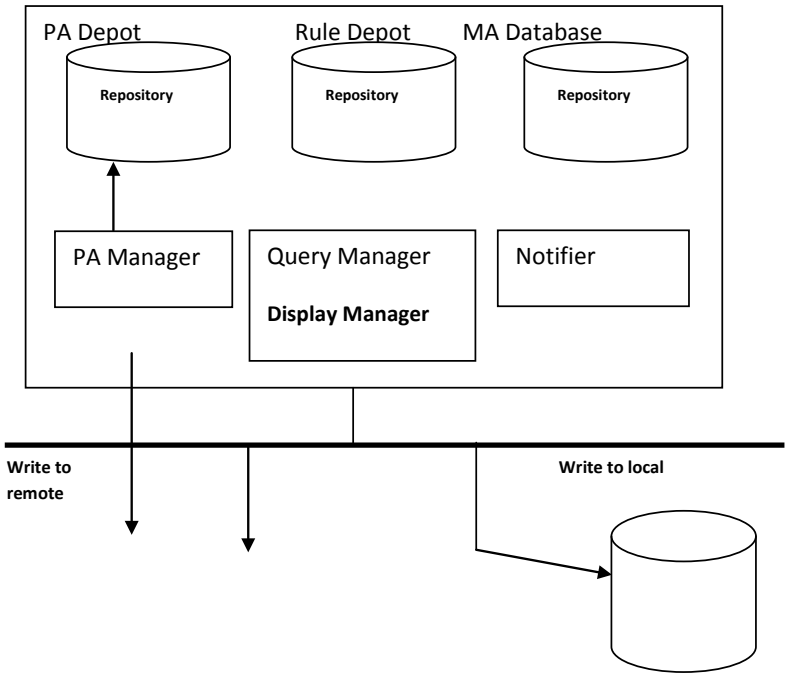

Figure 2.8: WISPA Architecture

Source: (Sannomiya et. al., 2001)

2.8.4 Metadata Management Systems

Metadata management system uses annotation to provide a linkage between a conceptualized structure of a specific domain, Neurosciences, and web documents of the domain. It was aimed at addressing and solving research problems associated with the need for querying, sharing and exchanging of heterogeneous data (data from different sources) generated from experiments and observations in collaborative research environments vis-a-vis lack of global database schema structures and standard data integration approaches. Annotation Graph model was used for representing and querying concepts, annotations and web documents. Its conceptual architecture was used to manage and perform retrieval operations based on concepts, annotations and documents.

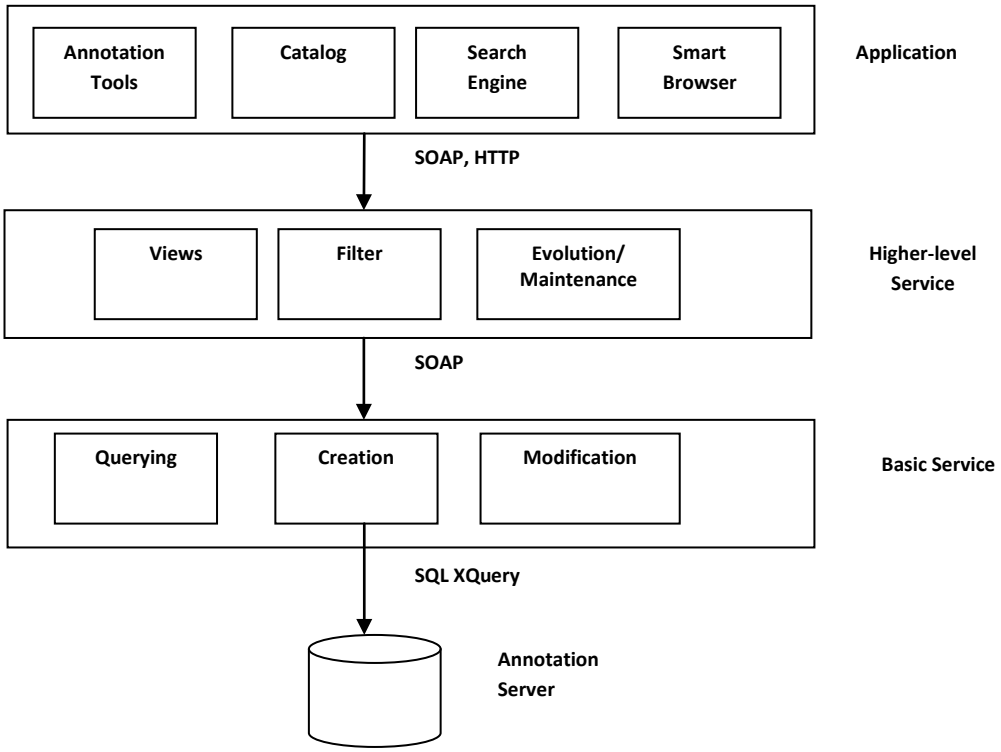

Figure 2.9: Conceptual Architecture of MAP

Source: (Bargeron et. al., 2001)

2.8.5 Triple-Note

The work of Yang et al. (2007) focused on the use of triple tagging scheme for creating semantic annotations in order to enrich web contents with machine readable metadata. Triple annotations, called triple tags consist of a set of human defined statements about web contents. The work provided formal description of triple tagging scheme for building triple tagging ontology using Semantic Web technologies. The scheme was implemented as a system called Triple-Note. A user of triple tagging system could annotate web documents for personal use or collaboratively with other users. The system provided functionalities such as browsing, querying and feedback functionalities. Figure 2.10 shows the adapted RDF graph of triple tagging ontology.

Triple-Note system architecture followed a client-server approach. The client side of the system was implemented as a Firefox extension and had an SVG-based triple tag browser. Users could

annotate, browse, explore, filter or query triple tags. The server side of the system on the other hand, had modules for extracting triple tags from sentences, query conversion from user queries to SPARQL queries and triple collection.

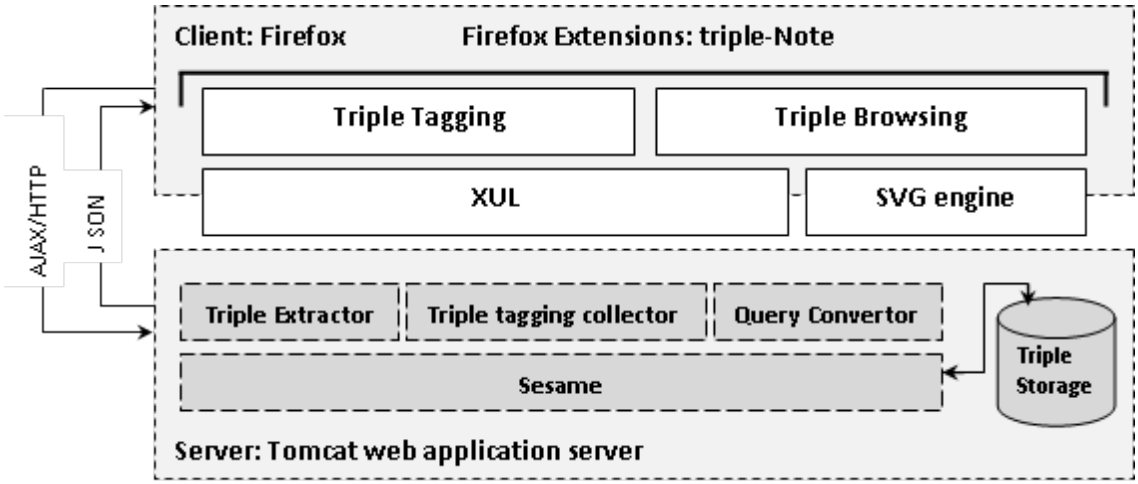

Figure 2.10: Adapted Triple-Note architecture

Source: (Yang et. al., 2007)

2.8.6 Pattern-based Annotation through Knowledge on the Web - PANKOW

PANKOW (Pattern-based Annotation through Knowledge on the Web), is a self annotating annotation model proposed in Cimiano et al. (2004) that uses unsupervised pattern based approach using linguistically motivated regular expressions to identify instance-concept relations on the Web, which serves as a source of big corpus. Eight patterns were linguistically defined for generating hypothesis phrases of which the first four were proposed and used by Hearst (1992). These are:

H1: <CONCEPT>s such as <INSTANCE>

H2: such <CONCEPT>s as <INSTANCE>

H3: <CONCEPT>s, (especially|including)<INSTANCE>

H4: <INSTANCE> (and|or) other <CONCEPT>s

DEFINITE1: the <INSTANCE><CONCEPT>

DEFINITE2: the <CONCEPT><INSTANCE>

APPOSITION: <INSTANCE>, a <CONCEPT>

COPULA: <INSTANCE> is a <CONCEPT>

Each identified instance was assigned to a concept that has maximal evidence deduced from Web statistics. PANKOW derived hypothesis phrases from identified noun phrases on the Web and candidate ontology using pre-defined linguistic patterns. The hypothesis phrases were used to query Google Application Programming Interface (Google API). Each phrase is categorised by assigning it to the concept having the highest document hits.

PANKOW was integrated into the CREAM framework by a plugin through which it has access to the ontology structure and to the document management system of the framework implementation.

The implementation of the plugin started with the scanning for the candidate proper nouns using a Part of Speech (POS) tagger. Heuristic that considered the intersection of the POS tagger categorization with the simple capitalized-words approach was used to get a higher precision for the candidate recognition.

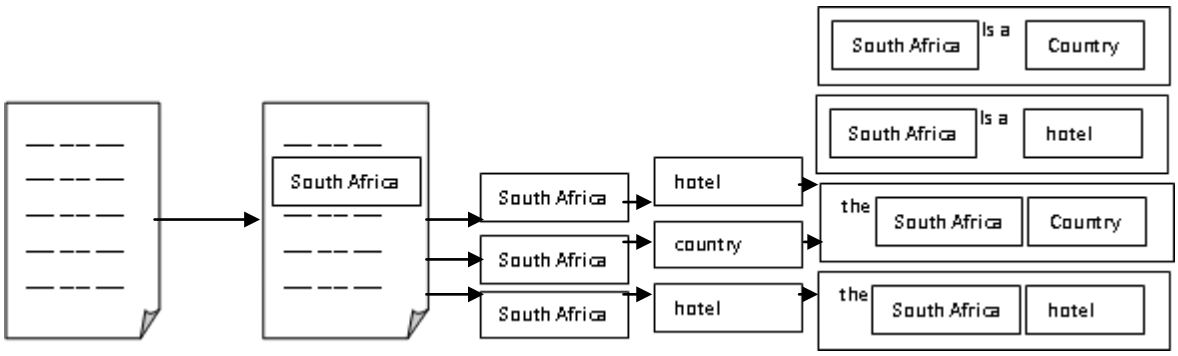

Figure 2.11: The adapted Process of PANKOV

Source: (Cimiano et. al., 2004)

While the use of ontology has played a significant role in standardizing semantic annotation in a collaborative environment, it however, limits the use of annotations to only agreed upon metadata in a specific domain. The users would be constrained to use the agreed terms only. It may inhibit understanding of how user's annotation could reflect user's personality. One of the goals of this study is to understand *why* a user creates annotation.

2.8.7 Annotation Model for Information Exchange

Annotation Model for Information Exchange (AMIE) is a model proposed in Robert (2007) for supporting information research in decision-making system in the domain of economic intelligence. The model consists of four main parts: annotator, document, annotation, and annotation context. While the model considers only the structure of annotation-object, it does not take into consideration the semantic and temporal annotation. Also, the term *context* as used in AMIE model refers to defining annotation implicitly or explicitly as well as stating the objectives of such annotation. It does not consider context in terms of user context and the use context of annotations. The "use context" refers to the domain-level use; what constitute a concept and its associated meaning in one domain may be different in another domain. The "user context" identifies the expertise of both the annotator and/or the use of such annotation. Annotations based on the instantiation of the concepts of these models could be used to annotate relevant documents and store the annotations possibly in an annotation repository. An annotator may wish to make an implicit annotation for private use. For example, a book reviewer may asterisk, underline and/or use question marks while reviewing a book. The meaning and the objectives of such annotation would be known only to the annotator.

2.8.8 ANNOTEA

Annotea is a web-based shared annotation modelled as a class of metadata based on RDF infrastructure. Annotations are stored in RDF databases and are accessible through an HTTP server. In Annotea, the RDF schema that defines the annotation properties is predefined. However, RDF provides support for extension of annotation types and/or relationship among the defined annotation types. The anchoring of annotation to document takes place within the client-Amaya web browser, and Apache running MYSQL database is used as a server. The concern here is that economic actors need to specify one of the annotation types to categorize the annotation they make. In a situation where any of the annotation types specified does not represent the objective of making annotation, one of the following could be done; either we find the closest annotation type to represent our intention or the RDF schema is extended to accommodate the structural changes. Table 2.2 compares some annotation models in terms of annotation representation.

Table 2.2: Comparison of some annotation models in terms of annotation representation

MODEL	APPROACH	ANCHORING	RETRIEVAL	Value/ Attribute-value
ANNOTEA (Kahan et al., 2003)	Uses RDF to define schema	Xlink/Xpointer technology	Xquery	Value
PAM (Yang et al., 04)	Ontology-enabled	Association & Spatial relations	Similarity clustering	Value
WISPA (Sannomiya et al., 2001)		Xlink/Xpointer technology	Xquery	Value
CDA (Gertz et al., 2001)	Graph Model	Xlink/Xpointer technology	Nodes & Edges traversal	Instance of concepts
ONTOMAT (Handschuh et al., 2002)	Ontological approach to define entities	Annotations added as instances of entities		Instance of entities
KIM (Kiryakov et al., 2003)	Ontological based (KIMO)			Proper noun
PANKOW (Cimiano et al., 2004)	Pattern-based, disa. by max. evidence			Named entities
AMIE (Robert, 2007)			EQUA^{2}TE	Value

Most of annotation models described in Table 2.2, viewed annotation as value-added information except in Gertz et al. (2001) where annotation is viewed as an instance of a concept. Therefore, there is a need to consider an annotation model that allows users to choose attributes of choice for concepts, and state or search for values for the chosen attributes. It is in this light that Attribute-Value-Pair (AVP) annotation model is being proposed. AVP representation allows objects to be represented with a set of attributes and values specified by the actors. It is believed that the ability of an actor to express his/her observation and/or contextualize document object as AVP annotation will improve significantly the effectiveness of such value-added information. This research study also believes that user query expressed as Attribute-Value Pair (AVP) will add semantic context to such query and this will greatly enhance identification of relevant information from large corpus such as World Wide Web. The proposed AVP search algorithm in this thesis will effectively assist user to retrieve contextualized information.

Attempt however, has been made in literature to extract attribute-value pairs from product description. The work of Probst(2006) focuses on extracting attributes and values from product description and associating the extracted attributes with the extracted values using Naïve Bayes combined with multi-view semi-supervised algorithm called *co-EM*. The attribute extracting system developed was made up of five modules: data collection; seed generation; attribute-value entity extraction; attribute-value relationship extraction; and user interaction. However, the goal of the work was to express each product as a set of attribute-value pairs for the purpose of comparing products. Wladmir et al. (2010) presented Wikipedia Attribute-Value Extractor (WAVE), for extracting attribute-value pairs from Wikipedia articles using a self-supervised approach for the purpose of generating infobox. Bakalov and Fuxman (2011) presented Structured Collective Attribute Discovery(SCAD) for extracting attributes from unstructured text. Their goal was to obtain structured data from unstructured text. These works do not consider the meaning of what the extracted attribute-value pairs represent. The major goal of this study is in representing annotation creation as attribute-value pair by users and in expressing user queries as attribute-value pair for information search. Before discussing the proposition of this study, it is pertinent to first discuss annotation relevance in EI process, as well as requirements the proposed annotation model should satisfy.

2.9 Annotation in Economic Intelligence Process

In this section, three areas of interest were identified: the use of annotation for eliciting knowledge for proper formulation of decision problem, annotation for assisting actors for retrieving relevant information once sources have been identified, and the use of annotation for interpreting analyzed information.

2.9.1 Annotation for Knowledge Elicitation

Decision-making process is often based on available strategic information. Such information, however, has to be sought, collected, processed with a view of eliciting knowledge relevant to the problem as well as representing both the collected information and elicited knowledge in a form that will facilitate information reuse. Decision based on such information alone may not be sufficient if the preferences of actors involved in problem resolution are not taken into consideration. Information of significance to one person may not be to another person depending on individual interest. Hence, availability of information is as important as the person who needs it or interprets it. Annotation-object is the expressed knowledge of annotator on the contextualized document. Knowledge elicitation is a process of obtaining, transforming and documenting information from identified information source such as human experts, and converting it to explicit knowledge. An actor's knowledge is elicited by capturing his/her interpretation from the document the actor has annotated. Such annotation may likely be based on his/her intellectual capacity, experience, cultural background and environment. Storing information on who annotates what and how personal characteristics influence annotations made could serve as a form of knowledge on the annotator. Annotation patterns of individual actor overtime could reveal interesting patterns that could assist in discovery of new knowledge about the actor.

2.9.2 Annotation in Information Retrieval

In Information Retrieval, the objective is to find information useful to user's need. The process entails the user expressing his/her need as query, the system maps the query to retrieved documents using keywords index search for example. In most cases, users have a long list of documents to look through; most of which may not be relevant to their needs. The needs of a user may vary depending on various factors. A major challenge in retrieving information automatically is that such retrieval is based on query submitted only. Personalized Information

Retrieval (PIR) is aimed at improving the retrieval process by taking into consideration the interests of individual user based on his/her history of past activity. A user may submit the same query at different times for different contexts. Using the history of past activity of a user for information retrieval may not meet the need of such user as his/her motive might be different. At any particular time, the motive behind submitting a query is only known to such user. User's query specifies *what*he/she is looking for. Documenting the *why*she/she needs such information as well as the domain context in form of annotation could prove very valuable in improving information retrieval.

2.9.3 Annotation as value-added information

A document often contains information author intends to pass across to readers. In most cases, such a document centres on a particular subject matter. Additional information that might not be known to the author or information the author might not have included in the document but proved valuable could be added in form of annotation. Marshall quoted Robert McCrum in (Marshall, 1998) on the account of annotations found in the books of Graham Greene's library. The quotation revealed that Graham added extra information on the margins, flyleaves and endpapers of those books. Such information may not have been provided in the books. The context could be totally different from the book context. Thus, such annotation made becomes value-added information. Information system contains pre-selected, sorted, and processed information for a particular use. Thus, information contents are structured according to targeted end use. Given such structured information, users can express information needs more precisely. They can specify values in relation to attributes of the information. However, users could be faced with the following challenges. A user needs to have a minimum knowledge of his/her needs by specifying at least a value of the stored attributes, since information search is still done by the principle of search by content. Problem arises where the user does not know what he/she is looking for or specify attribute that does not exist. Where a user specifies an attribute that does not exist, there would be a need for structural changes to accommodate new attributes. Such changes might be tedious, time consuming and costly to effect. Adding such information through annotation is a means of resolving such problem.

2.10 Requirements for Annotation Model

For effective use of annotation based on the three areas discussed in the preceding section, there is a need to design an effective annotation model. We are of the opinion that, in designing such an effective annotation model the following questions should be answered:

Who makes annotation? It is important to know who makes annotation and the possibility of how his/her personal characteristics influence the annotation made. Annotation-object is the expressed knowledge of annotator on the document or document object. Knowing the characteristics of the annotator is important. Annotation made by an annotator may be based on annotator's intellectual capacity, experience, cultural background and environment. Storing information on annotator and how his/her personal characteristics influence annotations made could serve as a form of knowledge on the annotator.

Annotation made concerns who? It may be created to have implicit or explicit meaning depending on targeted audience. Personal annotation could have implied meaning only to the annotator. The meaning is expected to be very explicit for public audience and agreed understanding of meaning among community of users.

Why the need for annotation? The motive of making annotation should be clear and expressly stated. The term *objective* has been used in Robert (2007) to designate the reason for making annotation. At any particular time, the motive of creating annotation is only known to such annotator.
It is therefore, necessary for an annotator to expressly state the reason for creating annotation.

When is annotation made? This concerns the date and time annotation is created. In a collaborative environment, it can be used to determine the most recent annotation made concerning a decision problem resolution.
What document has been annotated? Annotation is made on an existing document. Such document is referred to as annotated object. The object may be the document itself or its content. Existing annotation made is also regarded as a document that can be annotated.

How is annotation anchored to the annotated object? This is relevant in determining the relationship between annotation made and annotated object. Annotation created can be anchored

to the underlying document as inline annotation or as overlay depending on whether the actor has access right to modify the document or not. Considering the wealth of information that resides on the Web, this information is from different sources. Users may not be able to modify the structure of this information as contained in the web pages. Therefore, annotation model that does not take into consideration the structure of documents will prove more valuable.

