An efficient method to improve Information Recovery on Web

Dr. Muhammad Shahbaz [1],Dr. Syed Muhammad Ahsen[2], Farzeen Abbas[3], Muhammad Shaheen[4]
Syed Athar Masood[5]
[1,2,3, 4] Department of Computer Science & Engg, UET Lahore, Pakistan
[5]Department of Engineering Management, NUST College of E&ME, Rawalpindi Pakistan
[1] M.Shahbaz@uet.edu.pk, [2] Ahsencs@hotmai.com, [3]farzeenabbas@gmail.com, [4]shaheen@uet.edu.pk, [5] atharmasood2000@hotmail.com

Abstract: The information content on Web is very large and number of users is interacting with it in diverse manner and is growing fast. Information retrieval on web is concerned to capture precise and accurate content as requested in real time. To enhance the performance of Information Retrieval on web quality metrics are needed to be satisfied. The proposed research deals with introduction of such quality metrics, which can improve Information Retrieval systems on web. Different quality matrices are analyzed that are used for IR system. Information Retrieval metrics are already defined but they still could not make up with the relevancy requirement of users. Proposed Subsumption metrics is based on ontology to improve user query results and enhance the quality of retrieval. IR systems based on ontology are already in practise but they are not using any kind of metrics and they are specific to their respective domain. It is required to introduce such kind of metrics which is generic to all the systems and improves relevancy by incorporating Subsumption metrics. [Dr. Muhammad Shahbaz, Dr. Syed Muhammad Ahsan, Farzeen Abbas, Muhammad Shaheen. An An Efficient Method to Improve Information Recovery on Web. Journal of American Science 2011;7(7):13-23]. (ISSN: 1545-1003). http://www.americanscience.org.

Keywords: Subsumption Metrics, Information Recovery

Introduction

World Wide Web is a very diverse environment unlike usual situation, where traditional information retrieval systems function. Conventional IR systems function in extremely controlled, central and comparatively stable environment. New documents are added in the system, but in a restricted fashion. At times previous documents are detached or moved (for instance in case the "contemporary" database of a bibliographic IR system maintains documents from the last two years only, and older ones are saved to an archive). Also documents or their representations might change and mistakes could be corrected and problems could be solved. The major issue is that there ought to be quality metrics available to estimate and evaluate the performance of IR system based on web.

A classical IR system designed for Web based retrieval has to face different challenges correlated to the *functionalities and requirements* as following.

- Comprehensive coverage: Since the size of web is very large comprising of huge information base; the system must have the capability to retrieve data as required by user from all of the recourses within a complex structure of web.
- Effective discovery: The IR system must be efficient, fast and effective in order to respond significance relevance & ranking necessities and give valuable and useful search results.
- Contextual: The system must be capable to hold partial, imprecise, incomplete data & frequently depends on contextual info.
- Adaptive: The system must be adaptive in a sense of query updation & modifications, and also query interface and usability.
- Ability to learn: The system must have a learning element in order to obtain information regarding the user, the information content and structure, and user and system's interactions.
- Effective content delivery: The system must be capable to rank, recapitulate, and personalize information content for correctness in delivery.
- Easy to use: It must be robust, easy and flexible in order to use with greatly developed and smart, intelligent and simple user interfaces.

Consequently, quality metrics are required in order to provide proficient and effective parameters & measures for IR system based on web.

Evolutionary and Adaptive Web Based Information Retrieval System

The Internet and World Wide Web (WWW) are continuously becoming more dynamic and active in terms of information content and use. Information retrieval (IR) aims to keep up with this environment by manipulative intelligent systems that can deliver Web information content in real time to variety of wired and wireless devices. Evolutionary & adaptive systems are emerging as classical examples of such kind of systems. This paper gives one of attempts to collect and evaluate the type and nature of recent research on Web based IR using evolutionary & adaptive systems and proposes research incentives in

parallel to advancement and developments in Web environments. [1]

Criteria for Evaluating Information Retrieval Systems in Highly Dynamic Environments

The World Wide Web is complex environment as compared to usual setting, where typical information retrieval (IR) systems work or operate. Classical information retrieval systems operate in a well controlled, central and comparatively stable environment. Documents could be added, but in a controlled manner. Old documents are moved (such as in case of "current" database tables of a bibliographic information retrieval system maintains documents from previous two years, and the older are saved in an archive); also documents representations might change such as correction of mistakes and so on. The main point is that all of the processes functions in highly controlled manner. [2]

Semantic Metrics, Conceptual Metrics, and Ontology Metrics

Resemblances and diverseness b/w "semantic metrics" (i.e. metrics defined on a knowledge based IR system) & "conceptual metrics" (i.e. metrics defined based on a Latent Semantic Indexing IR system) are discussed in the paper. Prospective collaboration areas among research groups are recognized. Prospective application and cooperation areas of research area known as "ontology metrics" metrics computed on the base of ontologies that constitutes the part of an ontology software system, are also discussed. At present ontology metrics are considered using techniques that are similar to semantic metrics, but there are other semantic based extensions similar to conceptual metrics. [3]

Ontology based Web Crawler

The constraint of a web Crawler to facilitate downloading of relevant pages is a major confront in field of Information Retrieval (IR) Systems. Usage of link analysis algo such as page rank and Importance metrics have shown a novel move toward prioritizing URL queue to download highly relevant pages. In current paper, the amalgamation of above metrics along with latest metric known as **association metric** has been proposed. The **association metric** approximates the semantic substance of the URL that is based on domain based ontology; that in return reinforces the metric used for recommending URL queue. Proposed metric would solve the main problem of discovery the relevancy of pages prior to procedure of crawling to the best possible and optimal level. [4]

Problem Statement:

Keeping in view the swiftly increasing rate of web and also its applications; is creating a lot of aspects underneath contemplation. To deal with dynamically budding technology, it has turn out to be indispensable to enhance IR systems based on web.