In addition to providing answers to the above questions, the model should be able to satisfy the following functional requirements:

Structure: Ability of an economic intelligence actor to express his/her observation and/or contextualize document object of interest as attribute-value pair annotation could improve significantly the effectiveness of such valued-added information as opposed to adding annotation as atomic object only. It will provide a good basis for data restructuring, data mining, robust exploitation, knowledge elicitation among others. Since the needs of actors are evolving, it becomes practically impossible to anticipate all possible attributes at the design stage, hence, the need for flexible and robust means of adding attributes and values as the need of actors arise. Actors should have the possibilities of expressing their views without restriction. They should be able to choose the word, phrase or sentence that best suit what they intend to represent.

Communication: Annotation is being used to discuss topics of interest in different forumon the social web. In Wikipedia (n.d.), annotation is used for authoring documents. Tagging (annotation) is used in Delicious (n.d.) for categorizing web pages of interest by users. With Flickr (n.d.), group can jointly annotate pictures of interest. The objective is to have annotation as communicative acts among collaborators for sharing one another's perceptions in resolving a decision problem. Actors may share their perceptions of the decision problem or of any document through annotation in a collaborative environment. The conclusion arrived at about the object of deliberation (decision problem) could be added to the initial information source as annotation. Annotation could be used by economic actors to share one another's perception synchronously or asynchronously. With synchronous annotation, actors can communicate with one another in real time. It however, requires all participating actors to be online. With asynchronous annotation, the actors need not communicate in real time. Electronic mail or any other form could be used in the communicative acts. It however, requires a means of notifying the involving actors about the pending message as time may be of essence.

In EI process for example, the second stage of the process is transforming a decision problem into information search problem. Actors do communicate to clarify imbroglio in decision problem such as a concept, context or environment of the problem. Annotation could be used by watcher to add his/her interpretation to the document. Decision maker may validate watcher's interpretation through annotation. This is illustrated in Figure 2.12.

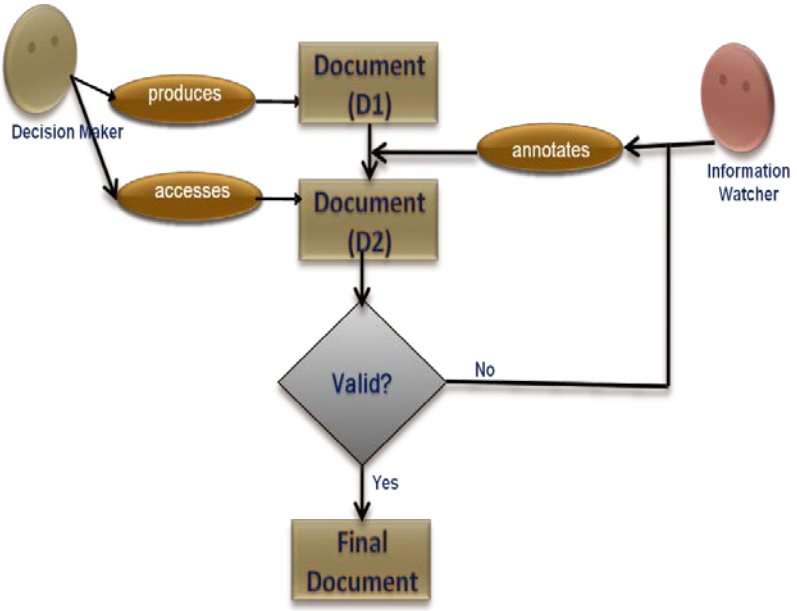

Figure 2.12: Diagram indicating structuring of initial demand between actors

Scalability: There should be the possibility of growth on annotation made. Annotations could become document for further annotations.

Reusable: Previous annotations made should be reusable. Actors should be able to adapt previous annotations to similar problems without necessarily modifying it. A modified annotation becomes another annotation.

Granularity: Actors should be able to add annotation to a document or document object of interest at any level of granularity. No restriction should be placed on what to annotate and where to place such annotation. There should be the possibility of adding annotations both at coarse-grain level and fine-grain level. For example, annotation might be added to document title, paragraph, sentence, phrase, text, word, image section etc.

Therefore, the design of annotation model for economic actors that can provide annotation capabilities for adding values to information is important as well as very necessary. The overall structure is illustrated in the Figure 2.13.

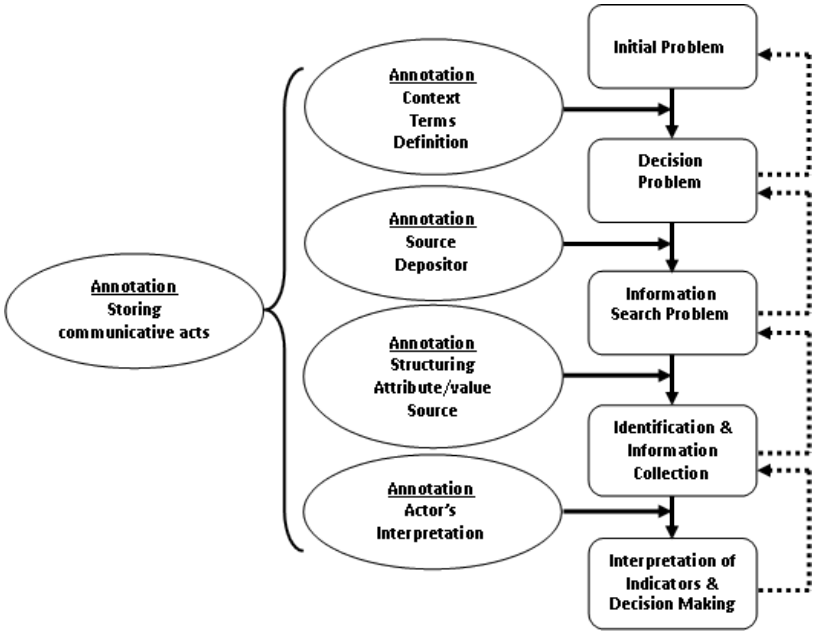

Figure 2.13: Annotation in Economic Intelligence

CHAPTER THREE

RESEARCH METHODOLOGY

3.0 Introduction

This Chapter discusses the research approaches used in answering the research questions stated in section 1.4. For the first research question: What annotation model could be developedthat will allow actors to load, annotate and store annotations taking into consideration the semantic context of annotators? **T**he study proposes Attribute-Value Pair (AVP) model for representing annotation. The AVP model is described in section 3.1 and the model is used to develop a prototype called Annotation Model and Tools for Economic Intelligence Actors (AMTEA). The general architecture of AMTEA is also discussed in this section. Section 3.3 describes annotation creation and document loading. For the second research question: How can this study realize the exploitation of the stored annotations and other relevant information that will enhance solving decision problem based on the context of the problem?The study uses the AMTEA system to discuss how this question is answered in section 3.4. The third question:What search algorithm is required that will enable actors to search for information of interest based on the actors' objective of search is answered in section 3.5.

3.1 Description of AVP Model

This study proposes Attribute-Value Pair (AVP) model for representing annotation. The model is divided into annotation_attribute and annotation_value. The annotation_value is the actual written expression of view of the user (called annotator) and the annotation_attribute component expresses the reason why the annotator has made the annotation. For example, if an actor finds a concept odour in a text, the actor may annotate the concept and state the objective of such annotation such as definition, cause, and effect, etc. The concept odour is referred to as annotation_value and definition, cause, effect are referred to as annotation_attribute. The annotation_attribute is for categorizing annotation_value(s). Let A_{avp} denote annotation

represented as AVP. A_{avp} from the ongoing discussion is a super set of annotation_attribute (attr) and annotation_value (val); i.e.

$$A_{avp} = \{(attr, val)\} \quad (3.1)$$

Two forms of annotations can be inferred: annotations made on the entire document (referred to as A_d) and annotations made on document-object (referred to as A_o). Examples of document-objects (d_o) are terms, phrase, sentences, etc. Therefore, the two forms of A_{avp} are A_d and A_o.

Based on the definition of annotation in Buckland (1991), this study differentiated annotation-object from annotation-process. Annotation discussed in the preceding paragraph is annotation-object. It is the actual annotation created by the actors. The process of creating annotation-object is referred to as annotation-process.

3.1.1. Formalism of AVP Annotation Model

The Resource Description Framework (RDF) used in this study, is a data model introduced by World Wide Web Consortium (W3C) for describing web resources. It is a part of the W3C's Semantic Web (SW) activity in which information expressed in a machine readable form could be used and exchanged both by human and machine agents. In information modelling, RDF describes resources as statements. Each statement is a triple consisting of subject, predicate and object. While the subject represents the resource being described, the predicate expresses a relationship between the subject and the object. The collection of RDF triples constitutes an RDF graph. The subjects and objects of the triples constitute the nodes of the graph while the predicates denote directed links between the nodes.

The AVP annotation modelled with RDF is formalized in the following definitions:

Definition 1

An AVP is RDF-like triple consisting of document being annotated (D), user-specified attribute (A), and annotation description as value (V). Therefore, $a \in AVP = \{d,t,v\}$ where $d \in D$, $t \in A$, and $v \in V$ are of type String. d is connected to v through t. V represents user meaning of D as contextualized by A and asserted by user U.

Definition 2

The set of all instances of AVP constitute annotation graph G. that is, G = $\{a_1, a_2, \ldots a_n\}$ such that $\forall a \in$ AVP, a = {d, t, v} with d and v representing the graph nodes and t as the label of the directed edge from d to v.

Definition 3

Annotation-process ($A_{process}$), is defined over four tuples as $A_{process}$ = (U, A_{avp}, $A_{anchor,}$ D) where D is the set of all documents, U is the set of annotators, and A_{avp} is the set of annotations expressed as AVP.

3.2 Annotation Schema

This section discusses annotation schema based on the identified annotation components. It explains the structure and interrelationship among the identified components which are annotator, annotation, document and anchor.

A user may access any of the stored documents. If he finds any of it relevant, may annotate the document or any section of the document using annotation component. For the annotation component, anno_access_mode property enables annotators to specify whether the annotation made is private, protected (for a community of users), or for public domain. User's interpretation is contained in annotation_value property categorized by the user stated annotation_attribute property. The annotation created, the anchor information and the information on the actor are stored in the annotation database. The study separates annotating the entire document from annotating parts of a document for the purpose of granularity. Anchoring an annotation to an entire document involves linking the annotation to the URI of the document. Annotation on a sectional part of a document requires the determination of both the positional and association relations between the annotation and the document object. Concept component represents the abstract form of the document object. This is necessary as annotators should be able to add values in form of annotation with or without prior knowledge of the underlining document design. Information stored in both annotation database and user database is a form of knowledge base. Knowledge on annotation pattern of an individual annotator or annotations made

concerning a decision problem can be extracted. The annotation schema is shown in the Figure 3.1. However, important attributes of each component are displayed in the diagram.

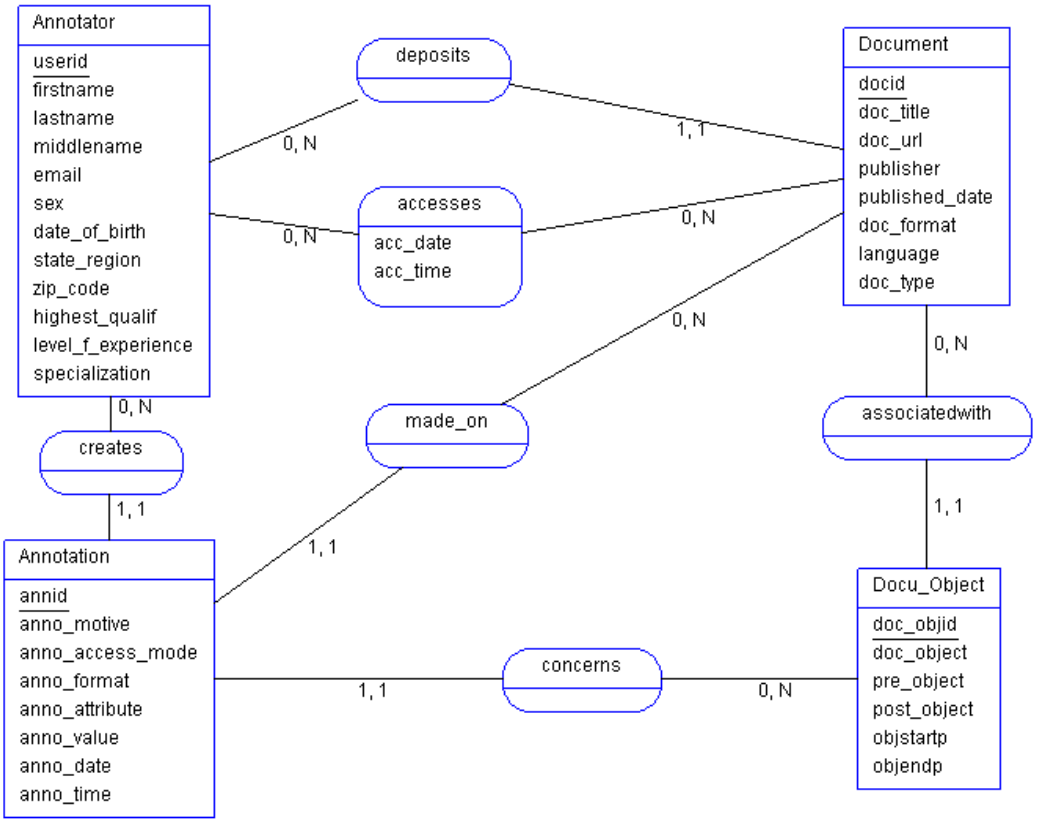

Figure 3.1: The Annotation Schema

3.2.1 Interpretation of the Annotation Schema

Actors in this context are regarded as annotators since the focus is on making annotations. Therefore, the term annotator shall henceforth be used to denote an Economic Actor.

3.2.1.1 Annotator and Document

An annotator wishing to perform annotation must have access to document he intends to annotates. Such document is retrieved from document repository. A document may be accessed by zero or more annotators. Where a document is accessed, the access date and access time are

noted. Likewise, an annotator may access zero or more documents. An annotator may deposit a document he finds interesting or he intends to annotate but not in the document repository into the repository. The schema constrains deposits of duplicate documents by ensuring that only one annotator can deposit a document. In a case of modified document, such document is regarded as another document. The Figure 3.2 below illustrates this argument.

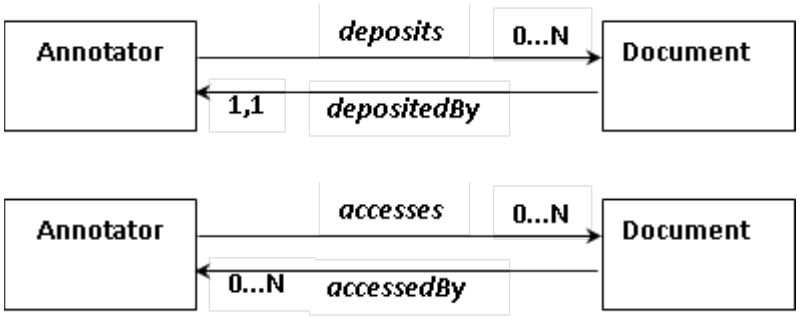

Figure 3.2: Relationship between Annotator and Document

The Figure 3.2 implies that:

An **Annotator***deposits* zero or more **Documents.**

Conversely, a **Document** is *deposited by* one and only one **Annotator**.

An **Annotator***accesses* zero or more **Documents.**

Conversely, a **Document** is *accessed by* zero or more **Annotators**.

3.2.1.2 Annotator and Annotation

An annotator may create zero or more annotations on retrieved document. However, an annotation is created by one and only one annotator. Similar annotations may be created by different annotators or even, by the same annotator. In this case, the annotations are different with each having a unique identifier.

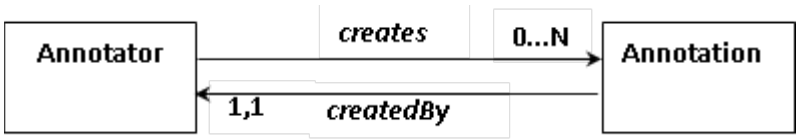

Figure 3.3: Relationship between Annotator and Annotation

The Figure 3.3 implies that:

An **Annotator***creates* zero or more **Annotation.**

Conversely, an **Annotation** is *created by* one and only one **Annotator**.

3.2.1.3 Annotation and Document

An annotation concerns one and only one document but a document may contain zero or more annotations. Similar annotations may be made on different documents. In this case, the annotations are stored differently with each having a unique identifier.

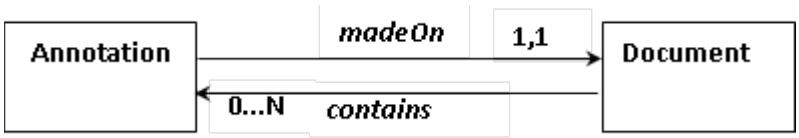

Figure 3.4: Relationship between Annotation and Document

The Figure 3.4 implies that:

An **Annotation** is *made on* one and only one **Document.**

Conversely, a **Document***contains* zero or more **Annotation**.

3.2.1.4 Annotation and DocumentObject

An annotator may annotate an object in the document. Each object is seen differently though they may be similar or the same. Therefore, annotation concerns one and only one object. Conversely, an object may have zero or more annotations.

Figure 3.5: Relationship between Annotation and DocumentObject

The Figure 3.5 implies that:

An **Annotation***concerns* one and only one **DocumentObject.**

Conversely, a **DocumentObject** can *have* zero or more **Annotation**.

3.2.1.5 Document and DocumentObject

A document has one or more objects but an object is associated with one and only one document. This constraint is imposed in order to uniquely identify each annotation made on each object. In this case, the annotations are stored differently with each having a unique identifier.

Figure 3.6: Relationship between Document and DocumentObject

The figure 3.6 implies that:

A **Document***can have* one or more **DocumentObject.**

Conversely, a **DocumentObject** is *associated with* one and only one **Annotation**.

With these formal definitions of AVP and the discussed annotation schema, the study used the AVP model to develop a prototype called Annotation Model and Tools for Economic Intelligence Actors (AMTEA). The general architecture of AMTEA is shown in Figure 3.7 followed by brief discussion of its functional components.

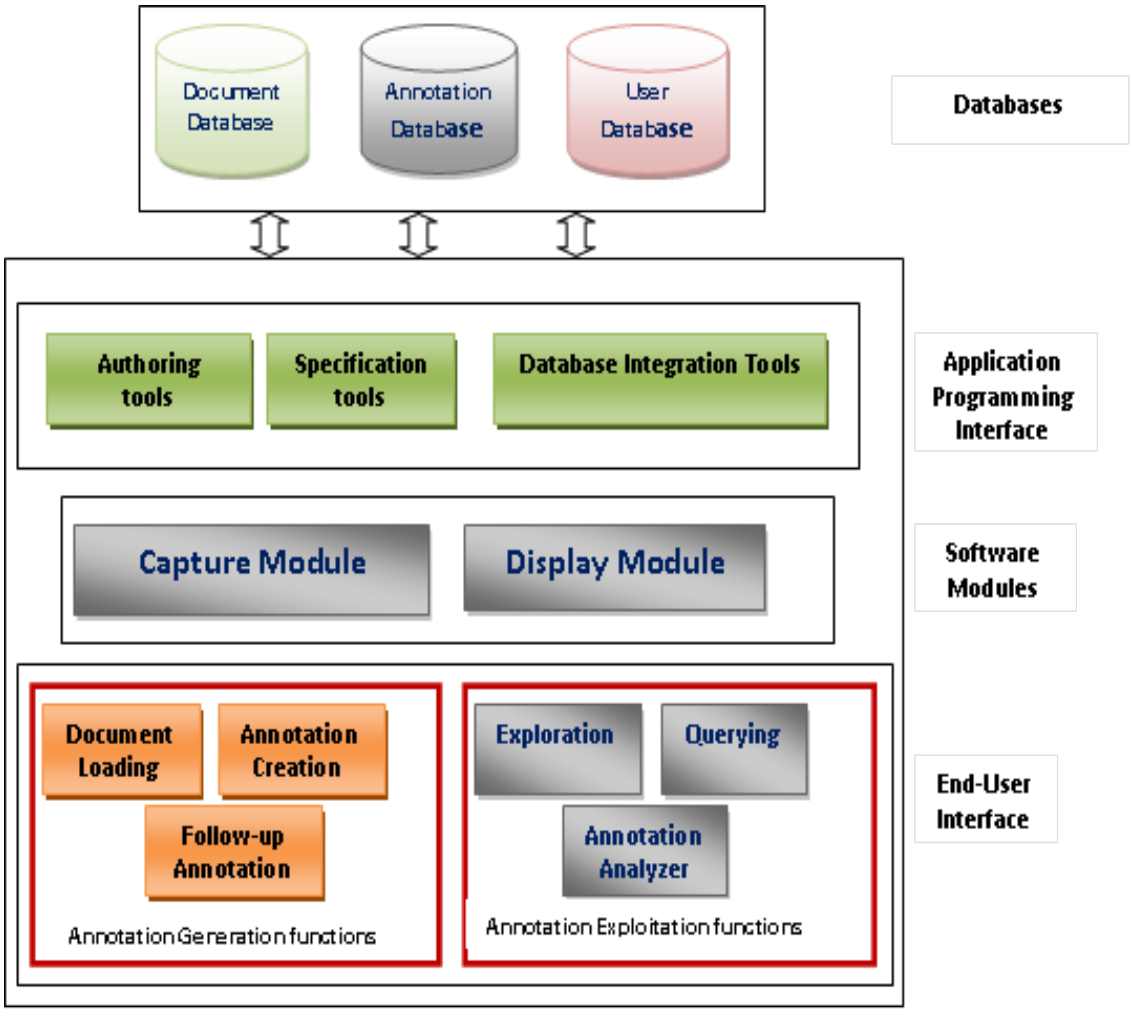

Figure 3.7: AMTEA Architecture

End-user Interface: This layer is divided into two functional components- annotation creation and annotation exploitation. Annotation generation components involve document uploading, creation of new annotation or follow-up annotation. Annotation, document or user's profile exploration through hyperlinks, querying and annotation analyser are contained in the exploitation components.

Software modules: These are modules for capturing annotation and display of stored annotations. The capture module handles the login, registration of new users, deposits of new documents and annotations created. Display module handles the necessary views of annotation, document and user's profile.

Application Programming Interface: This layer contains other APIs such as Google API for searching, Thesaurus API for controlled vocabulary in choosing attributes and other tools of relevance.

Databases: It contains relevant databases. In the proposed system, document database, annotator database and annotation database are identified.

Figure 3.8 shows the use case diagram for the AMTEA and Figure 3.9 shows activity diagram for annotation. They are modeled using Unified Modeling Language, UML.

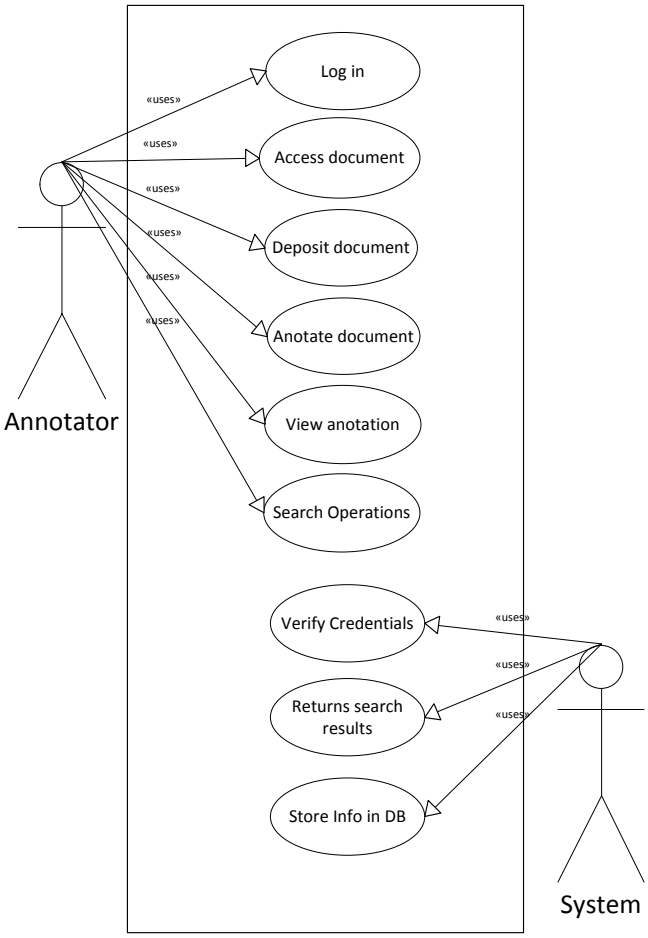

Figure 3.8: Use case diagram of AMTEA prototype.

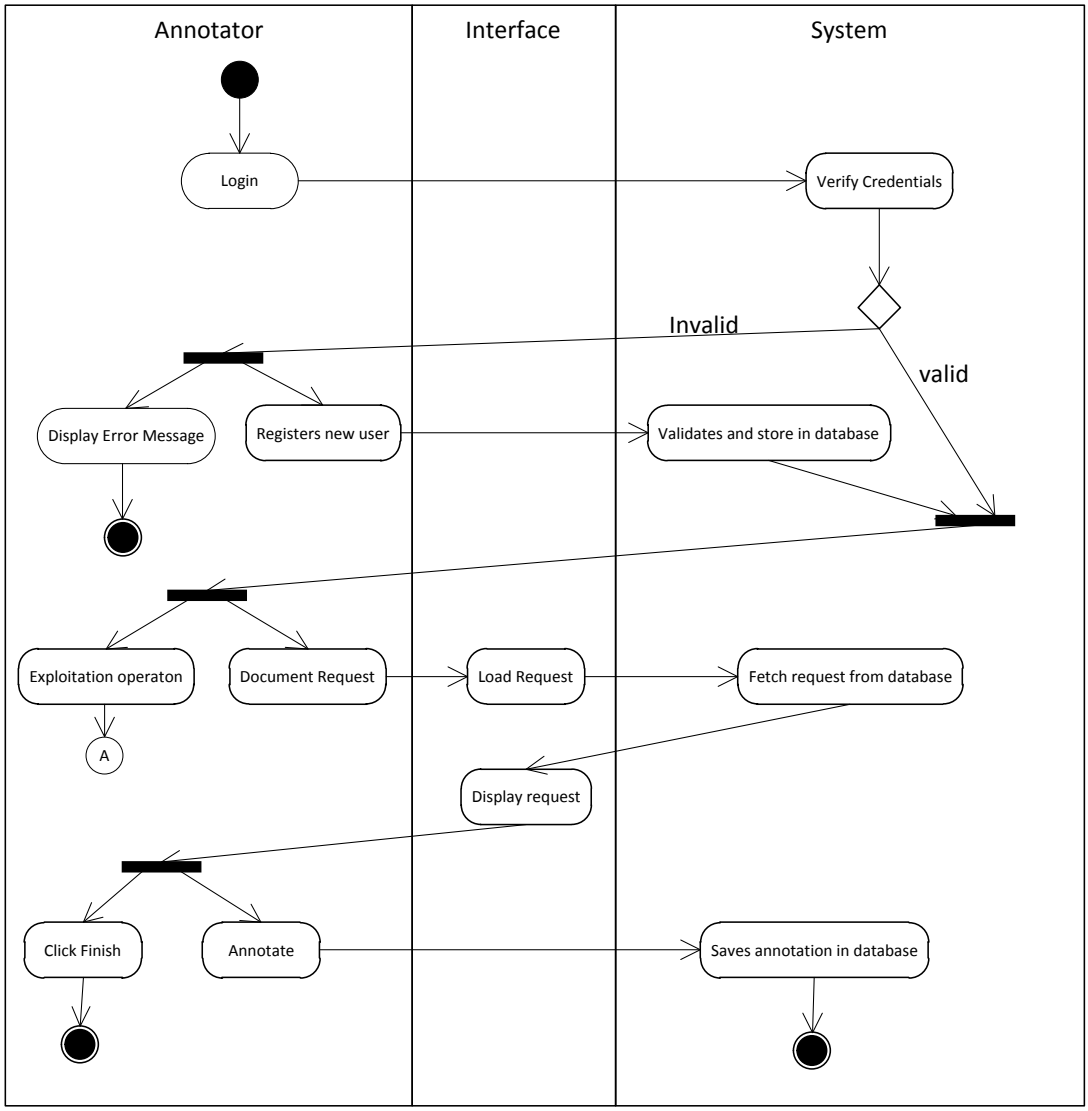

Figure 3.9: Activity diagram of AMTEA prototype showing annotation activity only.

Based on the functionalities of the architecture, the following sub-sections discuss annotation creation, annotation exploitation and AVP search methodologies.

3.3 Annotation Creation

This study refers to annotation creation as value-added information written and attached to a document or any part of a document's content (which will henceforth be called document-object)

by a user. The capture module in the architecture is used to create annotation. The module comprises of document loading, annotation creation and follow-up annotation described below.

Document uploading: The term document in this context refers to launching of web pages or uploading a saved document. In line with the definition of annotation, document to be annotated must have existed; hence, the annotator has the capability of browsing, launching or uploading documents from any source into the system.

Annotation creation: This sub component enables actors to select or highlight any part of opened document for annotation, annotates the selection with AVP approach and saves the annotation either into the local disk or annotation server. The annotation-object created could be a new annotation or extension of an annotation previously made on the same object.

Follow-up annotation: This refers to annotations made on the existing annotation. It enables actors to use annotation as communicative acts in concept clarification. Clarity and proper understanding of a decision problem among the actors is germane to the decision problem's resolution. Keeping the history of annotations in respect to a particular problem is very relevant.

3.4 Annotation Exploitation

Annotation Exploitation in this context covers exploration and querying of existing annotations. In annotation *exploration,* the user can explore stored information. Let A, D and P represent the sets of all annotations, annotated-documents, and annotators respectively. An actor can explore:

- All annotations (A)
- All annotated-documents (D)
- All annotators (P)
- All annotations made by a particular annotator in all annotated-documents ($A \cup D \cup p \in P$)
- All annotations contained in a particular document by all annotators ($A \cup P \cup d \in D$)

In *exploitation*, users may want to further refine their exploration to some specific information by querying the system. Querying operation may involve homogenous set of information or union of the sets of information. For instance, the user's query operation may be on annotation, annotator or document entities. The operation may also be union of these entities.

Exploitation function allows the user to select zero or more attributes for querying. It enables the user to select zero or more attributes for querying. Where no attribute is selected, the system assumes that the user does not know what he/she is looking for. It therefore, returns the contents of the system storage for possible exploration. It gives the user the minimum knowledge of what can be exploited. In a situation where the user wants to exploit the system by submitting one or more attributes, the query module is designed in such a way that any combination of attributes can be chosen. The function of annotation analyzer is to examine the chosen attributes, determine whether there exists any relationship among the attributes, and return results of such combination. Where no relationship exists among the attributes, the system returns the contents of each table. The **Display module** in the architecture is used for exploiting annotation. The module comprises of exploration, querying and annotation analyzer described below.

Figure 3.10 shows the activity diagram for query operations. Figures 3.11-3.13 show various sequence diagrams for the operation.