I am interested here to propose metrics that will increase the quality of web-based IR Systems. This proposed research deals with the problem of introducing such metrics, which will be used to improve the quality of web-based IR Systems.

Methodologies:

Following activity plan is laid to accomplish the required objective

1. Extensive review of the existing IR matrices required for non-web based system.
2. Review of methods/models used to maintain information in www.
3. Formalization to use non-web based IR matrices for web based IR systems
4. Case-study, evaluation and validation of metrics.
5. Development, testing and verifications of purposed solution.
6. Comparison between existing IR metrics and proposed metrics.
7. Conclusions over comparison results.

Proposed Solution

Metrics are set of variables or methods of quantitative and occasional evaluation of a process including the events to perform measurement. Metrics are specific to a domain and that is why they are only applicable in particular field of study for which they are designed. [5]

Since we are concerned about metrics of Information Retrieval System, first have a look at Information retrieval System. Information Retrieval System is the discipline for extracting information from documents that includes searching of documents and the meta data which explains the documents and also searching through databases that might be Relational databases or Hypertext as in case of World Wide Web. However, there is mystification among the terms document retrieval, text retrieval, information retrieval, data retrieval. Each of these disciplines is based on its own theory and technology. Information Retrieval is the most promising and emerging field that is interdisciplinary and have overlapping of computer science, maths, statistics, physics, linguistics and information sciences. [6]

Information Retrieval Systems based on ontology is of prime concern to get more structured and meaningful data or result. Ontology is a term originated from Philosophy. In computer sciences ontology is being used as a model for description of types, properties and relationships of objects. Though provided definition of ontology do varies, but it is assumed that real world and ontology features are analogous. [7]

Ontology based Information retrieval systems are already in practise now.

For geographical information retrieval of place by means of place name, ontology based Information retrieval system was developed which returns the results in form of documents images and records of the desired geographical place.[8]

Textpresso is text mining IR system for biological data using biological concepts e.g. genes, cell, cell group and biological processes that relate the objects. [9]

David Vallet, Miriam Fernández, and Pablo Castells proposed ontology based retrieval model for exploitation of ontologies in knowledge bases to imply semantic search. [10]

All of the IR systems discussed above are based on ontology but they are not using any kind of metrics and they are specific to their respective domain. It is required to introduce such kind of metrics which is generic to all the systems and improves relevancy by incorporating ontology in Information Retrieval Systems. Also ontologies introduced must be refined. For this purpose normalization techniques are employed.

Normalization is being used to discover attributes of semantics of ontology in the manuscript constitution. Five normalization forms are introduced to identify each and every class / subclass of ontology, eliminate redundancy and materialize each property instance.

SubSumption Metric

World Wide Web is a major interactive source of business & daily life information. The information content on Web is very large and number of users is interacting with it in diverse manner and is growing fast. Information retrieval on web is concerned to capture precise and accurate content as requested in real time. To enhance the performance of Information Retrieval on web quality metrics are needed to be satisfied. The proposed research deals with introduction of such quality metrics, which can improve Information Retrieval systems on web. Different quality matrices are analyzed that are used for IR system. Information Retrieval metrics are already defined but they still could not make up with the relevancy requirement of users. Proposed metrics is based on ontology to improve user query results and enhance the quality of retrieval. Hence ontology is essential part of projected subsumption metrics.

What ontology has in common in both computer science and in philosophy is the representation of entities, ideas, and events, along with their properties and relations, according to a system of categories. In both fields, one finds considerable work on problems of ontological relativity. Ontologies are used in Artificial Intelligence, the Semantic Web, Software Engineering, Biomedical Informatics, Library Science, and Information Architecture as a form of knowledge representation about the world or some part of it.

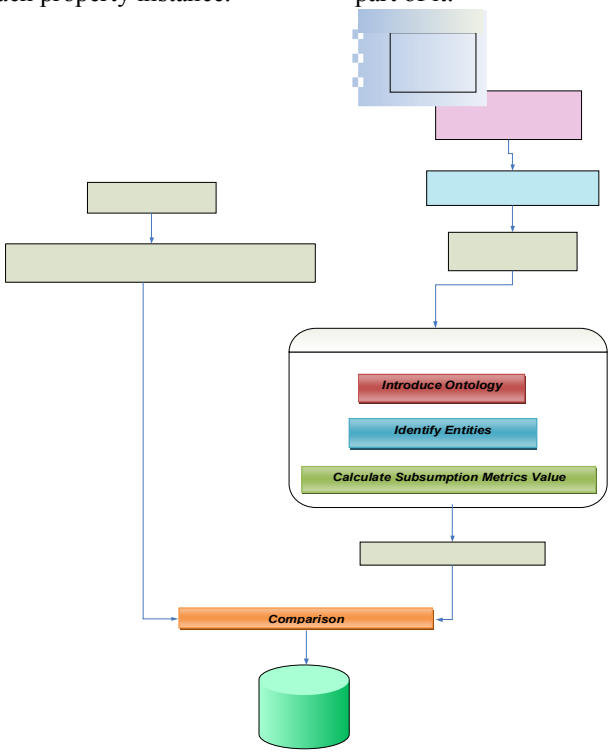


Figure 1 Subsumption metrics based information retrieval system

Following is the high level diagram of the information retrieval system with Subsumption metrics introduced.

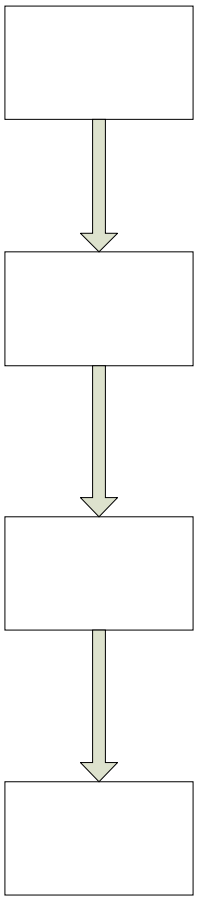


Figure 2: Information Retrieval System, High level diagram

Following are the steps needed to follow in order to get the value of SubSumption metric.

Step 1. Normalization of Ontological Metrics

Now it will be discussed that how normalization can be achieved. The basic purpose of normalization is to find the features of semantics of ontology within the document structure. [11]

The following normalization steps are applied:

1. Name all significant classes, so no unknown composite class descriptions are left
2. Name unidentified individuals
3. Turn up the subsumption hierarchy & normalize names/labels
4. Detect and find the deepest/cordial potential class or property
5. Normalize property instances

In the **first normalization** attempt is made to liberate unknown multipart class descriptions Keep in the view that it is likely to set up named classes that are insatiable. This is not the idea that the ontology becomes insatiable, but exclusively these recently introduced classes.

The **second normalization** removes unidentified individuals. It is assumed that every empty node that is of (declared or secondary) category individual requires to be regenerated with a Uniform Resource Identifier (URI) reference.

The **third normal form** turns up the subsumption hierarchy & normalizes labels. The subsumption formation now shapes a directed acyclic graph so as to signify the whole subsumption hierarchy of novel ontology.

So set of normal classes of ontology can be defined as following:

Every class that contributes in a subsumption axiom following the 3rd normalization of ontology is basically normal class of respective ontology. In case a property has more than one name, we choose one (or bring in a new name & position the equality). Each and every normal property names needed to be declared clearly to be equal to the entire additional property names they are equivalent to (so as to is, we materialize the sameness relations linking the normal property names & the non-normal ones). All incidences of non normal property names (also contained by axiom giving equality by means of normal property name, and in addition in annotation property occurrences) are substituted by normal property name. Similar is the case with the individuals. If there is a scenario that an individual has multiple names, we make a decision to introduce a normal one & describe explicitly equality or correspondence to the normal name, and then substitute every single instances of non-normal entity names by normal names (moreover inside the axiom giving sameness in the company of normal individual name, and as well as inside annotation property occurrences).

The **fourth normalization** intend towards touching the instantiations or occurrences to the deepest achievable level, since this suggests nearly all information explicitly (and obtaining instantiations and occurrences of upper levels is very economical for the reason of declared explicitness of hierarchy owing to 3rd normal form). This is not the reason that each occurrence or instance will fit in only one class, several instantiations will be essential in wide-ranging.

The **fifth normalization** concludes normalization of properties i.e. turns up property instances of symmetric & opposite properties, also removes the transitivity connection. This can be completed alike to the formation of subsumption hierarchy in 3rd normalization: following materializing every one of property instances, eliminate every unnecessary and

redundant in sub ontology, that includes just the property instances of each and every transitive properties, also the axioms giving the transitivity of respective properties. It is significant to state that normalization does not guide to canonical or basic normalized edition. Which represents that there might be a good deal of ontology that comes as a result of normalization of ontology. Frequently normalizations do not result in canonical or simple, unique results. There might be conjunctive normal forms.

Step 2. Identify Entities

Once ontology is defined in the system by normalization, System would identify the unique entities/ classes based upon their subsumption hierarchy.

Step 3. Calculate the Value of Subsumption Metrics

Subsumption metrics consist of two values i.e.

1. Number of classes or entities found in particular document
2. Frequency of occurrence of each of the class or entity found in the document based on ontology.

Step 4. Find Results

So end result of normalization would be subsumption metrics which defines number of classes identified after normalization. The more the number of classes/ entities; the less specific or irrelevant is the document or document is more generic. The lesser the number of classes/ entities the more specific is the document. By comparing the value of classes to the given search term in the query; we can find the desired document. Also if in any document there is only a single entity/ class and also its frequency is very less say one or two, these kinds of documents are treated as documents that only contain the definitions. They are relevant to user query but of less importance than the documents which have smaller the no. of entities and greater the no. of frequency of that particular entity/ class.

Proposed Information Retrieval System with improved relevancy

Following are the steps that are followed to in order to implement the design of proposed subsumption metric based information retrieval system. Here I have used GATE software in order to give the overview of proposed information retrieval system.