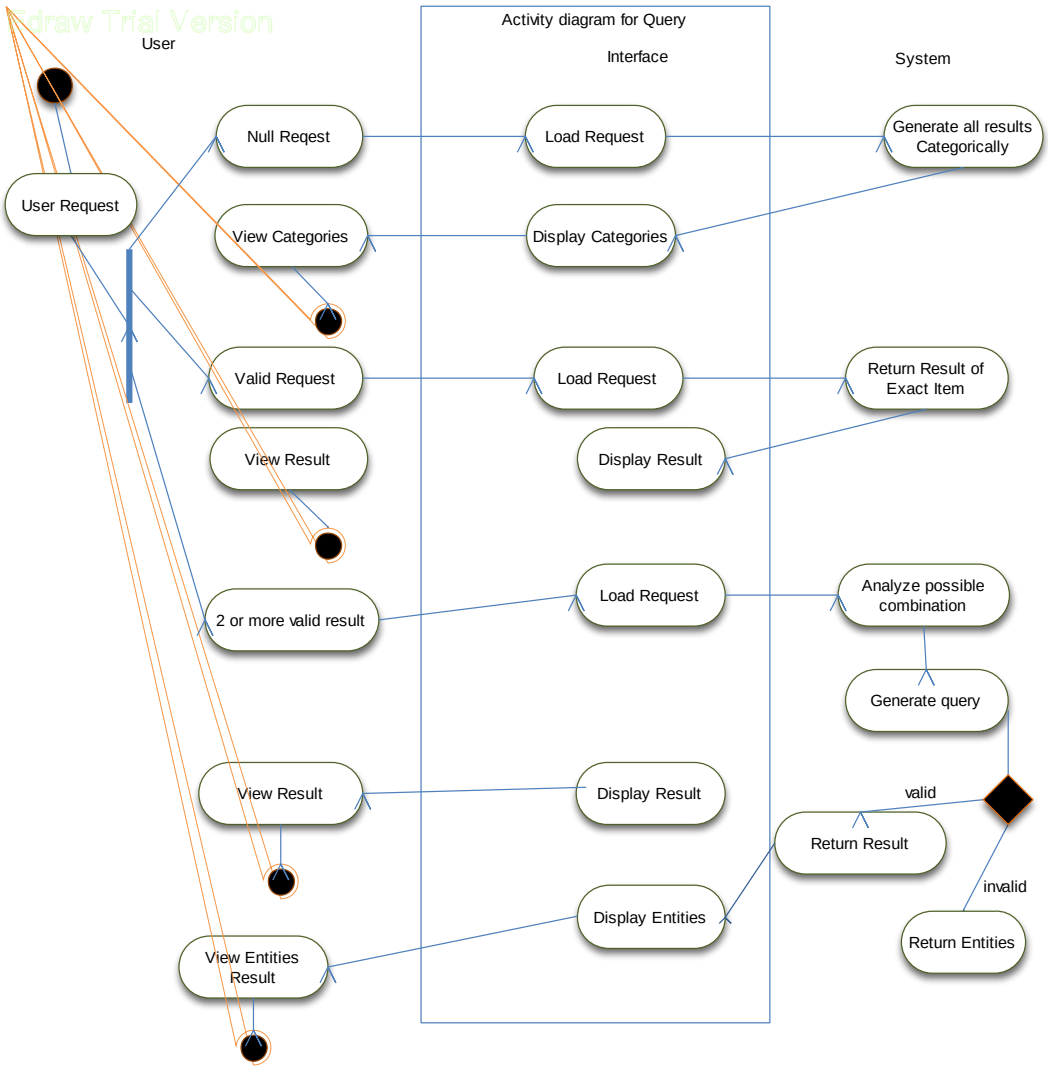

Figure 3.10: Activity diagram for query operations

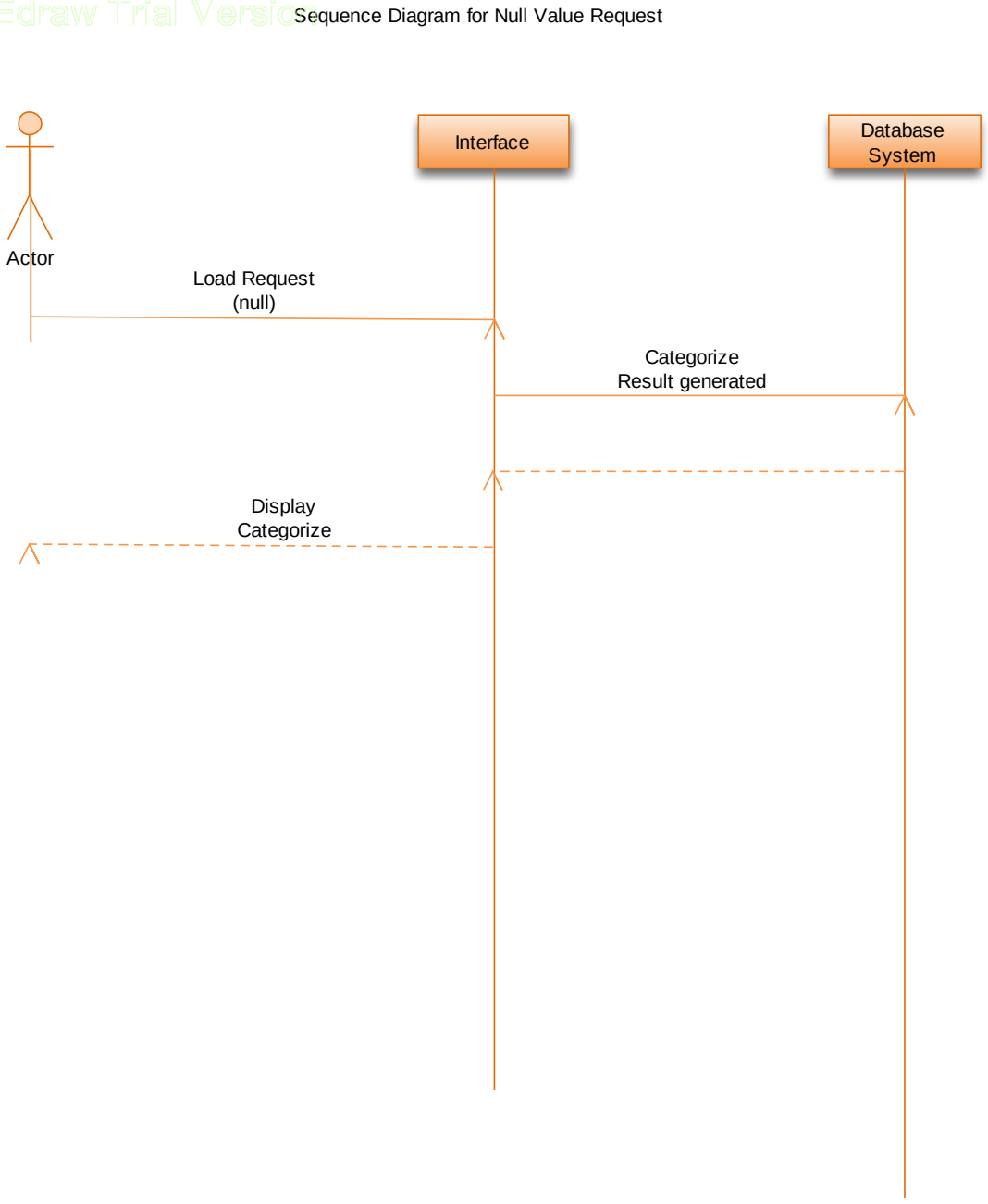

Figure 3.11: Sequence diagram for null value request

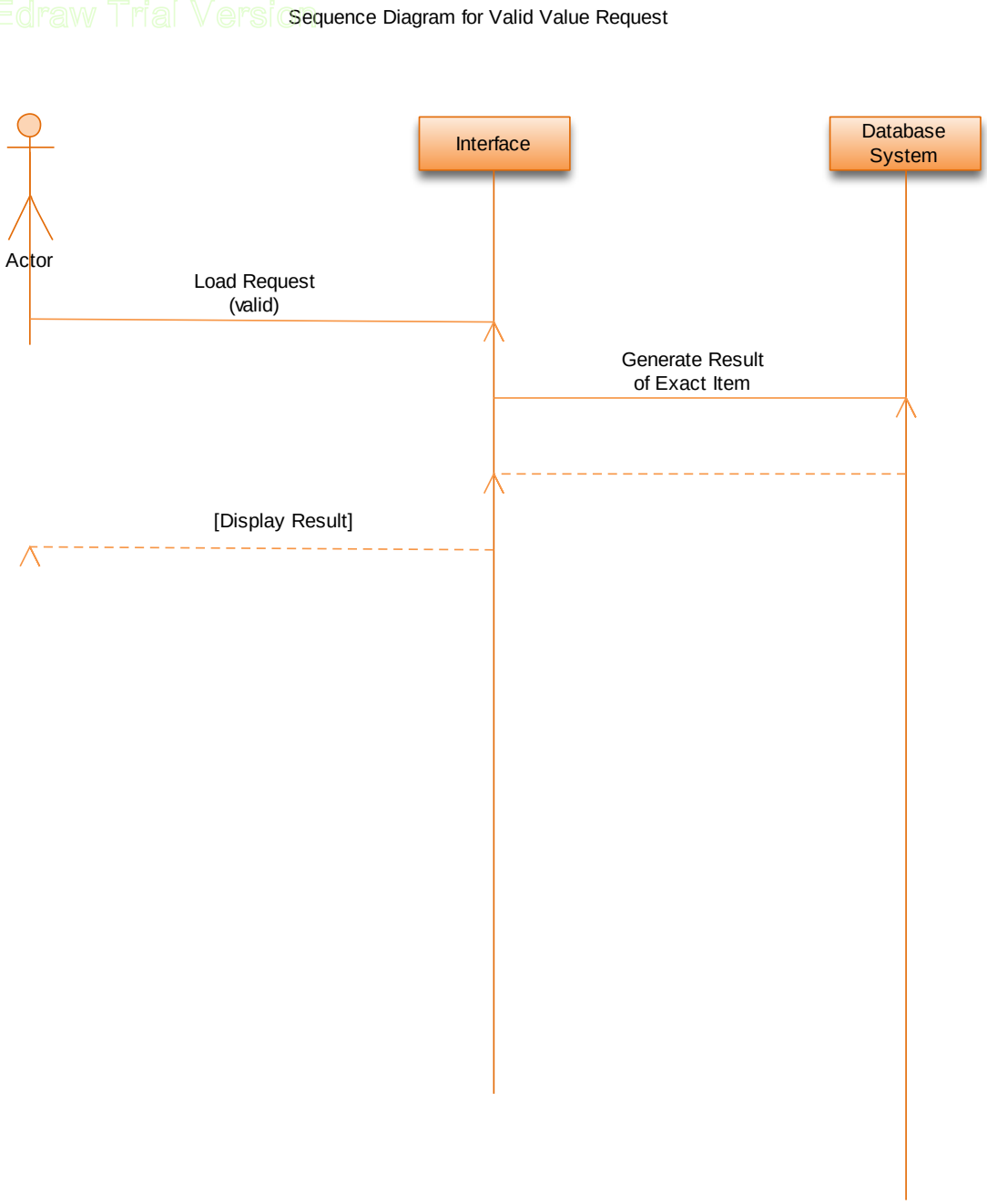

Figure 3.12: Sequence diagram for valid request

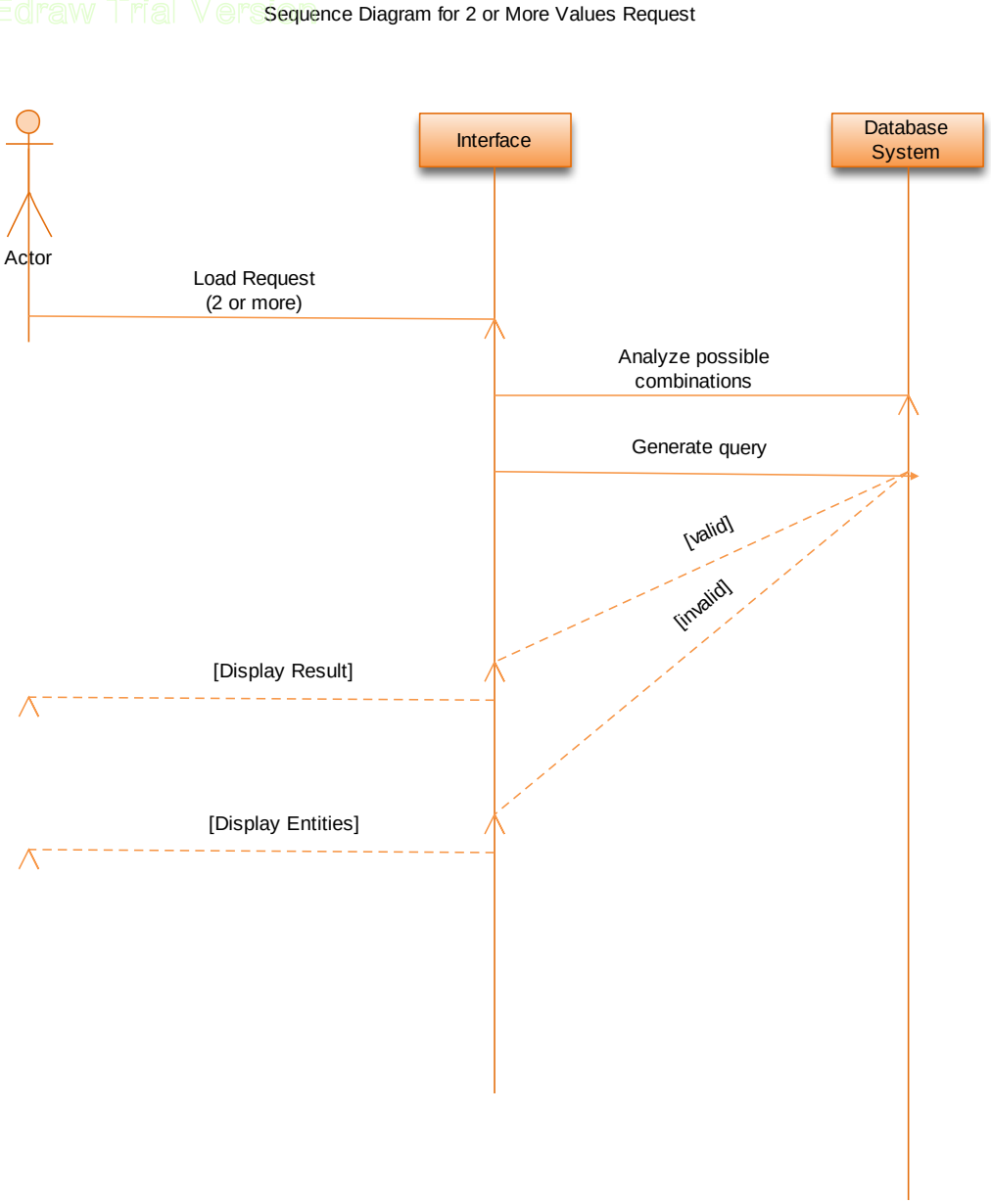

Figure 3.13: Sequence diagram for one or more valid value request

3.5 AVP Search

Information search is one of the major components of EI process. Search operation from large corpus such as World Wide Web (WWW) could be very challenging. A search technique that would allow actors to search for information of interest based on the objective of search will be

very desirable. This study therefore, considers a semi-automatic annotation solution called AVP search algorithm.

The AVP search algorithm allows the user to search for information of interest based on the semantic context of the user. With AVP search, the user expresses his/her query as attribute-value pair. The search algorithm tries to detect the actor's specified attribute from text and the associated value based on the context of the attribute. The attribute component is used for categorizing retrieved documents based on the user's objective of search. AVP search uses regular expression patterns to detect and retrieve values. The stages involved are summarily described thus:

- The user expresses his query as attribute-value pair. For example, "cause of body odour" is expressed as "cause" as attribute and "body odour" as value. This is called user defined attribute-value pair.

- The system retrieves similar attributes from annotation database if any, as well as retrieves synonyms of the attribute using a Big Huge Thesaurus (BHT) Application Program Interface (API) which is a plug-in to the system. The BHT API takes as input the attribute supplied by the user and generates the corresponding synonyms. Since the BHT API is a plug-in, the system does not bother on how the synonyms are generated. The system splits the synonyms from BHT API into noun and verb forms using JavaScript Object Notation (JSON) format. The result is presented to the user. The need for this process is to suggest to the user similar terms that might be used in web content describing the chosen attribute. The user is however, not constraint to choose any of the suggested synonyms.The section of php code below will generate the synonyms for the attribute "cause":

  $url = "http://words.bighugelabs.com/api/2/1b3093a931c02274fc86d862c333088e/cause/json";

  The resulting synonyms are:

  noun:{"syn":["reason","grounds","campaign","crusade","drive", "movement", "effort", "causal agent","causal agency","lawsuit","suit","case","causa","inception","justification","legal

proceeding","origin","origination","physical
entity","proceeding","proceedings","venture"]},

verb:{"syn":["do","make","induce","stimulate","have","get","create"]}

- System computes semantic relation of user query and generates semantic patterns from the query and synonyms of the attribute. The system does this by first stemming (a process by which a word is reduced to itsstem or root form) the attribute and then by constructing phrases from the chosen stemmed attributes (including zero or more synonyms that might have been chosen) and the specified value from the user-defined attribute-value pair.

  <ATTRIBUTE>s of<VALUE>
  <VALUE>is <stemmed (ATTRIBUTE)> by
  <VALUE>is <stemmed (ATTRIBUTE)> as

  Using the example of “cause of body odour”, the system generates the following patterns: “cause of body odour”, “body odour is caused by”, “body odour is stimulated by”, “reason for body odour”, “body odour is induced by”, etc.
- System queries search engine API (e.g. Google API) of choice with user selected patterns
- Web crawler crawls the Web pages using the URL links extracted from JSON results returned by Google API.
- System retrieves documents ranked in order of relevance based on documents with greater number of matched patterns.

Figure 3.14 and Figure 3.15 show the Activity diagram and the Sequence diagram for the AVP search respectively.

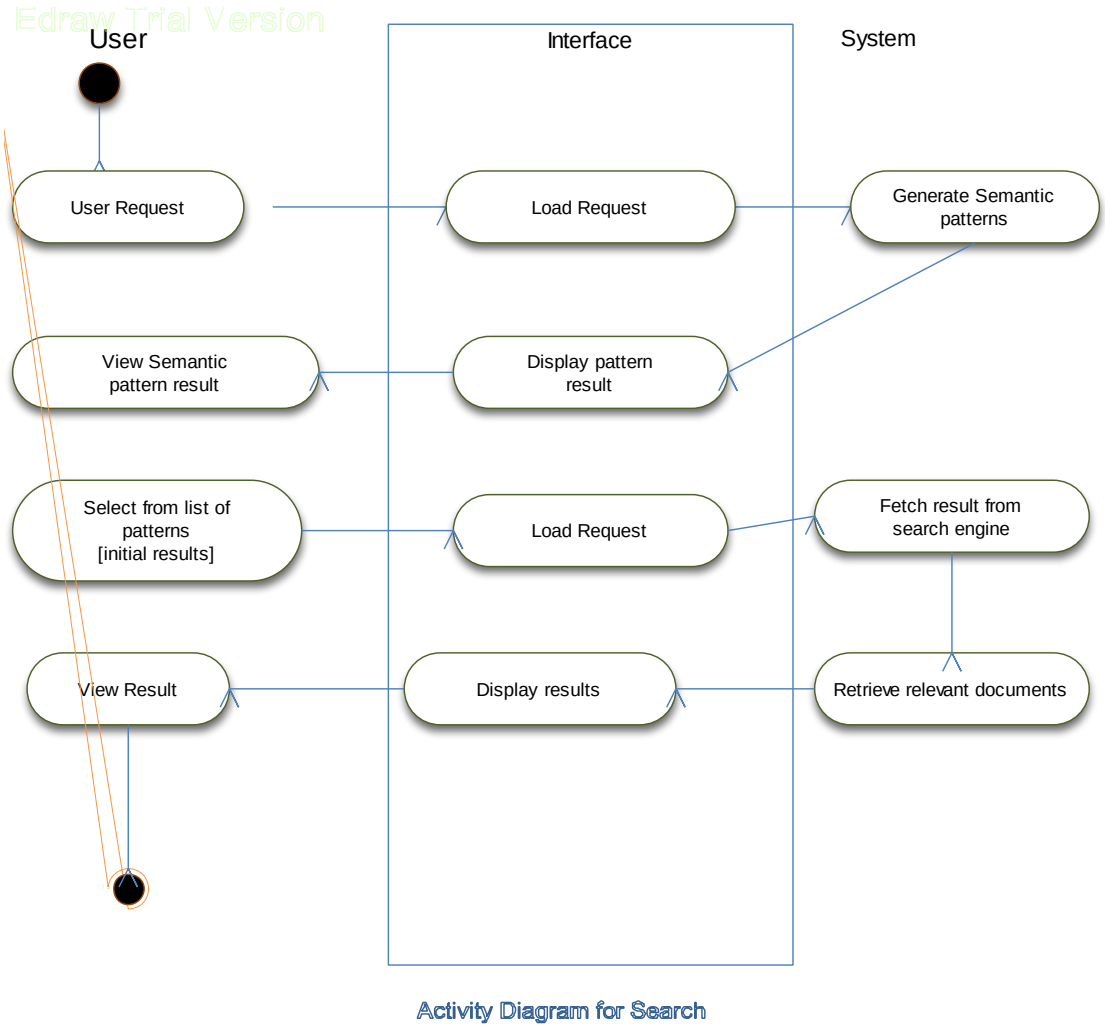

Figure 3.14: Activity diagram for AVP Search

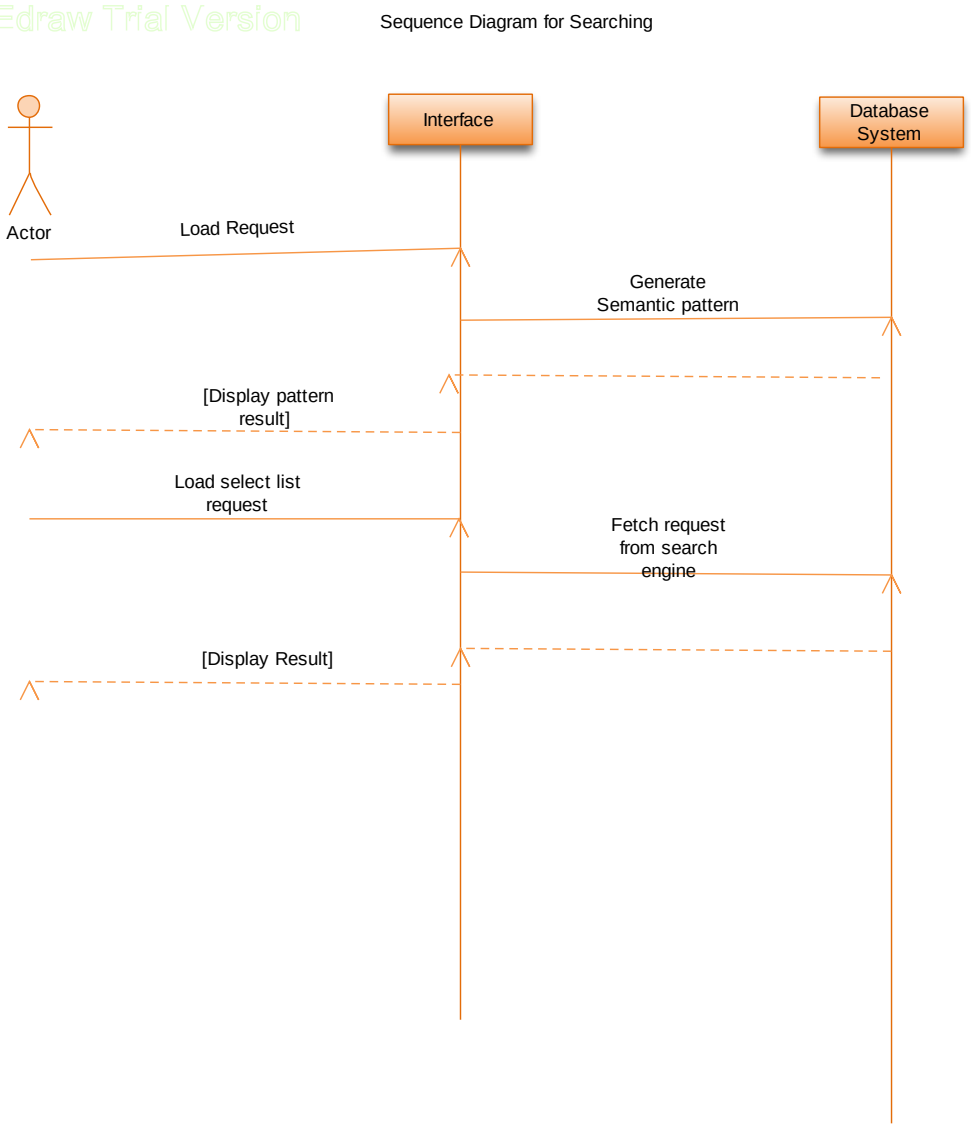

Figure 3.15: sequence diagram for AVP Search

3.6 Experimental Design of AMTEA Using AVP Annotation Model

The implementation of the AMTEA architecture is hereby described. The architecture is used to develop an annotation system called AMTEA Annotation system. The client-side of AMTEA Annotation system as shown in Figure 3.16 is divided into four segments named A1, A2, A3 and A4.

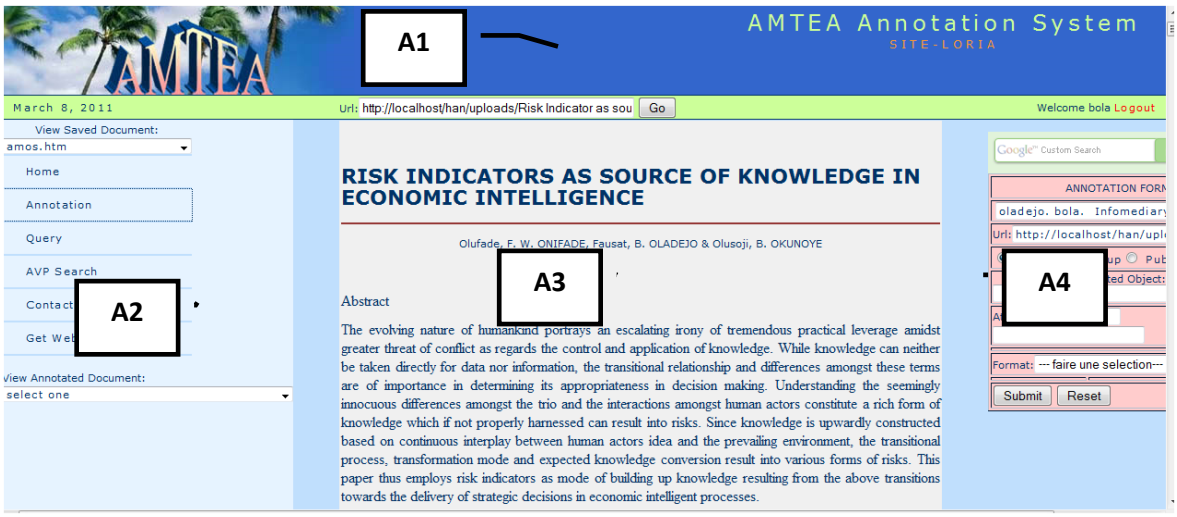

Figure 3.16: AMTEA Annotation System

Segment A1

This contains the bar for entering URL address, the login module, and the date. In order to perform annotation creation, a user needs to login as an existing user or registers as a new user. The relevance of login module is to retrieve the explicit information of the person using the system. When such information is not available, such user is requested to register. The user information that is captured is stored in Annotator database. The information is used for determining the influence of characteristics of annotator on the annotation made.

Figure 3.17 and Figure 3.18 show the login and registration forms.

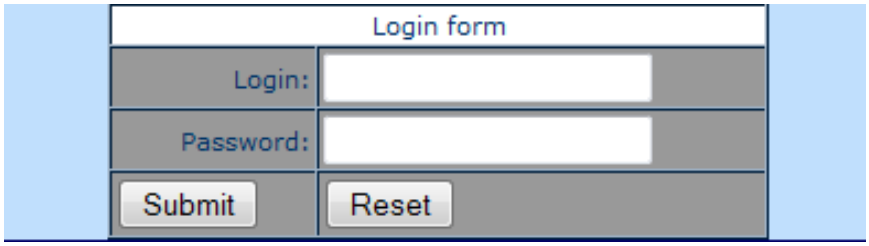

Figure 3.17: Login form for existing user

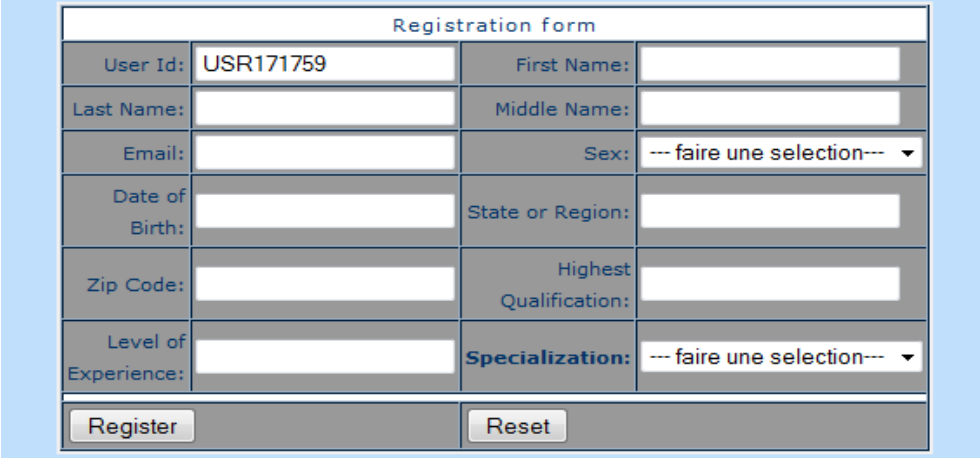

Figure 3.18: Registration form for new user

Segment A2

The functional menus are found in this segment. These include Annotation, Query, AVP search, View annotated Document, View Saved Document, and Get Webpage.To create a new annotation or make a follow-up annotation, the user, hitherto called an annotator, first uploads the document he/she intends to annotate into Segment A3, highlights the text to be annotated, and then clicks on the annotation menu. The system captures and displays on the annotation form the names of the annotator, URL of document being annotated, and the highlighted text. The annotator only needs to specify the attribute and value for the highlighted text. Also, he/shespecifies the access mode of the annotation- private, group or public. Figure 3.19 shows an example of displayed annotation form.