GATE (General Architecture for Text Engineering)

GATE [12] as architecture proposes that the modules of software systems which develops natural language could suitably be broken down hooked on variety of elements, identified as resources. Elements are reusable software lumps with distinct interfaces, and are accepted architectural form, that is being used in Sun's Java Beans & Microsoft's .Net, for instance. GATE components are specific kinds of Java Bean, and they come in three flavors:

- LanguageResources (LRs) correspond to entities e.g. lexicons, corpora or the ontologies;
- ProcessingResources (PRs) symbolize entities that are for the most part algorithmic, e.g. parsers, generators or ngram modellers; VisualResources (VRs) characterizes visualisation and editing components that play a part
- in GUIs.

These classifications can be indistinct in practice as required.

Jointly, the group of resources incorporated with GATE is known as CREOLE: Collection of REusable Objects for Language Engineering. All the resources are wraped up as Java Archive or 'JAR' files, Also several XML pattern data. The JAR & XML files are prepared so that they are available to GATE by placing them on a web server, or merely setting them in confined file space. While using GATE to build up language processing functionality for specific application, the developer exploits the development environment & the framework to assemble resources of the above mentioned three types. This possibly will involve programming, or the expansion of Language Resources e.g. grammars that are used by presented Processing Resources, or a combination of both can also be employed. The purpose of using development environment is to visualize the data structures created and consumed throughout processing, and also for the sake of debugging and performance measurement.

The GATE development environment is similar to systems akin to Mathematica for Mathematicians and JBuilder for Java programmers i.e. it gives a well-situated graphical environment for examining and expansion of language processing software.

When a suitable set of resources are developed, they could be entrenched in the target client application with the help of GATE framework.

a) Add required plug-in. In current scenario I have added Ontology Plug-In

1. Plug-in ontology

OWL file .ontology

OWL is a file required when the information in documents needs to be processed by machine/ applications, as compared to scenarios where the information only needs to be given to humans. OWL is used to (explicitly) signify the meaning of terms in glossary and also relationships among those terms. This presentation of terms/ keywords and their inter-relationships is known as ontology. OWL has

facilities to express syntax and semantics as compared to XML, RDF, and RDF-S, therefore OWL ability to characterize machine interpretable data on Web. OWL is a modification of DAML+OIL web ontology language which incorporates features and improvements DAML+OIL.

OWL has been developed to convene the requirements for a Web Ontology Language.OWL is component of W3C recommendations stack linked to the Semantic Web:

- XML gives syntax for structured web pages/ documents, but does not impose any semantic limitation on documents meaning.
- XML Schema is a used to restrict structure of XML documents and extend XML with data types.
- RDF is a data-model for objects known as resources and relations among them, gives a simple semantics model for this data.
- RDF Schema is basically vocabulary for explaining properties and also classes of RDF resources, along with semantics for generalization hierarchies of these properties and classes.
- OWL gives addition to more vocabulary for properties discripption and classes, relations between classes, cardinality, equality, properties, characteristics of properties and last but not the least enumerated classes.

2. Identify Entities

Example of Ontology

```
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:owl
="http://www.w3.org/2002/07/owl#"
xmlns="http://www.mydomain.org/african">
<owl:Ontology rdf:about="">
<owl:VersionInfo>
Server load management example version 1.2, 20 April 2010
</owl:VersionInfo>
</owl:Ontology>
<owl:Class rdf:ID="server load managemant">
<rdfs:comment>load management of server is a class</rdfs:comment>
</owl:Class>
<owl:Class rdf:ID="layer 4 load balancing">
<rdfs:comment> layer 4 load balancing is type of server load management </rdfs:comment>
<rdfs:subClassOf rdf:resource="#server load management"/>
</owl:Class>
<owl:Class rdf:ID="layer 7 load balancing">
<rdfs:comment> layer 7 load balancing is type of server load management </rdfs:comment>
<rdfs:subClassOf rdf:resource="#server load management"/>
</owl:Class>
<owl:Class rdf:ID="Global server load balancing">
<rdfs:comment> Global server load balancing is part of Server load management
</rdfs:comment>
<rdfs:subClassOf>
<owl:Restriction>
<owl:onProperty rdf:resource="#is-part-of"/>
<ow:allValuesFrom rdf:resource="# Server load management "/>
</owl:Restriction>
</rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:ID="Application acceleration">
Web Ontology Language: OWL 19
<rdfs:comment>
Application acceleration steps up application delivery
</rdfs:comment>
<owl:intersectionOf
rdf:parsetype="Collection">
<owl:Class rdf:about="#application delivery"/>
<owl:Restriction>
<owl:onProperty rdf:resource="#steps up"/>
<owl:allValuesFrom>
<owl:unionOf rdf:parsetype="Collection">
<owl:Class rdf:about="#application delivery"/>
<owl:Restriction>
<owl:onProperty rdf:resource="#is-part-of"/>
<owl:allValuesFrom rdf:resource="#Server load management"/>
</owl:Restriction>
</owl:unionOf>
</owl:allValuesFrom>
</owl:Restriction>
</owl:intersectionOf>
</owl:Class>
</rdf:RDF>
```

b) Add the document that is needed to be processed such as we can consider a text document but it is required that document must be of .txt format.

3. Apply rule in order to calculate subsumption metrics values.

JAPE – Java Annotation Patterns Engine is basically used to insert various rules in GATE plugIn. JAPE gives finite state transfer over annotations that are based on standard expressions. Regular expressions are used to character strings, but here they are applied

to a much more complex data structures. The result is in many cases matching procedure in non-deterministic: when structure in the graph being matched requires more as compared to regular automaton to recognize, the JAPE selects a substitute at random.