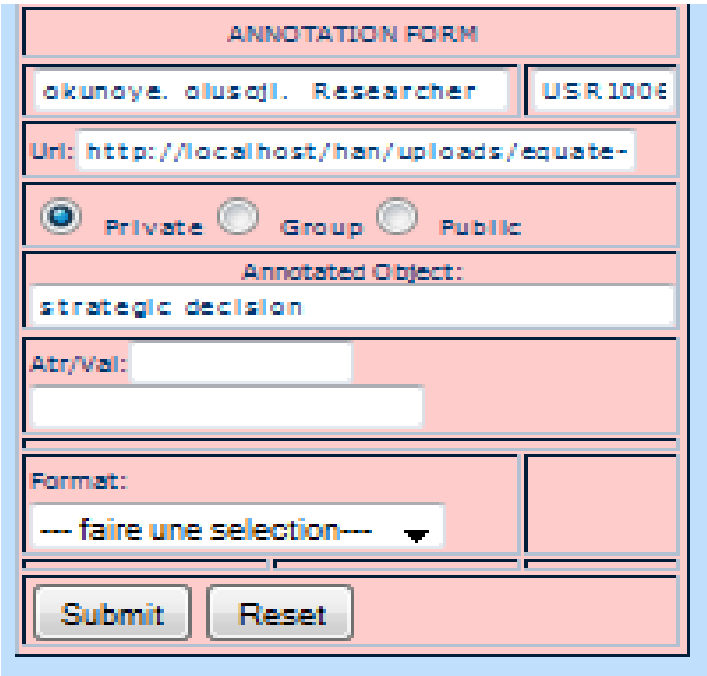

Figure 3.19: Annotation form

In the query menu, users can explore and exploit the system storage through querying. The query module is implemented using the EQuA^{2}Te architecture (David& Thiery, 2003). It enables the user to select zero or more attributes for querying. Where no attribute is selected, the system assumes that the user does not know what he/she is looking for. It therefore, returns the contents of the system storage for possible exploration. It gives the user the minimum knowledge of what can be exploited. In a situation where the user wants to exploit the system by submitting one or more attributes, the query module is designed in such a way that any combination attributes can be chosen. The function of annotation analyzer is to examine the chosen attributes, determine whether there exists any relationship among the attributes, and return results of such combination. Where no relationship exists among the attributes, the system returns the contents of each table.

The view annotated document menu contains list of saved documents that have been annotated. Whenever any of such documents is loaded into Segment A3, annotations made on the document as well as the names of the annotators is listed in segment A4. If a user clicks on any of the listed annotations, the system displays the positional link of the annotation on the document. A mouse

over the annotation link will reveal all stored information about the annotation. Figure 3.20 shows a displayed annotated document, list of annotation made and the display of a chosen annotation link.

Figure 3.20: Annotated document showing annotation made

The AVP search algorithm allows user to search for information of interest based on the semantic context of the user. With AVP search, the user expresses his/her query as attribute-value pair. The attribute component is used for categorizing retrieved documents based on the user's objective of search. See section 3.5 for the description of the stages involved.

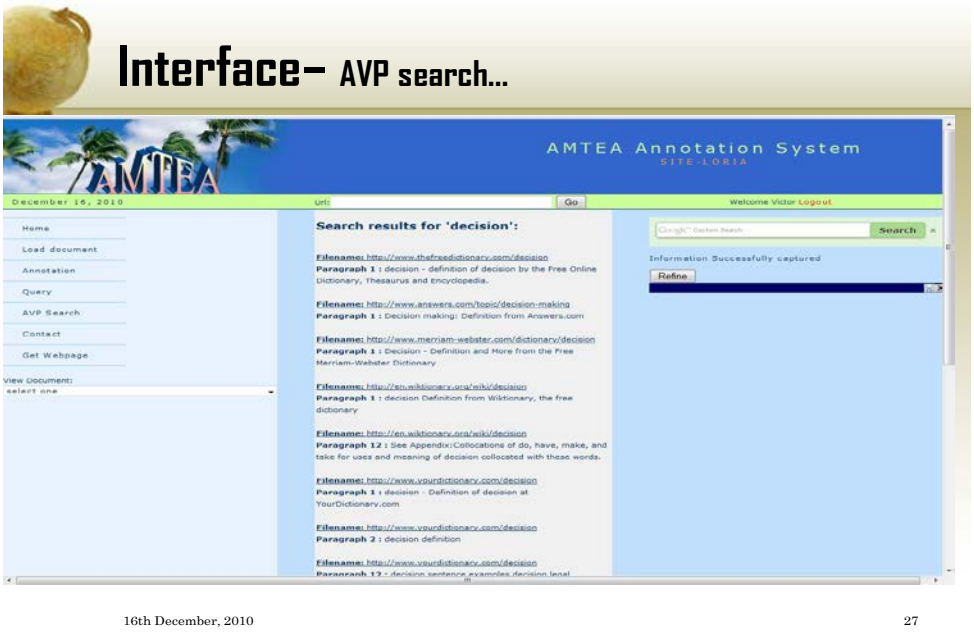

Figure 3.21: AVP Search displaying result

Segment A3

This is the main annotating pane. The pane holds documents to be annotated by users. It is also the pane where annotated documents and the associated annotations are displayed.

Segment A4

The segment holds list of annotations for a given loaded annotated documents. It also holds a user launched web page for browsing and navigation. However, if a user intends annotating the web page, he/she would need to click on **get webpage module** that will load the page into segment A3. Results returned when the system is queried are also displayed in this segment.

A decision problem often starts from a decision maker who gives initial demand as a document. This is stored in the document database. The watcher retrieves the document through annotation tool provided. He/she then calls annotation tool to add his/her annotation to the document. User's interpretation is contained in *attribute and value* properties. Anchoring an annotation to an entire document involves linking the annotation to the Uniform Resource Identifier of the document. Annotations made are stored in annotation database. Figure 3.22 shows the screen shots of the display of annotation tool in AMTEA environment. In Figure 3.23, annotations made on the displayed document is indicated as A, position of the annotations and the actual annotation

indicated as B and C respectively are shown. EI actors can query stored documents, annotations, annotations on a particular document, annotations made by a particular actor.

Figure 3.22: Annotation tool in AMTEA Environment

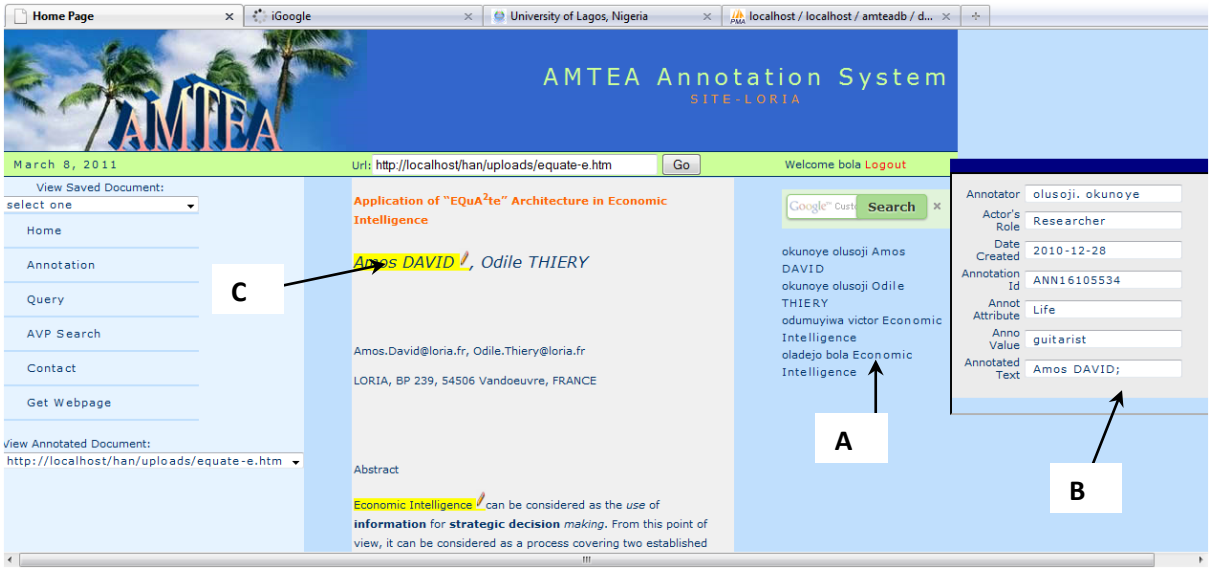

Figure 3.23: Annotation creation and exploitation in AMTEA Environment

3.7 Implementation Scenarios

In order to test or use the AMTEA system for annotation creation, exploitation and search, two scenarios are considered.

3.7.1 Scenario 1

The idea behind the use of AVP search algorithm is not only to annotate candidate attributes and their corresponding concepts alone, but also to search for possible values associated with such concepts based on the identified attributes by users. For instance, a user who wishes to search for the concept “odour” might likely get better results if he/she specifies the attribute of such concept when searching from a large corpus like Web. A system that considers attribute-concept pair pattern instead of search by keywords might likely get better results. AVP search algorithm implemented in AMTEA uses the latter method. A search for causes of body odour scenario is used as a test case. The concept is “body odour” and the exact attribute is “causes”. Some students (human agents) were asked to annotate “causes of body odour” from the first fifty links returned by Google API in order to determine the relevance of each link.

The interest is in getting relevant answers (termed values in this study) to the phrase “causes of body odour”. Similar phrases such as “body odour is caused by”, “body odour is the smell caused by”, “causes for change in body odour”, “body odour is due to”, “smell caused by”, “Causes of Changes in Body Odour”, “cause of sweating and body odour”, etc., were identified. The students discovered repeating links. By counting each link once in the cause of annotation, thirty-three links were found to be unique. The unique links are shown in Table 3.1. The column marked “X” symbolizes the links on which AMTEA system returned results and the column marked “Y” symbolizes the links on which human annotators marked relevant. The “N” mark indicates that the links were not found to be relevant with regards to the query.

Table 3.1: List of first 50 links returned by Google API for the query "causes of body odour"

Rank	HA*	AMTEA	URL
1.01	Y	X	http://www.netdoctor.co.uk/hilaryjones/embarrassingprobs/bodyodour.htm
1.02	Y	X	http://www.seasonsindia.com/healthfitness/bodyodour_sea.htm
1.03	N	N	http://www.medicalnewstoday.com/articles/173478.php
1.04	Y	X	http://en.wikipedia.org/wiki/Body_odor
1.05	Y	X	http://www.bbc.co.uk/health/physical_health/conditions/bodyodour2.shtml
1.06	Y	X	http://remedies.iloveindia.com/symptoms-body-odor.html
1.07	Y	X	http://www.naturalnews.com/004417.html
1.08	Y	X	http://prohealthnigeria.com/viewtopic.php?f=317&t=485
1.09	Y	X	http://www.globalhealingcenter.com/miscellaneous-health-and-wellness/body-odor
1.10	Y	X	http://www.shvoong.com/medicine-and-health/1735149-body-odour-causes-rid/
2.01	Y	X	http://serendip.brynmawr.edu/exchange/node/1752
2.02	Y	N	www.34-menopause-symptoms.com/body-odor.htm
2.03	Y	X	http://www.livestrong.com/article/13849-body-odor/#ixzz1cRrKfoL5
2.04	Y	N	http://www.hyperhidrosisweb.com/causes-of-body-odor.html
2.05	Y	X	http://www.mayoclinic.com/health/sweating-and-body-odor/DS00305/DSECTION=causes
2.06	Y	X	http://www.dherbs.com/articles/body-odor-198.html
2.07	Y	X	http://tribune.com.ng/sat/index.php/ask-the-doctor/3257-body-odour.html

2.08	Y	X	http://health.howstuffworks.com/wellness/men/sweating-odor/medicines-cause-body-odor.htm
2.09	N	N	www.webmd.com/healthy-beauty/preventing-body-odor
2.10	N	N	http://www.besthealthmag.ca/get-healthy/home-remedies/natural-home-remedies-body-odour
3.01	Y	X	http://www.homemademedicine.com/home-remedies-body-odor.html
3.02	Y	X	http://www.nhs.uk/conditions/Body-odour/Pages/Introduction.aspx
3.03	Y	X	http://www.cyh.com/HealthTopics/HealthTopicDetails.aspx?p=243&np=292&id=2387
3.04	Y	X	http://www.askmen.com/sports/health/43_mens_health.html
3.05	Y	X	http://www.earthclinic.com/CURES/body_odor.html
3.06	N	N	http://www.stylishandtrendy.com/beauty/how-to-beauty/how-to-get-rid-of-body-odor-causes-of-body-odor-natural-remedies/
3.07	N	N	http://www.child.net/articles/body-odor/
3.08			http://www.crystalspring.co.uk/body-odour-what-causes-a-26.html
3.09	Y	X	http://www.drkrider.com/Topics/Symptoms/Odor,%20body.htm
3.10	N	X	http://www.bodyodor777.com/myths.html
4.01	Y	X	https://zenulife.com/blog/body-odour-causes-body-odour-treatment
4.02	Y	X	www.livingfood101.com/diseases/bodyodor.html
4.03	Y	X	http://nigeriafilms.com/news/13259/50/causes-and-symptoms-of-body-odour.html

3.7.2 Scenario 2

Assuming the perennial traffic congestion being experienced in most part of Lagos State in Nigeria becomes worrisome to the Governor of the state from economical point of view. The Governor needs relevant information that could assist him in taking decision that could resolve this problem. He, as the decision maker, poses this problem as a decision problem. In the context of EI, the problem, "*What is the effect of traffic congestion in Lagos state?*" is specified in terms of stake. The central object of the problem is **traffic congestion**. Assuming that the signal perceived by the Governor includes late resumption to offices and decline in productivity level. He therefore, held a meeting with other EI actors to seek for information on the subject. AMTEA system could be used to obtain values for already defined attributes of stake.

Some students, like in the first scenario, were asked to annotate "***effect of traffic congestion"*** from the first fifty links returned by Google API in order to determine the relevance of each link. By removing the repeated links, forty-eight distinct links were identified. This is shown in Table 3.2. The column marked "X" symbolizes the links on which AMTEA system return results and the column marked "Y" symbolizes the links on which human annotators marked relevant. Symbol "I" represents links where implied attributes were used. The "N" mark indicates that the links were not found to be relevant with regards to the query.

Table 3.2: List of first 50 links returned by Google API for the query "effects of traffic congestion"

RANK	HA	AMTEA	URL
1.01	I		reason.org/files/ps371_growth_gridlock_cities_full_study.pdf
1.02	Y	X	http://myfundi.co.za/e/Causes_and_effects_of_Traffic_congestion
1.03	Y	X	traveltips.usatoday.com/effects-traffic-congestion-61043.html
1.04	Y	X	en.wikipedia.org/wiki/Traffic_congestion
1.05	Y	X	eprints.ucl.ac.uk/1259/
1.06	Y	X	wiki.answers.com/Q/What_are_the_main_effects_of_traffic_congestion
1.07	I		www.google.com.ng/url?sa=t&rct=j&q=effect of traffic congestion&source=web&cd=7&ved=0CGgQFjAG&url=http%3A%2F%2Fwritefix.com%2Fframes%2Ftraffic.ppt&ei=tMXMT6vPIoOi0QWzwqDjAQ&usg=AFQjCNHkTPT79rSZts25-f1hUQu9ZuoaXw&cad=rja
1.08	Y	X	www.eurojournals.com/ejss_16_03_10.pdf

1.09	N	N	www.lattimore.id.au/2007/09/26/traffic-congestion-the-impact-of-parents-with-children/
1.10	L		www.sciencedirect.com/science/article/pii/S0001457509000797
2.01	N	N	www-01.ibm.com/industries/government/ieg/pdf/feeling_the_pain.pdf
2.02	Y	X	www.ehow.com/list_6308569_effects-traffic-congestion.html
2.03	Y	X	www.lpcb.org/lpcb-downloads/isohdm_rue/1995_bennett_greenwood_congestion_fuel.pdf
2.04	N	N	www.wistrans.org/cfire/2011/03/2010-umr/
2.05	N	N	www.ncbi.nlm.nih.gov/pubmed/19540969
2.06	N	N	www.transportationconstructioncoalition.org/Docs/TCC-Harvard-Traffic-Congestion-Report-Final.pdf
2.07	Y	X	www.hartgengroup.net/Projects/National/USA/cong_region_performance/2009-06-22_Final_Summary_PS371.pdf
2.08	N	N	download.sue-mot.org/Conference-2007/Papers/Ogunbodede.pdf
2.09	N	N	www.ops.fhwa.dot.gov/congestion_report_04/executive_summary.htm
2.10	N	N	www.washingtonpolicy.org/publications/opinion/regions-transportation-and-land-use-policies-have-little-effect-traffic-congest
3.01	N	N	www.my3q.com/research/jide/30594.phtml
3.02	Y	X	ro.uow.edu.au/cgi/viewcontent.cgi?article=1013
3.03	N	N	www.its.pdx.edu/upload_docs/1248894217Pm3E7UBzWC.pdf
3.04	N	N	wiki.answers.com/Q/What_are_some_ways_a_traffic_jam_affects_the_society_and_environment
3.05	I		online.wsj.com/article/SB10001424052970203733504577024000381790904.html
3.06	Y	X	www.edrgroup.com/library/freight/traffic-congestion-effects-on-supply-chains.html
3.07	N	N	gao.gov/assets/590/587833.pdf
3.08	I		www-03.ibm.com/press/attachments/28320.pdf
3.09	L		www.tandfonline.com/doi/abs/10.1080/13675569908901576#preview
3.10	N	N	www.ghanaweb.com/GhanaHomePage/NewsArchive/artikel.php?ID=221855
4.01	Y	X	www.resc.org.ps/The Impact of Traffic Congestion and Public Transit on Air Pollution.pdf
4.02	N	N	cts.virginia.edu/docs/UVACTS-5-14-68.pdf
4.03	L		reason.org/studies/show/gridlock-and-growth-the-effect

4.04	N	N	www.google.com.ng/url?sa=t&rct=j&q=effect of traffic congestion&source=web&cd=34&ved=0CFcQFjADOB4&url=http%3A%2F%2Fwww.dbpartnership.org%2Fdocuments%2F%3Fid%3D44&ei=Rt_MT9ryG-Wf0QXXt7W3AQ&usg=AFQjCNE7Anm21pVpUGzkYPpuOk_QnFrFeA&cad=rja
4.05	N	N	americandreamcoalition.org/landuse/denseair.pdf
4.06	Y	X	www.aensionline.com/jasr/jasr/2010/529-542.pdf
4.07	N	N	www.vtpi.org/tca/tca0505.pdf
4.08	L		www.rand.org/topics/traffic-congestion.html
4.10	N	N	www.tomtom.com/landing_pages/trafficmanifesto/index-project.php?Lid=1
5.01	N	N	eprints.port.ac.uk/930/
5.02	Y	X	cdn.intechopen.com/pdfs/18515/InTech-Traffic_congestion_effects_on_supply_chains_accounting_for_behavioral_elements_in_planning_and_economic_impact_models.pdf
5.04	N	N	www.uctc.net/papers/846.pdf
5.05	I		www.ops.fhwa.dot.gov/congestion_report/chapter2.htm
5.06	I		www.internationaltransportforum.org/jtrc/CongestionSummary.pdf
5.07	N	N	seattletimes.nwsource.com/html/opinion/2018116500_guest02cox.html
5.09	N	N	www.ce.utexas.edu/prof/bhat/abstracts/bhat_sardesai_trptb_rev.pdf
5.10	N	N	www.completestreets.org/complete-streets-fundamentals/factsheets/ease-congestion/

3.8 Evaluation Metrics

In Information Retrieval (IR), the accepted practice for evaluating retrieved results is to calculate the precision and the recall metrics. Precision is the number of relevant documents a search retrieves divided by the total number of documents retrieved, while Recall is the number of relevant documents retrieved divided by the total number of existing relevant documents that have been retrieved. However, the two measures are orthogonal. High precision implies that every retrieved document is relevant but one might not have retrieved all relevant documents. Conversely, high recall symbolizes retrieval of large number of documents but the retrieval may include irrelevant documents.

Precisionand Recall are therefore, used as evaluation measures for evaluating AMTEA system against human annotations. Suppose **n** documents are returned as search results, the evaluation measures are defined as follow:

$$Precision = \frac{relevant\ documents\ \cap retrieved\ documents}{retrieved\ documents} \quad 3.2$$

$$Recall = \frac{relevant\ documents\ \cap retrieved\ documents}{relevant\ documents} \quad 3.3$$

Documents retrieved using search engines are ranked in terms of relevance. Therefore, calculation of recall and precision values for these retrieved sets of documents is defined based on the ranking. This approach called ***precision at rank p*** is often used to compare different retrieval algorithms or search engines. If the precision for a ranking at rank position *p* for a retrieval algorithm Ais higher than the precision for another retrieval algorithm B at the same rank position *p*, then A is said to perform better than B, The recall of A will also be higher than recall of B.