A JAPE grammar is comprised of set of phases. Each of phrases is composed of a set of pattern/ action rules. Also the phases run in sequence and form a flow of finite state transformers over annotations. The left hand side (LHS) of rules is comprised of an annotation pattern which have regular expression operators e.g. *, ?, +. The right hand side (RHS) is made up of annotation operation statements. Comments/ Annotations are matched on LHS of a rule and can be referred to on the RHS with the help of tags which are attached to pattern essentials.

At the start of each grammar, quite a few options can be placed:

- Control: defines the techniques of rule matching
- Debug: when it is set to true and the grammar is working in Appelt mode. If there is more than one match than conflicts will be presented.

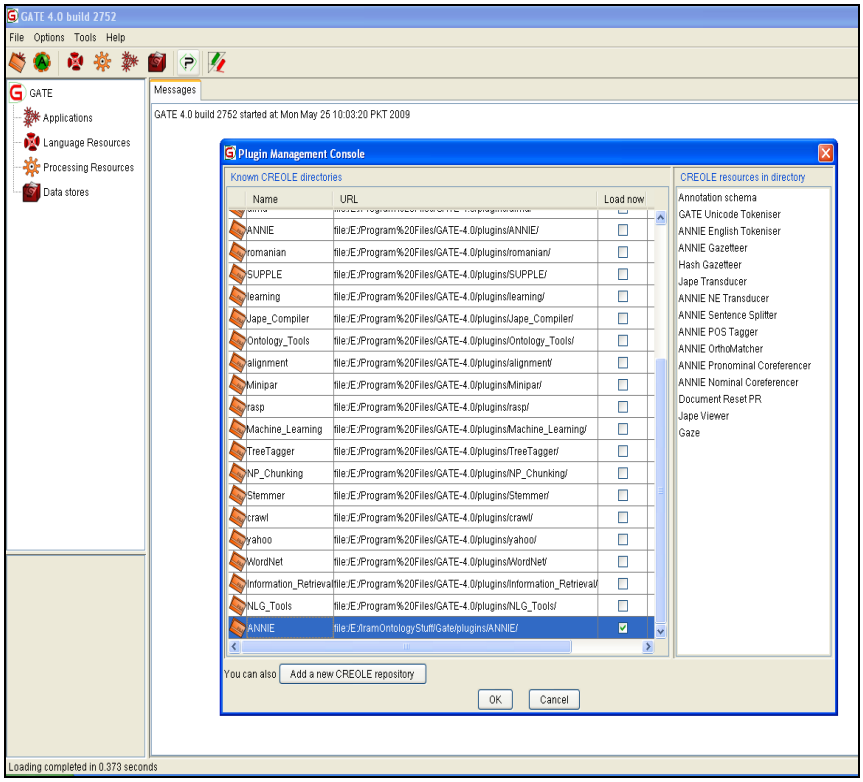
c) Input annotations are defined at the beginning of each grammar. If no comments/ annotations are clear, all annotations will be matched. There are three ways in which pattern can be defined:

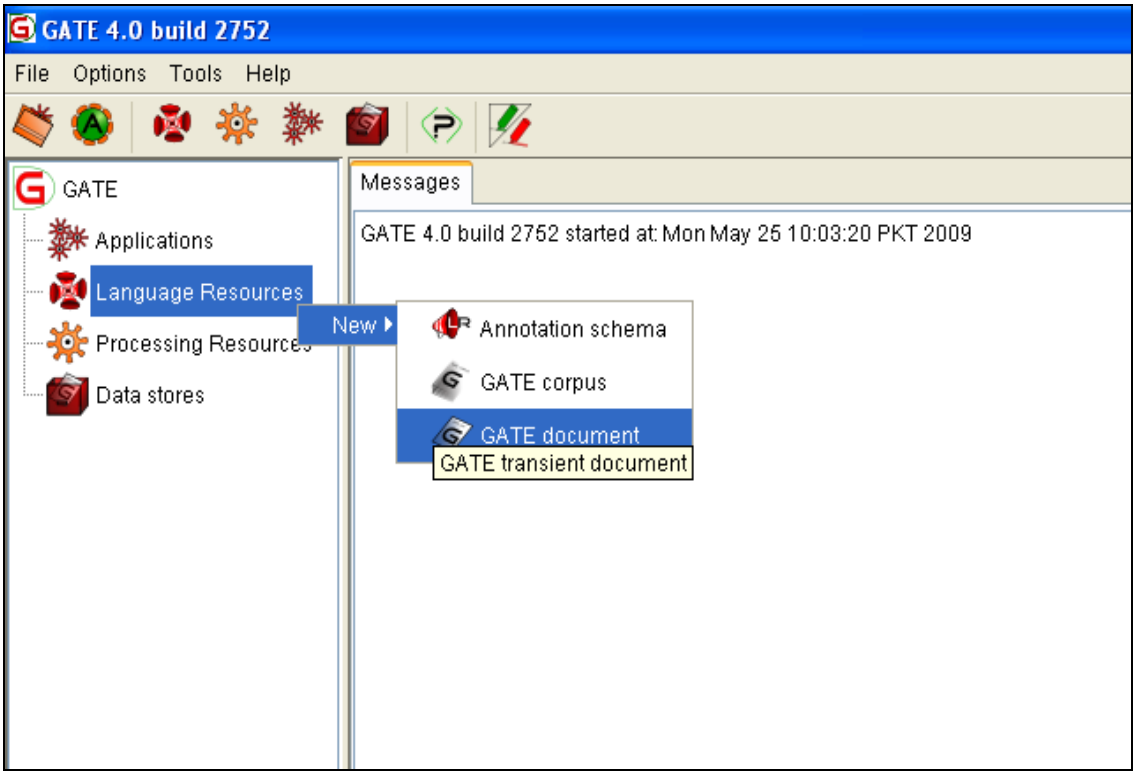
- Define a string of text, for example. {Token.string == "the"}
- Identify an annotation formerly assigned from gazetteer, lookup, tokeniser, or other module.
- give the attributes of annotation),that is . {Token.kind == string}