Also, the effectiveness of a specific retrieved algorithm across a collection of queries is measured by calculating precision values at standard recall levels from 0.0 to 1.0 in increment of 0.1. Standard recall levels are used as the basis for averaging effectiveness across queries. In order to obtain these precision values, interpolation is required. Interpolation is the technique for calculating a new point from two existing data points.

If r_j, where $j \in (0, 1, 2, \dots, 10)$ is a reference to the standard recall j-th, then, the precision at r_j, $P(r_j)$ is defined thus:

$$P(r_j) = \max r_j \ \leq r \ \leq \ r_{j+1} \quad 3.4$$

The effectiveness of retrieved algorithm for the ranked results obtained for multiple queries is measured by calculating the average precision values. Mean Average Precision (MAP) is the most widely used effectiveness measure, in literature. It is calculated as follows:

$$\text{MAP}(r) \quad = \sum_{k=1}^{N_q} \frac{P_k\,(r)}{N_q} \qquad 3.5$$

Where N_q is the number of queries used, $P_k\,(r)$ is the precision at recall level r for the k-th query

The essence of these metrics is to determine the effectiveness of AMTEA system against human agents. The two performance methods will be adopted. First, precision at rank position will be measured for the comparison and second, precision at *standard* recall levels will be calculated for each and the MAP values compared together.

CHAPTER FOUR

PRESENTATION AND ANALYSIS OF RESULTS

4.1 Results Obtained and Discussion of Results for Scenario 1

Using the scenario, if P-Hand R-H represent precisionand recallfor human annotation respectively and P-AMTEA and R-AMTEA for precisionand recall based on the system annotation, Table 4.1 therefore shows the results obtained from a ranked precision and recall for both human and the system.

Table 4.1: Precision and Recall for results obtained in scenario 1

RANK	P-H	R-H	F-H	P-AMTEA	R-AMTEA
Top 5	4/5 = 0.8000	4/33 = 0.1212	0.2105	4/5 = 0.8000	4/33 = 0.1212
Top 10	9/10 = 0.9000	9/33 = 0.2727	0.4186	9/10 = 0.9000	9/33 = 0.2727
Top 15	14/15 = 0.9333	14/33 = 0.4242	0.5833	12/15 = 0.8000	12/33 = 0.3636
Top 20	17/20 = 0.8500	17/33 = 0.5152	0.6415	15/20 = 0.7500	15/33 = 0.4545
Top 25	22/25 = 0.8800	22/33 = 0.6667	0.7586	20/25 = 0.8000	20/33 = 0.6061
Top 30	23/30 = 0.7667	23/33 = 0.6970	0.7302	22/30 = 0.7333	22/33 = 0.6667
Top 35	26/35 = 0.7429	26/33 = 0.7879	0.7647	25/35 = 0.7143	25/33 = 0.7576
Top 40	28/40 = 0.7000	28/33 = 0.8485	0.7671	27/40 = 0.6750	27/33 = 0.8182
Top 45	28/45 = 0.6222	28/33 = 0.8485	0.7179	28/45 = 0.6222	28/33 = 0.8485
Top 50	28/50 = 0.5600	28/33 = 0.8485	0.6747	28/50 = 0.5600	28/33 = 0.8485

Figure 4.1 and Figure 4.2 show the precision/recall graph for human and system annotations respectively.

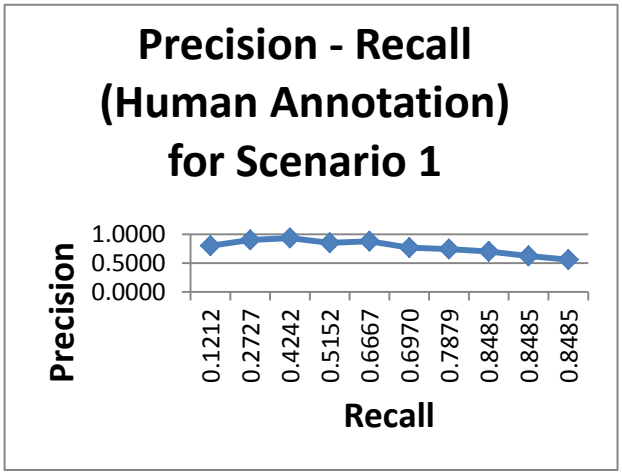

Figure 4.1: Precision/Recall for human annotation in scenario 1

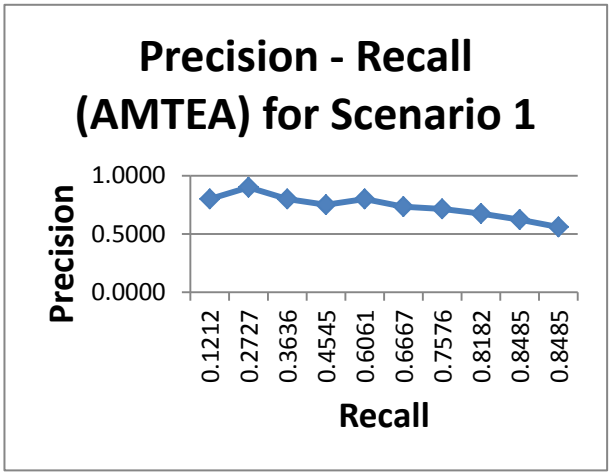

Figure 4.2: Precision/Recall using AMTEA system in scenario 1

From Table 4.1, the precision and recall values, 0.8000 and 0.1212 respectively for the top 5 retrieved documents are the same for both the human agent and the AMTEA system. The implication is that both the human agent and the AMTEA found four of the retrieved five documents relevant for the query "*causes of body odour*";likewise for the top 10, top 45 and top 50. However, there is a slight difference in the top 15 retrieved documents. The precision and recall values at this rank position 15 are 0.9333 and 0.4242 for human agent and 0.8000 and 0.6061 for the AMTEA system. The human agent found fourteen of the retrieved fifteen documents relevant to the query while the AMTEA system found twelve of the documents relevant. The implication is that the precision value of the human agent at rank position 15 is slightly higher than that of the AMTEA system at the same position. The difference is due to the retrieval of documents with ***implied*** attributes by human agent. For example, the selection for similar phrases such as "***body odour is caused by"***, ***"body odour is the smell caused by"***, ***"body*** odour is due to", "smell caused by", "Causes of Changes in Body Odour", give rise to the difference. Where similar attributes were explicitly stated, the results returned by AMTEA system is found to be the same to that of human agent. This was demonstrated in scenario 2.

For the purpose of comparison, standard recall is plotted against P-H and P-AMTEA. Table 4.2 shows interpolated precision and recall for the two forms of annotation created. The resulting graph is shown in Figure 4.3.

Table 4.2: Interpolated precision and recall for scenario 1

Standard Recall	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
P-H	0.80	0.80	0.90	0.93	0.93	0.85	0.77	0.74	0.70	0	0
P-AMTEA	0.80	0.80	0.90	0.80	0.75	0.75	0.80	0.71	0.68	0	0

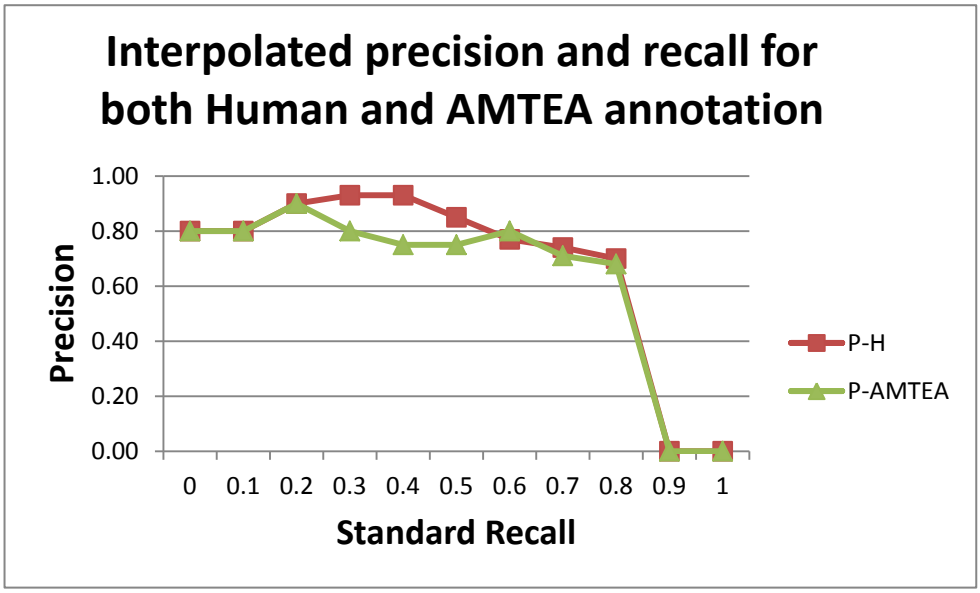

Figure 4.3: Interpolated precision and recall for the two forms of annotation creation

It can be seen from Table 4.2 that there is a marginal difference in precision values of the human agent and the AMTEA system at the 30% to 80% recall. There is 0% precision at recall greater than 90%. At 60% recall, the AMTEA precision value is 0.8 which is higher than 0.77 precision value of the human agent.

4.2 Results Obtained and Discussion of Results for Scenario 2

Table 4.3 shows a ranked precision and recall for both human and system results obtained using scenario 2. Figure 4.4 and Figure 4.5 show the precision/recall graph for human and system annotations respectively.

Table4.3: Precision and Recall for results obtained in scenario 2

RANK	P-H	R-H	F-H	P-AMTEA	R-AMTEA
Top 5	5/5 = 1.0000	5/47 = 0.1064	0.1923	5/5 = 1.0000	5/47 = 0.1064
Top 10	7/10 = 0.7000	7/47 = 0.1489	0.2456	8/10 = 0.8000	8/47 = 0.1702
Top 15	9/15 = 0.6000	9/47 = 0.1915	0.2903	10/15 = 0.6667	10/47 = 0.2128
Top 20	10/20 = 0.5000	10/47 = 0.2128	0.2985	11/20 = 0.5500	11/47 = 0.234
Top 25	12/25 = 0.4800	12/47 = 0.2553	0.3333	13/25 = 0.5200	13/47 = 0.2766
Top 30	14/30 = 0.4667	14/47 = 0.2979	0.3637	14/30 = 0.4667	14/47 = 0.2979
Top 35	15/35 = 0.4286	15/47 = 0.3191	0.3658	16/35 = 0.4571	16/47 = 0.3404
Top 40	17/40 = 0.4250	17/47 = 0.3617	0.3908	17/40 = 0.4250	17/47 = 0.3617
Top 45	18/45 = 0.4000	18/47 = 0.3830	0.3913	19/45 = 0.4222	19/47 = 0.4043
Top 50	19/50 = 0.3800	19/47 = 0.4043	0.3918	20/50 = 0.4000	20/47 = 0.4255

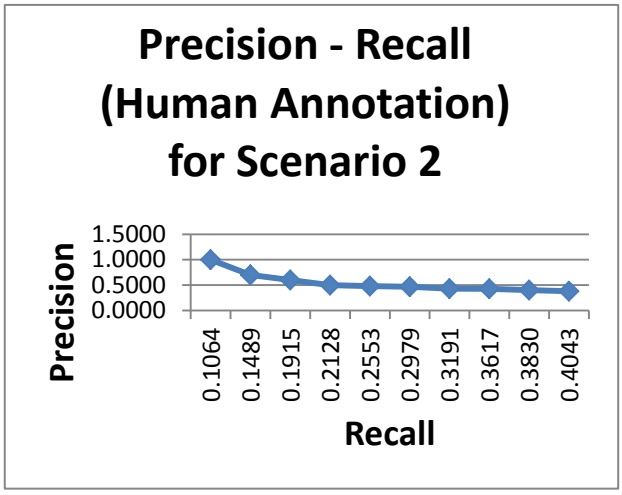

Figure 4.4: Precision/Recall for human annotation in scenario 2

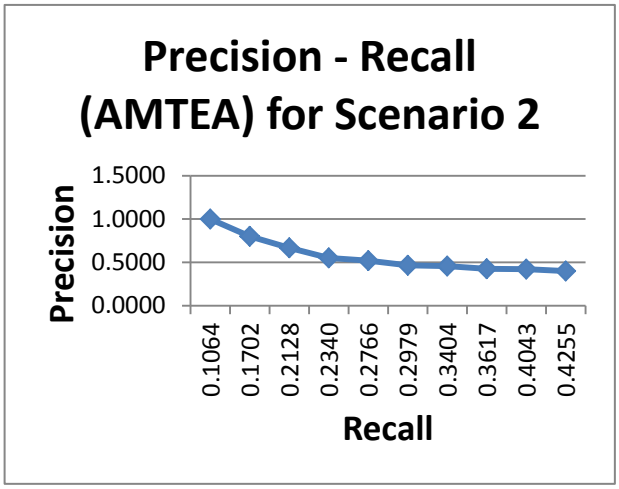

Figure 4.5: Precision/Recall using AMTEA system in scenario 2

From Table 4.3, the precision and recall values, 1.0000 and 0.1064 respectively for the top 5 retrieved documents are the same for both the human agent and the AMTEA system. However, the precision value of 0.8 for the AMTEA system at rank position 10 is greater than that of human agent, which is 0.7. The table revealed that the precision values for AMTEA system is higher. The implication is that the system was able to identified relevant documents in the collection of the retrieved documents where all needed attributes were explicitly stated. Also, for the purpose of comparison, standard recall is plotted against P-H and P-AMTEA for the scenario 2. Table 4.4 shows interpolated precision and recall for the two forms of annotation created.The resulting graph is shown in Figure 4.6.

Table 4.4: Interpolated precision and recall for scenario 2

Standard Recall	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
P-H	1.00	1.00	0.50	0.43	0.38	0	0	0	0	0	0
P-AMTEA	1.00	1.00	0.67	0.46	0.42	0	0	0	0	0	0

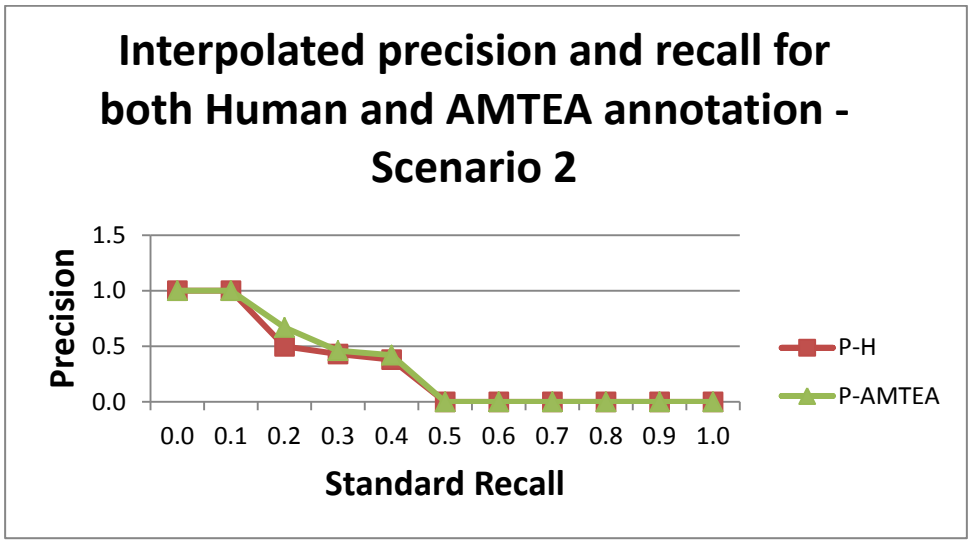

Figure 4.6: Interpolated precision and recall for the two forms of annotation creation in Scenario2

From Table 4.4, at 0% and 10% recall, both the human agent and the AMTEA system have 100% precision. The precision values for both at recall greater than 40% is zero. The precision vale of the AMTEA system is higher than that of human agent at 20%, 30% and 40% recall.

4.3 Measurement of the Effectiveness of AMTEA System against the Human Agent

The effectiveness of Human agent for the ranked results obtained for the two queries (scenario 1 and scenario 2) is measured by calculating its Mean Average Precision (MAP). Table 4.5 shows the precision values for scenario 1 and scenario 2 at standard recall levels for the human agent. Also shown is the average precision value for the two scenarios.

Let MAP(HA) and MAP(AMTEA) represent the Mean Average Precisions for human agent and the AMTEA system respectively.

Table 4.5: Precision values for scenarios 1 and 2 at standard recall levels for the human agent

Standard Recall	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
P-H (scenario1)	0.800	0.800	0.900	0.9333	0.9333	0.850	0.7667	0.7429	0.700	0	0
P-H (scenario2)	1.000	1.000	0.500	0.4286	0.3800	0	0	0	0	0	0
Average Precision	0.900	0.900	0.700	0.6810	0.6567	0.425	0.3834	0.3715	0.35	0	0

From Table 4.5, using equation 3.5, the Mean Average Precision (MAP) is calculated as follows:

$$\text{MAP(HA)} = \frac{0.900 + 0.900 + 0.700 + 0.6810 + 0.6567 + 0.425 + 0.3834 + 0.3715 + 0.35}{11}$$

$$= 0.487945$$

The effectiveness of AMTEA system for the ranked results obtained for the two queries (scenario 1 and scenario 2) is measured by calculating its Mean Average Precision (MAP). Table 4.6 shows the precision values for scenario 1 and scenario 2 at standard recall levels for the AMTEA system. Also shown is the average precision value for the two scenarios.

Table 4.6: Precision values for scenarios 1 and 2 at standard recall levels for the AMTEA System

Standard Recall	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
P-AMTEA (scenario1)	0.8	0.8	0.9	0.8	0.75	0.75	0.8	0.7143	0.675	0	0
P-AMTEA (scenario2)	1	1	0.6667	0.4571	0.4222	0	0	0	0	0	0
Average Precision	0.9	0.9	0.78335	0.62855	0.5861	0.375	0.4	0.35715	0.3375	0	0

$$\text{MAP(AMTEA)}$$

$$= \frac{0.900 + 0.900 + 0.78335 + 0.62855 + 0.5861 + 0.375 + 0.4 + 0.35715 + 0.3375}{11}$$

$$= 0.478877$$

From the calculated MAPs, computing the percentage of MAP (AMTEA) to MAP(HA):

$$\frac{MAP(AMTEA)}{MAP(HA)} * 100\% = \frac{0.478877}{0.487945} = 98.14\% \qquad 4.1$$

It is shown that AMTEA system was able to detect over 98% of retrieved documents which the human agent considered to be relevant to stated queries.

4.4 Summary of Findings

Objective 1: To design and develop an annotation model for creating and storing annotations taking into consideration the semantic context of annotators.

Finding: The thesis was able to design a robust AVP (Attribute-Value Pair) annotation model and the development of a prototype, AMTEA based on the AVP annotation model. The model is designed to capture the intention of actors. The prototype, AMTEA allows actors to load, annotate and store annotations that take into consideration the semantic context of such annotator.

Objective 2: To devise a mechanism for exploiting stored annotations based on the context of problem.

Finding:The objective is achieved by designing query for exploiting stored annotations. The query module is implemented using the EQuA^{2}Te architecture.

Objective 3: To develop a search algorithm that will allow actors to search for information of interest based on the objective of the search.

Finding: The study developed AVP search algorithm that allows user to search for information of interest based on the actor's search objective. The AVP search enables actors to express their search as attribute-value pair.

CHAPTER FIVE
CONCLUSION AND RECOMMENDATION

5.1 Conclusion

The first research objective of this thesis is to design and develop an annotation model for creating annotation by Economic Intelligence actors taking into consideration the semantic context of annotators. This goal is achieved in this thesis by designing a robust AVP (Attribute-Value Pair) annotation model in Chapter Three and the development of a prototype, AMTEA based on the AVP annotation model. AVP annotation model is designed to capture the intention of actors. AMTEA that allows actors to load, annotate and store annotations that take into consideration the semantic context of such annotators. To the best of our knowledge, no AVP annotation model exists in the context of EI.

The second research objective of this thesis is on how best to store created annotations in such a way that will make annotation exploration and exploitation actor-centred. Following the design of AVP annotation model, this goal is achieved by designing query for exploiting stored annotations. The query module is implemented using the EQuA^{2}Te architecture.

The AVP search algorithm allows the user to search for information based on the semantic context of the user. With AVP search, the user expresses his/her query as attribute-value pair.

Finally, the study developed AVP search algorithm that allows the user to search for information of interest based on the actor's search objective. With AVP search, the actor could express his/her search as attribute-value pair. The attribute component is used for categorizing retrieved documents based on the user's objective of search. The value component is the consequence information extracted from corpus based on the actor's specified attribute. This is a novel approach we introduced in EI context.

5.2 Achievements and Contributions

The study has contributed to knowledge in the following areas:

1. Designing a robust annotation model, AVP, and development of a prototype, AMTEA that allows actors to load, annotate and store annotations that take into consideration the semantic context of such annotations.

2. Ability to exploit stored annotations and other relevant information that could enhance solving decision problems based on the context of such problems.

3. The design of AVP search algorithm that allows users to search for information based on the semantic context of search.

5.3 Recommendation for Further Studies

Despite the achievements of this thesis, presentation of results obtained from AVP search is still in textual form. It is believed that a visual representation of obtained results will enhance user's comprehensibility.

The AVP search algorithm introduced in this thesis considers sequential crawling of retrieved documents as well as sequential pattern matching of user chosen phrases. It is believed that parallel crawling approach as well as parallel pattern matching will enhance the performance of the algorithm. A study in this direction is necessary and desirable.

This work presented AVP annotation model and its implemented prototype, AMTEA annotation system, annotation tools for creating and exploiting annotations in the context of EI. Annotation is represented as Attribute-Value pair (AVP) so as to facilitate information reuse and the extensible nature of annotation. Both the logical model and the architectural components of the system were presented. Emphasis was laid on annotation creation, exploration and exploitation as well as on AVP search. Two scenarios were used to illustrate the use of the system. Finally, the system was evaluated in relation to human annotation and results obtained were presented.

REFERENCES

Afolabi B., (2007). La conception et l'adaptation de la structure d'un système d'intelligence économique par l'observation des comportements de l'utilisateur, Thèse de doctorat, Université Nancy 2

Afolabi, B. and Thiery, O. (2006). Using Users' Expectations to Adapt Business Intelligence Systems, Ninth International ISKO Conference on Knowledge Organization for a Global Learning Society, , Vienna, Austria pp.4 - 7.

Bakalov, A., Fuxman, A., Talukdar, P. and Chakrabarti, S. (2011). SCAD: Collective Discovery of Attribute Values. Proceedings of the 20th World Wide Web Conference (WWW-11), pages 447–456, Hyderabad, India.

Bargeron, D., Gupta, A., and Brush, A. J. B. (2001).A Common Annotation Framework. US Patent App 10119544. Retrieved from http://research.microsoft.com/apps/pubs/default.aspx?id=69900

Bhagwat, D., Chiticariu, L., Tan, W.-C., and Vijayvargiya, G. (2005).An annotation management system for relational databases. The VLDB Journal, 14(4), 373-396. VLDB Endowment. Retrieved from http://www.springerlink.com/index/10.1007/s00778-005-0156-6

Bhatnagar, N., Juliano, B. A., and Renner, R. S. (2007) Data Annotation Models and Annotation Query Language. Proceedings of World Academic of Science, Engineering and Technology Volume 21.

Bodain, Y., and Robert, J.-M. (2007). Developing a robust authoring annotation system for the Semantic Web. Seventh IEEE International Conference on Advanced Learning Technologies ICALT 2007, (Icalt), 391-395. Ieee. Retrieved from http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4281043

Bouaka, N. (2004). Développement d'un modèle pour l'explicitation d'un modèle décisionnel : un outil d'aide à la décision dans un contexte d'intelligence économique. Thèse de doctorat en science de l'"Information et de la Communication, Université Nancy 2.

Bouaka, N. and David, A. (2004): “A Proposal of a Decision Maker Problem for a Decision Maker Problem for a Better Understanding of information Needs.” IEEE Explore, (2004), pp 551-552 http://ieexplore.ieee.org/iel5/9145/29024/01307879.pdf

Brandão, W. C., de Moura, E. S., Silva, A. S., and Ziviani, N. (2010). A self-supervised approach for extraction of attribute-value pairs from wikipedia articles. In Int’l Symp.on String Processing and Information Retrieval. pp. 279–289.

Briciu, S., Vrîncianu, M., and Mihai, F. (2009). Towards a New Approach of the Economic Intelligence Process: Basic Concepts, Analysis Methods and Informational Tools. Theoretical Applied Economics, General Association of Economists in Romania. Pp.21-34. Retrieved March 29, 2010 from http://store.ectap.ro/articole/379.pdf

Brush, A. J. B. (2002).Annotating Digital Documents for Asynchronous Collaboration.Significance.ACM Press. Retrieved from http://portal.acm.org/citation.cfm?doid=506443.506472

Brusilovsky, P. (1997). Efficient Techniques for Adaptive Hypermedia.(C. Nicholas and J. Mayfield, Eds.)Intelligent Hypertext, Springer-Verlag.1326, pp.12-30. Retrieved from http://www.springerlink.com/content/d4445498h2847507/?p=e386261816c0457c8756cc37b4d3fc87&pi=20

Buckland, M. K. (1991). Information as thing.Journal of the American Society for Information Science, John Wiley and Sons. 42(5), pp. 351-360. Retrieved from http://doi.wiley.com/10.1002/%28SICI%291097-4571%28199106%2942%3A5%3C351%3A%3AAID-ASI5%3E3.0.CO%3B2-3

Cadiz, J. J., Gupta, A., and Grudin, J. (2000).Using Web annotations for asynchronous collaboration around documents. (D. G. Durand, Ed.)Proceedings of the 2000 ACM conference on Computer supported cooperative work CSCW 00, 309-318. ACM Press. Retrieved from http://portal.acm.org/citation.cfm?doid=358916.359002

Chiticariu, L., Tan, W.-C., and Vijayvargiya, G. (2005).DBNotes : A Post-It System for Relational Databases based on Provenance. (F. Özcan, Ed.)Online, 942-944. ACM Press. Retrieved from http://portal.acm.org/citation.cfm?doid=1066157.1066296

Cimiano, P., Handschuh, S., and Staab, S. (2004a).Towards the self-annotating web.In Proceedings of the 13th World Wide Web Conference, pp. 462–471.

David A. and Thiery O. (2003), "L'Architecture EQuA2te et son Application à l'intelligence économique" Proceeding on the Conférence on "Intelligence Economique : Recherches et Applications.

David, A and Thiery, O. (2001). Prise en compte du profil de l'utilisateur dans un système d'information stratégique, Congrès VSST '01, Barcelona, October 2001 cited in David, 2010 on page196

David, A. (2008). L"information pertinente en intelligence économique.Sous la direction de Papy, Fabrice, Problématiques émergeantes dans les sciences de l'information.Paris : Hermès, p. 209-231.

David, A. (2010). Relevance Information in Economic Intelligence; A book chapter in Information Science edited by Paprice Papy, John Wiley, USA, 2010 page195

Delicious, (n.d.). Retrieved August 3, 2013 from the Wikipedia:http://www.delicious.com

Euzenat, J. (2002). Eight questions about Semantic Web annotations. IEEE Intelligent Systems.IEEE. Retrieved from http://www.bibsonomy.org/bibtex/2b5343cabcb6f276e39e5eb59aee0ea9a/lysander07

Flickr, (n.d.). Retrieved August 3, 2013 from the Wikipedia:http://www.flickr.com/

Folksonomy.(n.d.). Retrieved June 2, 2013 from the Wikipedia: http://en.wikipedia.org/w/index.php?title=Folksonomy&oldid=557718796

Geerts, F., Kementsietsidis, A., and Milano, D. (2006). MONDRIAN: Annotating and Querying Databases through Colors and Blocks. 22nd International Conference on Data Engineering ICDE06, 1(c), 82-82.Ieee. Retrieved from http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1617450

Gertz, M. and Sattler , K.(2001) A Model and Architecture for Conceptualized Data Annotations Technical Report, Department of Computer Science, University of California, Davis.

Goria, S. (2006).Proposition d‘une démarche d‘aide à l‘expression des problèmes de recherche d‘informations dans un contexte d‘intelligence territoriale.Thèse de doctorat, Université Nancy 2.

Handschuh, S., Staab, S., and Volz, R. (2002).On Deep Annotation.Management, 17(3), pg.431.ACM Press. Retrieved from http://portal.acm.org/citation.cfm?doid=775152.775214

Hearst, M. A. (1992). Automatic acquisition of hyponyms from large text corpora.Proceedings of the 14th conference on Computational linguistics (Vol. II, pp. 539-545).Association for Computational Linguistics. Retrieved from http://portal.acm.org/citation.cfm?doid=992133.992154

Kahan, J. (2002). Annotea: an open RDF infrastructure for shared Web annotations. Computer Networks, 39(5), 589-608. Elsevier. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/S1389128602002207

Kiryakov, A., Popov, B., Terziev, I., Manov, D., and Ognyanoff, D. (2004).Semantic annotation, indexing, and retrieval.(D. Fensel, K. Sycara, and J. Mylopoulos, Eds.)Web Semantics Science Services and Agents on the World Wide Web, 2(1), pp.49-79.Elsevier. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/S1570826804000162

Kislin P. and Bouaka, N. (2002). From decision-problem to problem solving in Economic Intelligence process, ICTEI 2002, University of Ibadan, Nigeria, pp. 6-8

Kislin, P. (2007). Modélisation du problème informationnel du veilleur dans la démarche d’intelligence économique. Thèse de doctorat en Sciences de l‟‟information et de la communication. Université Nancy 2.

Knauf, A. (2007). Caracterisation des rôles du coordinateur-animateur : émergence d’un acteur nécessaire à la mise en pratique d’un dispositif régional d’intelligence économique. Thèse de doctorat en Sciences de l‟‟information et de la communication. Université Nancy 2, Octobre 2007.

Knauf, A. and David, A. (2004): “The Role of the informediary in EI process”. Proceedings of the World 8th Multi-conference on Sytemic, Cybernetics and Informatic- WMSCI’04.

Li, Q., Labrinidis, A., and Chrysanthis, P. K. (2008).User-Centric Annotation Management for Biological Data. Management, 72, pp. 54-61.

Li, Q., Labrinidis, A., and Chrysanthis, P. K. (2008).ViP : a User-centric View-based Annotation Framework for Scientific Data. In Vitro, 1-18. Retrieved from http://www.springerlink.com/index/2518742707776707.pdf

MacMullen, W. J. (2005). "Annotation as Process, Thing, and Knowledge: Multidomain studies of structured data annotation". SILS Technical Report TR-2005-02. Chapel Hill: University of North Carolina, School of Information and Library Science, Technical Report Series.

Marshall C. C. (1998) "The Future of Annotation in a Digital (Paper) World" 35th Annual GSLIS Clinic: Successes and Failures of Digital Libraries, University of Illinois at Urbana-Champaign. Retrieved on 11/09/2010.

Martin, H. and Mark, T. K. (2005). An Assessment of Tag Presentation Techniques Retrieved from http://www2007.org/htmlposters/poster988/

Martre, H. (1994). Intelligence économique et stratégie des entreprises. Paris : La Documentation française.

Menéndez, A., Atanes, E., Alonso, J., Merino, C., Bourgogne, P., Geffroy, P., Popkowska, M., Cornbill, J., Dobson, J., Ison, K., Noble, E., Turner, A., Giachetti, C., Chericoni, S., Diana, V., Amati, L., And Mancini, M. A.(2002). Economic intelligence: A guide for beginners and practitioners. CETISME partnership, 2002.Retrieved March 21, 2008 from http://www.madrimasd.org/Queesmadrimasd/Socios_Europeos/descripcionproyectos/Documentos/CETISME-ETI-guide-english.pdf

Odumuyiwa, V. (2010). La Gestion de la Recherche Collaborative d'Information Dans le Cadre du Processus d'Intelligence Economique. Thèse de doctorat en science de l'‟Information et de la Communication, Université Nancy 2.

Odumuyiwa, V. and David, A. (2008). Collaborative Information Retrieval among Economic Intelligence Actors, Proceedings of the fourth International Conference on Collaboration Technologies (CollabTech 2008), Wakayama, Japan, pp. 21-26

Onifade, O. F. W. (2010). A Model for Information Risk Management in Economic Intelligence Systems.Thesis for the award of co-supervised doctorate degree université Nancy 2, France and University of Ibadan, Nigeria.

Onifade, O.F.W. , Oladejo B.F., and Okunoye, O.B. (2009). Risk Indicators as a Source of Knowledge in Economic Intelligence (EI), Symposium d'analyse et de réflexion sur les: Models, methods, Ingéniere de l'intelligence Compétitive, Beaulieu Sur Mer, pp. 25-26.

Ontology.(n.d.). Retrieved August 3, 2013 from the Wikipedia:http://en.wikipedia.org/w/index.php?title=Ontology_(information_science)&oldid=566838854

Pohl, J. (2000). Adapting to the Information Age, InterSymp-2000: 12th International Conference on Systems Research, Informatics and Cybernetics Baden Baden, Germany, July 31-August 4, 2000

Probst, K., Ghani, R., Krema, M., Fano, A., & Liu, Y. (2007). Extracting and Using Attribute-Value Pairs from Product Descriptions on the Web. In B. Berendt, A. Hotho, D. Mladenic, & G. Semeraro (Eds.), From Web to Social Web Discovering and Deploying User and Content Profiles (Vol. 4737, pp. 41-60). Springer-Verlag. Retrieved from http://portal.acm.org/citation.cfm?id=1422162

Revelli, C. (1998). Intelligence stratégique sur Internet. 2ème Ed, Paris : Dunod.

Robert C. A. and David, A. (2006). Annotation and its application to information research in economic intelligence. Journal of Advances in Knowledge Organization, 35-39. Retrieved from http://hal.inria.fr/inria-00089729/en/,

Robert, C. A. (2007). L'annotation pour la recherche d'information dans le contexte d'intelligence économique. Thèse de doctorat en sciences de l"Information et de la Communication. Université Nancy 2.

Sannomiya, T., Amagasa, T., Yoshikawa, M., and Uemura, S. (2001).A framework for sharing personal annotations on web resources using XML.(M. E. Orlowska and M Yoshikawa, Eds.)Aust Comput Sci Commun, 23(6), 40-48. {IEEE} {C}omputer {S}ociety {P}ress. Retrieved from http://portal.acm.org/citation.cfm?id=545625#

Sekine, S. (2004).Named Entity: History and Future. Project notes New York University. Retrieved April 2011 from http://www.cs.nyu.edu/~sekine/papers/NEsurvey200402.pdf.

Simonnot, B. (2010). Relevance in Information Science different models, one theory? A book chapter in Information Science edited by Paprice Papy, John Wiley, USA, 2010 pg145.

Tag cloud, (n.d.). Retrieved August 3, 2013 from the Wikipedia: http://en.wikipedia.org/w/index.php?title=Tag_cloud&oldid=567383753

Wang, X. &Raghupathy, S., 2007. Ink Annotations and their Anchoring in Heterogeneous Digital Documents. In Ninth International Conference on Document Analysis and Recognition ICDAR 2007. IEEE, pp. 163-167. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4378696

Wikipedia, (n.d.). Retrieved August 3, 2013 from the Wikipedia:http://en.wikipedia.org/wiki/Main_Page

Windle R., and Warren S. (n. d.). Collaborative problem solving: steps in the process. Retrieved August 2010 from http://www.directionservice.org/cadre/section5.cfm

Yang, J., Matsuo, Y., and Ishizuka, M. (2007).An Augmented Tagging Scheme with Triple Tagging and Collective Filtering. IEEEWICACM International Conference on Web Intelligence WI07, pp. 35-38. Ieee. Retrieved from http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4427063

Yang, S. J. H., Chen, I. Y.-ling, and Shao, N. W. Y. (2004). Ontology Enabled Annotation and Knowledge Management for Collaborative Learning in Virtual Learning Community. Educational Technology and Society, 7(4), 70-81. IEEE COMPUTER SOC, LEARNING TECHNOLOGY TASK FORCE. Retrieved from http://www.eric.ed.gov/ERICWebPortal/detail?accno=EJ853746

Yee, K.-ping. (2002). CritLink: Advanced Hyperlinks Enable Public Annotation on the Web.. Demo to the CSCW 2002 Conference, New Orleans. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.5.5050&rep=rep1&type=pdf

APPENDIX I

SAMPLE SOURCE CODE

Listing1 - The Homepage

```
<?php
        ob_start();
        session_start();
        include_once('linkdoc/init.php');
?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">

<head>
<title>Home Page</title>
<meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1" />
<link rel="stylesheet" href="ann_cs.css" type="text/css" />
<script type="text/javascript" src="jsfile/annjs.js"></script>
<script language="javascript" src="jsfile/findreplace.js"></script>
<script language="javascript" src="jsfile/ajaxwin.js"></script>
<script language="JavaScript" src="jsfile/amdate.js"></script>
</head>
<body bgcolor="#C0DFFD">
<div id="anwin" style="position:absolute;background-
color:#EBEBEB;left:0px;top:0px;display:none" <div align="right" style="background-
color:navy"><img src="image1/close.gif" onclick="closeit('anwin')"></div><div
id="anwincontent" style="height:100%"><iframe id="cframe" src="" width=100% height=100%
contenteditable='true'></iframe></div></div>

<table width="1000" border="0" cellspacing="0" cellpadding="0">
<tr bgcolor="#3366CC">
<td width="382" colspan="3" rowspan="2"><img src="image1/amtea.jpg" alt="Header image"
width="382" height="127" border="0" /></td>
<td width="378" height="63" colspan="3" id="logo" valign="bottom" align="center"
nowrap="nowrap">AMTEA Annotation System </td>
<td width="100%"> </td>
</tr>
<tr bgcolor="#3366CC">
```

```
<td height="64" colspan="3" id="tagline" valign="top" align="center">SITE-LORIA</td>
        <td width="100%"> </td>
</tr>
<tr>
<td colspan="7" bgcolor="#003366"><img src="image1/mm_spacer.gif" alt="" width="1"
height="1" border="0" /></td>
</tr>

<tr bgcolor="#CCFF99"><?php // shown?>
        <td colspan="2" id="dateformat" height="25">  <script
language="JavaScript" type="text/javascript">
document.write(TODAY);     </script>
        </td>
        <td colspan="2" height="25"><label>Url:<input type="text" name="turl" id="turl"
size="45" />
                <input type="button" name="Go" value="Go"
onclick="javascript:lwin(document.getElementById('turl').value,'frcontent');" />
        </td>
        <td colspan="3" class="bodyText" align="center" id="wel" ><?php include
'wel.php';?></td>

</tr>
<tr>
<td colspan="7" bgcolor="#003366"><img src="image1/mm_spacer.gif" alt="" width="1"
height="1" border="0" /></td>
</tr>

<tr>
<td width="165" valign="top" bgcolor="#E6F3FF">
        <table border="0" cellspacing="0" cellpadding="0" width="165" id="navigation">
<tr>
<td width="165" align="center">
                        <div>
                        <form name="saveddoc" method="">
                                <label class="bodyText">View Saved Document:</label>
                <select id="durl" name="durl"  onchange="jan('durl','turl'); " class="navText">
                                <option value="" >select one</option>
                                <?php  include 'linkdoc/viewsaveddoc.php';
                                ?>
                                </select>
```

```
                    </form>
                </td>
</tr>
<tr>
                <td width="165"><a
href="javascript:showan('linkdoc/homepage.php','maincontent');"
class="navText">Home</a></td>
</tr>
<tr>
                <td width="165"><a
href="javascript:showann(document.getElementById('turl').value,
'linkdoc/annform.php','annot')" class="navText">Annotation</a></td>
</tr>
<tr>
                <td width="165"><a href="javascript:showan('linkdoc/searchalg.php',
'maincontent');" class="navText">Query</a></td>
</tr>
                <tr>
                <td width="165"><a href="javascript:showan('linkdoc/att_synwnbnn.php',
'maincontent');" class="navText">AVP Search</a></td>
</tr>
                <tr>
                <td width="165"><a href="javascript:lwin('http://site.loria.fr/','frcontent')"
class="navText">Contact</a></td>
</tr>
<tr>
                <td width="165"><a
href="javascript:getifrContent('frcontent','maincontent'),closeit('icontent')" class="navText">Get
Webpage</a></td>
</tr>
                <tr>
                <td width="165"><a href="javascript:showan('linkdoc/usingopen.php',
'maincontent');" class="navText">Test Open</a></td>

</tr>
</table>
        <br />
                    <div>
                    <form name="sel2" method="">
                        <label class="bodyText">View Annotated Document:</label>
```

```
                <select id="durl" name="durl"
onchange="showann(this.value,'linkdoc/getann.php','annot'),lwin(this.value,'frcontent'),
document.getElementById('turl').value=this.value; " class="navText">
                                        <option value="" >select one</option>
                                        <?php  include 'linkdoc/viewannotateddoc.php';
                                        ?>
                                        </select>
                                        </form>
                                </div>
</div>
        <div id="annotm"></div>
         <br /> </td>
<td width="50"><img src="image1/mm_spacer.gif" alt="" width="50" height="1" border="0"
/></td>
<td width="400" colspan="2" valign="top"><img src="image1/mm_spacer.gif" alt=""
width="305" height="1" border="0" /><br />

        <table border="0" cellspacing="1" cellpadding="1" width="400" bgcolor="#F0F0F0">
<tr>
<td class="bodyText">

                <div id="maincontent" class="bodyText" onmouseup="getSelText()">
                        <?php
                                include("linkdoc/att_synwnbwrk2.php");
                                include("linkdoc/homepage.php");
                         ?>
                </div>
                </td>
                </tr>

</table>
            <br /></td>
<td width="50"><img src="image1/mm_spacer.gif" alt="" width="50" height="1" border="0"
/></td>
<td width="190" valign="top"><br />
<div id="gsh">
<div id="cse" style="width: 100%;">Loading</div>
        <script src="http://www.google.com/jsapi" type="text/javascript"></script>
        <script type="text/javascript">
        google.load('search', '1', {language : 'en', style : google.loader.themes.GREENSKY});
```

```
        google.setOnLoadCallback(function() {
var customSearchControl = new
google.search.CustomSearchControl('012250348746701193287:u2dmhlobkqs');
customSearchControl.setResultSetSize(google.search.Search.FILTERED_CSE_RESULTSET);
customSearchControl.draw('cse');
  }, true);
</script>
</div>

<div id="annot">
<?php    include 'linkdoc/success.php';
include("linkdoc/searchdisp.php");
include 'linkdoc/reg_success.php';
 ?>      </div>
<div id="content">
        <form ><div align="right"style="background-color:#000066" ><img
src="image1/save.gif" onclick="showann(document.getElementById('turl').value,
'linkdoc/savepage.php','annot')"><img src="image1/close.gif"
onclick="closeit('content')"></div></form><iframe  id="frcontent" width="400" height="400"
contenteditable="true" allowtransparency="true" frameborder="0" ></iframe></div>

                        </td></tr>
                </table>
                </td>
        <td width="100%"> </td>
</tr>
<tr>
<td width=""> </td>
<td width=""> </td>
        <td width=""> </td>
<td width=""> </td>
<td width=""> </td>
<td width=""> </td>
        <td width=""> </td>
</tr>
</table>
</body>
</html>
```