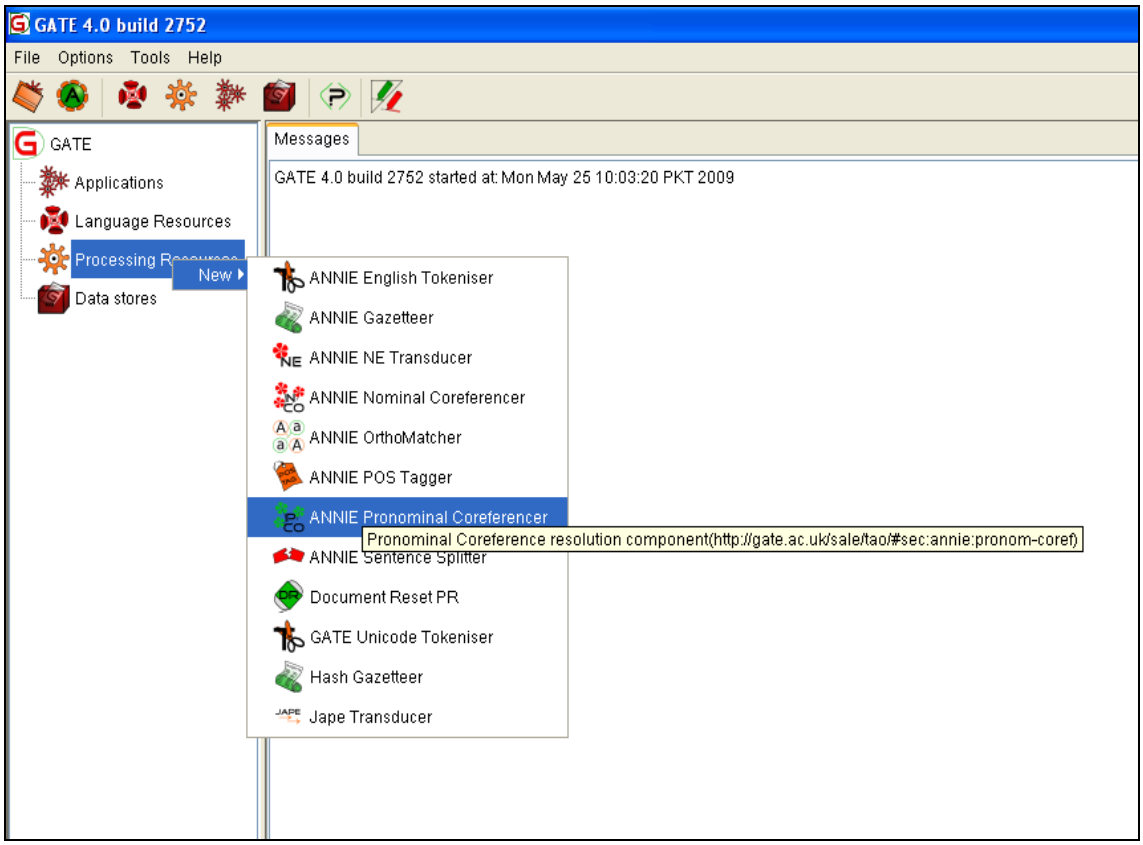
Rule : ServerLoadManagement
Priority : 1
(
{Lookup.majorType == LoadManagement}
) : loadManagement
(
{TempServer}
) : server
→

:loadManagement.LoadManagement = {rule = "ServerLoadManagement"},
:server.Server = {kind = "Server", rule = "ServerLoadManagement"}

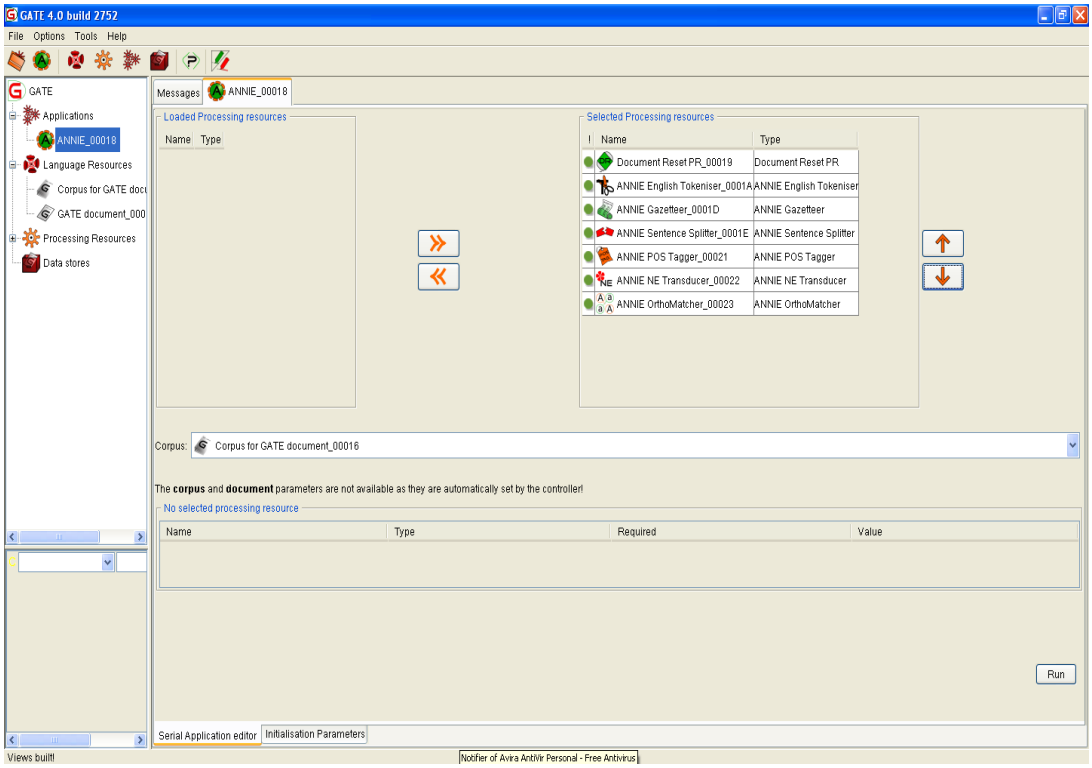
d) Add the required Processing Resources from the list.




GATE 4.0 build 2752
File Options Tools Help
GATE
Applications
Language Resources
Processing Resources
Data stores
Messages
GATE 4.0 build 2752 started at: Mon May 25 10:03:20 PKT 2009
New
Annotation schema
GATE corpus
GATE document
GATE transient document

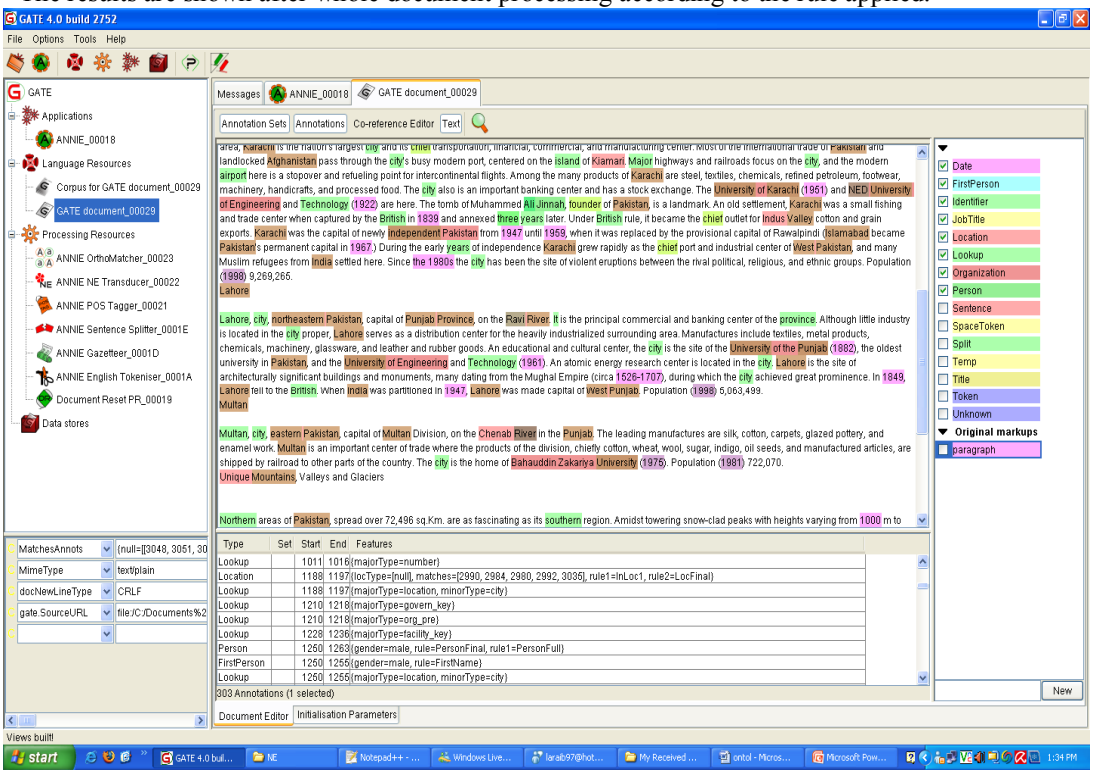

GATE 4.0 build 2752
File Options Tools Help
GATE
Applications
Language Resources
Processing Resources
Data stores
New
Messages
GATE 4.0 build 2752 started at: Mon May 25 10:03:20 PKT 2009
ANNIE English Tokeniser
ANNIE Gazetteer
ANNIE NE Transducer
ANNIE Nominal Coreferencer
ANNIE OrthoMatcher
ANNIE POS Tagger
ANNIE Pronominal Coreferencer
Pronominal Coreference resolution component(http://gate.ac.uk/sale/tao/#sec:annie:pronom-coref)
ANNIE Sentence Splitter
Document Reset PR
GATE Unicode Tokeniser
Hash Gazetteer
Jape Transducer

e) Add the application and run the Processing Resources.