Listing2 – Sample Annotation Form Source Code

```
<?php
session_start();
	$q=$_GET['q'];

if( !isset($_SESSION['SESS_LAST_NAME']))
	{	header("Location: loginfrm.php");}
	$ann =   "ANN". rand(5, 100).date('His');

?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
<meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1" />
<title>Annotation</title>
</head>

<body>
<form action="" method="post" name="frmav" id="frmav" >
<table width="200" border="" align="center" bgcolor="#FFCCCC">
<tr align="center">
	<td  class="bodyText" colspan="3">ANNOTATION FORM</td>
</tr>
	<tr>
<td colspan="2" align="center"><input size="25" name="txtsurname" type="text"
class="navText" id="txtsurname"  value ="<?php echo  $_SESSION['SESS_LAST_NAME'];?>.
<?php echo  $_SESSION['SESS_FIRSR_NAME'];?>.  <?php echo
$_SESSION['SESS_SPECIAL']; ?>" /></td>
	<td colspan="1" align="center"><input size="5" name="nameid" type="text"
class="navText" id="nameid"  value ="<?php echo  $_SESSION['SESS_MEMBER_ID'];?>"
/></td>
</tr>

<tr>
<td colspan="3" >Url:<input size="30" name="url" type="text" id="url"  class="navText" value
="<?php echo $q; ?>"/></td>
</tr>
```

```
	<tr>
<td colspan="3" class="navText"><input type="radio" name="antype" id="antype"
value="private"  checked=""/>
		Private<input type="radio" name="antype" id="antype" value="community" />
		Group<input type="radio" name="antype" id="antype" value="public" />
		Public	</td>
	</tr>
	<tr>
<td colspan="3" align="center">Annotated Object:<input size="35" name="needle" type="text"
id="needle"  class="navText"/></td>
</tr>
<tr>
<td colspan="3"  class="bodyText">Atr/Val:<input name="attr" type="text" size="10"
class="navText"><input name="val" size="20" type="text"  class="navText"></td>
</tr>
<tr><td colspan="3" id="ann" class="bodyText" ></td></tr>
	<tr>
<td  class="bodyText" colspan="2">Format:<select name="anformat" id="anformat"
onChange="Choix(this.value,'anformat','display_fld2')">
	<option value="null">--- faire une selection---</option>
	<option value="Text">Text</option>
	<option value="Image">Image</option>
	<option value="Voice">Voice</option>
	<option value="other">Others</option>
</select></td>
<td id="display_fld2"  class="bodyText"></td>
	</tr>
<tr>
<td colspan="1" class="bodyText"><input name="pre" size="5" id="pre" type="hidden"
class="navText" ></td>
<td colspan="1" class="bodyText"><input name="anid" size="10" id="anid" type="hidden"
class="navText" value="<?php echo $ann; ?>"/></td>
	<td colspan="1" class="bodyText"><input name="pos" size="5" id="pos" type="hidden"
class="navText" ></td>
</tr>
<tr>
<td colspan="3" class="bodyText"><input name="annsubmit" type="submit" value="Submit"
id="annsubmit" />
<input name="annreset" type="reset" value="Reset" id="annreset" /></td>
</tr>
```

```
</table>
</form>

</body>
</html>
```

Listing 3 – Sample Query Source Code

```
<?php
include '../Conn/conn.php';
print "<html><head><title>My Query Engine</title><link rel='stylesheet' href='../ann_cs.css'
type='text/css'/>";
print "<script language='javascript' src='../jsfile/annjs.js'></script></head><body>";
?>
<form name="qfrm" id="qfrm" method="post" action="" target="_self">
                                <div>
                                <span class="navText">Search opt1
</span>
                                <select name="Attribute1" id="Attribute1" class="memorize">
                                        <option value="">Select one</option>
                                        <?php
                                $result = mysql_list_tables("amteadb");
                                for ($i = 0; $i < mysql_num_rows($result); $i++)
                                {
                                        $re = mysql_tablename($result, $i);

if ($re == 'annobj'||$re == 'annobject'||$re == 'annobjecta'||$re == 'annobjecta2'||$re ==
'annobjectb'||$re == 'annobjectuser'||$re == 'annobjectusera'||$re == 'annobjectusera1'||$re ==
'annobjectusera2'||$re == 'annobjectuserb'||$re == 'atttable'||$re == 'attribute2'||$re ==
'accesses'||$re == 'attributetable'||$re == 'belongs'||$re == 'newtbl'||$re == 'pix'||$re ==
'signaltb'||$re == 'subject_domain'||$re == 'testlogin'||$re == 'occ'||$re == 'att'||$re == 'syn'||$re ==
'att_syn')
                                        {continue;}
                                        else{
                                        $query = mysql_query("select  * from $re");
                                        $num_fields = mysql_num_fields($query);
                                        $row = mysql_fetch_array($query);
                                        for ($ii = 0; $ii < $num_fields; $ii++) {
                                        $r = mysql_field_name($query, $ii);
```

```
                                        if($r== 'userid'|| $r=='docid'||
$r=='annid'||$r=='doc_objid'){continue;}else{
                        ?>
<option class="navText" value="<?php echo $re."@-@".$r;?>"><?php echo $re."@-
@".$r;?></option>
<?php }}}
                                    } //end for loop
                                    ?>
                                        </select>
                </div>
<div>
                                    <span class="navText">Search opt2
                                    </span>
                                    <select name="Attribute2" class="memorize">
                                        <option value="">Select one</option>
                                        <?php
                                    $result = mysql_list_tables("amteadb");
                                    for ($i = 0; $i < mysql_num_rows($result); $i++)
                                    {
                                        $re = mysql_tablename($result, $i);

if ($re == 'annobj'||$re == 'annobject'||$re == 'annobjecta'||$re == 'annobjecta2'||$re ==
'annobjectb'||$re == 'annobjectuser'||$re == 'annobjectusera'||$re == 'annobjectusera1'||$re ==
'annobjectusera2'||$re == 'annobjectuserb'||$re == 'atttable'||$re == 'attribute2'||$re ==
'accesses'||$re == 'attributetable'||$re == 'belongs'||$re == 'newtbl'||$re == 'pix'||$re ==
'signaltb'||$re == 'subject_domain'||$re == 'testlogin'||$re == 'occ'||$re == 'att'||$re == 'syn'||$re ==
'att_syn')
                                        {continue;}
                                        else{
                                        $query = mysql_query("select  * from $re");
                                        $num_fields = mysql_num_fields($query);
                                        $row = mysql_fetch_array($query);
                                        for ($ii = 0; $ii < $num_fields; $ii++) {
                                        $r = mysql_field_name($query, $ii);
                                        if($r== 'userid'|| $r=='docid'||
$r=='annid'||$r=='doc_objid'){continue;}else{
                        ?>
<option class="navText" value="<?php echo $re."@-@".$r;?>"><?php echo $re."@-
@".$r;?></option>
<?php }}}
```

```
                                } //end for loopsession_write_close();
                                ?>
                                </select>
                                </div><div>
                                <span class="navText">Search opt3 </span>

                                <select name="Attribute3" class="memorize">
<option value="">Select one</option>
                                        <?php
                                $result = mysql_list_tables("amteadb");
                                for ($i = 0; $i < mysql_num_rows($result); $i++)
                                {
                                        $re = mysql_tablename($result, $i);
if ($re == 'annobj'||$re == 'annobject'||$re == 'annobjecta'||$re == 'annobjecta2'||$re ==
'annobjectb'||$re == 'annobjectuser'||$re == 'annobjectusera'||$re == 'annobjectusera1'||$re ==
'annobjectusera2'||$re == 'annobjectuserb'||$re == 'atttable'||$re == 'attribute2'||$re ==
'accesses'||$re == 'attributetable'||$re == 'belongs'||$re == 'newtbl'||$re == 'pix'||$re ==
'signaltb'||$re == 'subject_domain'||$re == 'testlogin'||$re == 'occ'||$re == 'att'||$re == 'syn'||$re ==
'att_syn')
                                        {continue;}
                                        else{
                                        $query = mysql_query("select  * from $re");
                                        $num_fields = mysql_num_fields($query);
                                        $row = mysql_fetch_array($query);
                                        for ($ii = 0; $ii < $num_fields; $ii++) {
                                        $r = mysql_field_name($query, $ii);
                                        if($r== 'userid'|| $r=='docid'||
$r=='annid'||$r=='doc_objid'){continue;}else{
                                ?>
<option class="navText" value="<?php echo $re."@-@".$r;?>"><?php echo $re."@-
@".$r;?></option>
<?php }}}
                                } //end for loop
                                ?>
                                </select>

                        <input type="reset" name="Reset" value="Reset">
                        <input type="submit" name="Submitt" value="submit">
</form>
</body></html>
```

Listing 4 – Sample Search Source Code

```
<?php
print "<html><head><title>My Search Engine</title><link rel='stylesheet' href='../ann_cs.css'
type='text/css'/>";
print "<script language='javascript' src='../jsfile/ajax2.js'></script></head><body>";

function atval($at,$va)
{
include 'projectX/important_classes/html_url_parser.php';
include 'projectX/stemmer.php';


echo '<br/>';
          echo $at."XXX".$va;
          $str = $va." of ".$at;
   $page = 'http://search.yahoo.com/search?p='.str_replace(' ', '+', $str);
   $reader = new HtmlReader();

             $opts = array(
   'http' => array(
      'proxy' => 'tcp://asoju.unilag.edu.ng:3128',
      'request_fulluri' => true,
      'header' => "Proxy-Authorization: Basic $auth",
   ),
);

   $html = $reader->getPageContent2($page,$opts);
   $htmlDoc = new HtmlDocument($html);
   $body = $htmlDoc->getResultBody();

   $links = $body->grabLinks();
   $stemmerObject = new Stemmer();
     $stemmedDef = $stemmerObject->stem($at);
     $stemmedWord = $stemmerObject->stem($va);
echo "<br/><br/> STEM OF WORD: ".$stemmedDef."<br/><br/>";

   $i = 1;
foreach ($links as $url) {
```

```
      $page = $url->url;
      $reader = new HtmlReader();
      $html = $reader->getPageContent2($page,$opts) or "";
if($html=="")
      {
continue;
      }
      $htmlDoc = new HtmlDocument($html);
      $body = $htmlDoc->getBody();
      $cleanBody = $body->getStrippedBody();
      $sentences = preg_split('/(?<=[.?!;])\s+(?=\p{Lu})/', $cleanBody->getContent());

foreach ($sentences as $str) {
if(stripos($str, $stemmedDef)!== FALSE){
if(stripos($str, $strings[1])!==FALSE)
          {
echo "<bold style='color:blue'> ".$i." </bold>".$str."<br/>"
                    . "<br/>"
                    . "<a href='".$url->url."'>".$url->url."</a>";
echo "<br/><br/>**********************************************<br/>";
                      $i++;
          }

       }
   }
   }


}

function getmicrotime()
{
list($usec, $sec) = explode(" ",microtime());
return ((float)$usec + (float)$sec);
}
function findexts ($filename)
 {
          return substr($filename, strrpos($filename, '.')+1);
 }
function attval($url,$at,$j)
```

```
{
	$fr = fopen('avpResult.txt','w');
	$cont = file_get_contents($url);
	$cont = strip_tags($cont);
	$cont = ereg_replace('/&\w;/', '', $cont);

	$sentn = preg_split ('/(?<=[.!?;])\s+(?=\p{Lu})/i',$cont,-1,PREG_SPLIT_NO_EMPTY);


		$pat = "/((.*)($at)+\s+(?:[is|can be|has been]*)\s+($va)+(.*))/i"; //|
	$k=1;
	foreach($sentn as $line) {
	if(trim($line)==='') continue;
		if (preg_match_all($pat, $line, $l)) {
			foreach($l[0] as $link){
				if(!empty($link)){
				print "<u><b>Filename ($j):</b> $url</u><br/><b>Paragraph $k :
</b>". $link;print "<br/>";
				fwrite($fr,'Filename ('.$j.') '.$url.'\n Paragraph $k : '. $link);
				fwrite($fr, '\n');
				  }
				else continue;
			  } print "<br/>";
		} $k++;
	}
	fclose($fr);
 }
function satt(){
	if( isset ($_POST['sel'] ))
	{
	  $a = $_POST['deflemma'];
	  $k=1;  $ary=array();
		foreach($a as $f)
		{
		print "<tr><td colspan='2'  class='bodyText'>Similar attributes:<input type='text'
size='20' name='".$f."' value='".$f."'></td></tr> ";
		  array_push($ary,$f);
		  $k++;
		}
			$at = "";
```

```
                foreach($ary as $abbr) {
                  $at = $at . (empty($at) ? " : '|').$abbr ;
                }
                return $at;
          }
}
if( isset ($_POST['ret'] ))
{
  $attrb = addslashes( $_POST['attrb'] );
  $val = ( $_POST['val'] );
  $fw = fopen('syndata.txt', 'w');
        $url =
"http://words.bighugelabs.com/api/2/1b3093a931c02274fc86d862c333088e/".$attrb."/json";
        $html="";


        $handle = fopen($url, "rb");
while (!feof($handle)) {
 $html .= fread($handle, 8192);
}
fclose($handle);


        $jso = json_decode($html);
        $jnoun=$jso->noun->syn;
                $jverb=$jso->verb->syn;
print "<form method='post'>";
print "<table width='200' border='1'>";
        print "<tr><td  class='bodyText'><input name='deflemma[]' id='deflemma[]'
type='checkbox' value='$attrb of $val ' checked='checked' />".$attrb."</td></tr>";

        foreach ( $jnoun as $trend )
        {
          print "<tr><td  class='bodyText'><input name='deflemma[]' id='deflemma[]'
type='checkbox' value='{$trend} of $val '  />{$trend}</td></tr>";

        fwrite($fw,$trend .'\r\n ');
  }
print "<br/>";
        foreach ( $jverb as $treverb )
```

```
        {
        print "<tr><td  class='bodyText'><input name='deflemma[]' id='deflemma[]'
type='checkbox' value='{$treverb}'  />{$treverb} </td></tr>";
        fwrite($fw,$treverb .'\r\n ');
  }
fclose($fw);
        print "<tr><td colspan='2' class='bodyText' align='center'><input type='submit'
value='select' name='sel'></td></tr> ";
        print "</table>";
        print "</form><br/>";
}
if( isset ($_POST['retr'] ))
{
  $start_time = getmicrotime();

  $va = addslashes( $_POST['va'] );
  $trb = addslashes( $_POST['att'] );
  $trb = explode ('|',$trb);
        $aa = preg_replace("/of/","",$trb[0]);
        $a = $aa.' ' .$va;
  $k=1;

foreach ($trb as $att)
  {
  $attt = preg_replace('/\s/', '+', $att);
  $atbing = preg_replace('/\s/', '%20', $att);
print "<h2>Search results for '".$attt."':</h2><br/>";

if($_POST['sourc']=='web' && $_POST['sch']=='google'){
                //searchgoogle($attt, $att);
                atval($att,$va);

}else if($_POST['sourc']=='Info'){
  $inf  = ( $_POST['pt'] );

        if ($handle = opendir($inf))
        {
        $k=0; $farray=array();
                while (false !== ($file = readdir($handle)))
                {
```

```
                        $et = findexts($file);
                        if($et=='htm'||$et=='html'||$et=='php'||$et=='txt')
                        {
                                $fname = $file;
                                $f = $inf.$fname;
                                array_push($farray,$f);
                                $k++;
                                attval($f,$att,$k);
                        atval($att,$va);
                        }
                }
                closedir($handle);
        }print "<br/>".$k;

}else{
echo "Choose Information Source<br/>";
}
$k++;
}
  $end_time = getmicrotime();

print "  "."  query executed in ".(substr($end_time-$start_time,0,5))." seconds.";
}
if( isset ($_POST['sel'] ))
{
print "<form method='post' target='_self'>";
print "<table width='500' border='1'>";
print "<td colspan='2' class='bodyText'>Information Source:<select name='sourc' id='sourc'
onChange=\"Selec(this.value,'sourc','pat')\" class='navText'>
        <option value='null'>--- Select Web or Info system---</option>
        <option value='web'>From Web</option>
        <option value='Info'>Infomation System</option>
</select>
</td>";
        print " <tr><td colspan='1' class='navText'><input type='radio' name='sch' id='sch'
value='google' checked="/>Google<input type='radio' name='sch' id='sch'
value='bing'/>Bing</td>
                        <td colspan='1' class='navText'><input type='radio' name='sch'
id='sch' value='yahoo' />Yahoo</td></tr>";
```

```
print "<tr id='pat' ></tr>";
print "<tr><td colspan='2' class='bodyText'>Value:<input type='text' size='34' name='va' id='va'
></td></tr> ";
print "<tr><td colspan='2' class='bodyText'>Attributes:<input type='text' size='34' name='att'
id='att' value='".satt()."' ></td></tr> ";
print "<tr><td colspan='2' class='bodyText' align='center'><input type='submit' value='Retrieve'
name='retr'></td></tr> ";
print "</table>";

print "</form><br/>";
 }
print "</body></html><br/>";
?>
```

APPENDIX II

PUBLICATIONS FROM THE THESIS

Olusoji OKUNOYE, Fausat OLADEJO and Victor ODUMUYIWA: "Dynamic Knowledge Capitalization through Annotation among Economic Intelligence Actors in a Collaborative Environment". VSST 2010 Conference, Toulouse, France; October 25 – 29, 2010.

Olusoji. B. Okunoye, Amos David and Charles Uwadia: AMTEA: Tool for Creating and Exploiting Annotations in the Context of Economic Intelligence (Competitive Intelligence). *IEEE IRI 2010 Conference, Las Vegas, USA; August 4 – 6, 2010.*

Olufade F. W. ONIFADE, Olusoji B. OKUNOYE and Amos DAVID: Embedded fuzzontological model for document interpretation in attribute-value-pair annotation in economic intelligent systems; KMO 2010 Conference, Veszprém, Hungary, May 18-19, 2010

O.F.W. Onifade, B.F. Oladejo and O.B. Okunoye, “Risk Indicators as a Source of Knowledge in Economic Intelligence (EI)”, Symposium d’analyse et de réflexion sur les: Models, methods, Ingéniere de l’intelligence Compétitive, Beaulieu Sur Mer, 25-26 Nov 2009.

Victor ODUMUYIWA, Soji OKUNOYE & Amos DAVID "*Knowledge creation and enhancement through collaborative information retrieval* », 1st conference of the British chapter of ISKO, June 22 and 23, 2009, London, UKA framework for developing a Robust Annotation Tools for Economic Actors