4. Results

The results are shown after whole document processing according to the rule applied.



As case study five documents were processed [13], [14], [15], [16], [17], [18] .
Then the value of subsumption metrics was calculated which included no. of entities/ classes in each document and their respective frequencies. Based upon the achieved results from subsumption metrics value, relevancy ranking was made.Then results were obtained by quering the google search engine.

Finally documents were given to MS students to evaluate the document relevancy against the query. It was found that documents relevancy ranking based on subsumption metrics were more according to user's requirements as compared to the one found using traditional search system e.g. google.

Table 6.1 Subsumption metrics values of five documents

Document / Entitity:Frequency	Document 1	Document 2	Document 3	Document 4	Document 5
$E_1 : F_1$	$D_1E_1 : 2$	$D_2E_1 : 4$	$D_3E_1 : 1$	$D_4E_1 : 1$	$D_5E_1 : 1$
$E_2 : F_2$	$D_1E_2 : 3$	$D_2E_2 : 1$	$D_3E_2 : 1$	$D_4E_2 : 1$	$D_5E_2 : 5$
$E_3 : F_3$	$D_1E_3 : 1$	$D_2E_3 : 1$	$D_3E_3 : 1$		$D_5E_3 : 1$
$E_4 : F_4$	$D_1E_4 : 1$	$D_2E_4 : 1$	$D_3E_4 : 1$		$D_5E_4 : 1$
$E_5 : F_5$	$D_1E_5 : 1$	$D_2E_5 : 1$	$D_3E_5 : 1$		
$E_6 : F_6$	$D_1E_6 : 2$		$D_3E_6 : 1$		
$E_7 : F_7$	$D_1E_7 : 1$		$D_3E_7 : 1$		
$E_8 : F_8$	$D_1E_8 : 1$				

Highlighted row shows the frequency of desired entity in respective documents.
As shown in the table; Document 5 is the most revelant document retrieved against the user query when subsumption metrics was introduced in the system. While in case of simple search Document 1 was graded as the most promising result, Document 2 as 2nd and so on upto Document 5 as shown in the table 1. But by implying proposed system Document 1 was ranked at the 2nd position, Document 4 at 3rd , Document 3 at 3rd , Document 2 at 4th and Document 5 at 1st position. This shows by introducing subsumption merics in the Information Retrieval systems, quality of retrieval can be improved greatly.

References

[1] Web-Based Evolutionary and Adaptive Information Retrieval, Ibrahim Kushchu, Member IEEE,IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION, VOL. 9, NO. 2, APRIL 2005

[2] Criteria for Evaluating Information Retrieval Systems in Highly Dynamic Environments, Judit Bar-Ilan, School of Library, Archive and Information Studies The Hebrew University of Jerusalem

[3] Semantic Metrics, Conceptual Metrics, and Ontology Metrics: An Analysis of Software Quality Using IR-basedSystems, Potential Applications andCollaborations, Letha H. Etzkorn Computer Science Department University of Alabama in Huntsville Huntsville, AL 35899 USA

[4] G.Aghila Asst.Professor, Department of CSE & IT, "Ontology-based Web Crawler", Pondicherry Engineering College, Pondicherry, India.

[5] Software Metrics
http://en.wikipedia.org/wiki/Software_metric

[6] Information Retrieval
www.answers.com/topic/information-retrieval

[7] Ontology
http://www-ksl.stanford.edu/kst/what-is-an-ontology.html

[8] Christopher B. Jones , Harith Alani , Douglas Tudhope, "Geographical Information Retrieval with Ontologies of Place", 2001 ,Spatial Information Theory, LNCS 2205

[9] Muller HM, Kenny EE, Sternberg PW, "Textpresso: an ontology-based information retrieval and extraction system for biological literature", Division of Biology and Howard Hughes Medical Institute, California Institute of Technology, Pasadena, California, USA. Sep 2004

[10] Pablo Castells, Miriam Fernandez, David Vallet, "An Adaptation of the Vector-Space Model for Ontology-Based Information Retrieval", IEEE Transactions on Knowledge and Data

Engineering. Volume 19, Issue 2 (February 2007)
[11] "The Semantic Web: Research and Applications", Chapter 6, " How to design better ontology metrics" by D Vrandecic, Y Sure, ESWC 2007, Innsbruck, Austria, June 3-7 , ISSN0302-9743 (Print) 1611-3349 (Online), Volume 4519/2007
[12] GATE (General Architecture for Text Engineering)
http://gate.ac.uk/
[13] Netscaler
http://http.cdnlayer.com/softlayerweb/NetScaler_SL_WP.pdf
[14] Citrix Server Load Management (Part 1)
http://www.virtualizationadmin.com/articles-tutorials/terminalservices/general/citrix-presentation-server-load-management-part1.html
[15] Citrix Server Load Management (Part 3)
http://www.virtualizationadmin.com/articles-tutorials/terminalservices/general/citrix-presentation-server-load-management-part3.html
[16] Server Load Management Solutions
http://www.dataracks.net/services_loadmanagement.php
[17] Citrix Load Mamagement
http://support.citrix.com/proddocs/index.jsp?topic=/xenapp6-w2k8/lm-wrapper-v2.html
[18] Netscaler VPX
https://secureportal.citrix.com/MyCitrix/login/EvalLand.aspx?downloadid=1857216&LandingFrom=1238.

1/3/2011