

# Explaining The Search Results Using Ontology

Gayathri T

<sup>1</sup>(Computer Science, Anna University, Coimbatore, India, gayucs13@gmail.com)

**Abstract**— The Semantic Web is a development of the current web in which it incorporates semantic substance in pages. The primary objective of semantic web is to push the nature of the current web by transforming its substance into machine reasonable structure. Consequently, the development of semantic web is to have semantic level data in the web. These days, individuals use diverse decisive word based web search tools to discover the significant data they require from the web. Be that as it may a hefty portion of the words are polysemous. At the point when these words are utilized to question a web search tool, it shows the Search Result Records (Srrs) with diverse implications. The Srrs with comparable implications are gathered together and commented. Semantic Annotation is the procedure of adding the semantic metadata to web assets. Thus the gathered Srrs are commented and create an outline which portrays the data in Srrs. Be that as it may the programmed semantic annotation is a huge test in the semantic web. Here metaphysics and information based representation are utilized to explain the website pages.

**Keywords**— Semantic Web , Semantic Annotation, Ontology

## 1. INTRODUCTION

A substantial volume of information accessible in the web is in unstructured or semi-organized organization. Consequently the information in the web will be proposed to be skimmed by people, not by machines. Most of the clients use catchphrase based web crawlers to recover data from the web. Since there is an emotional increment in the web pages pivotal word based web crawlers can't help the clients to discover the most applicable and exact data in a productive way. The semantic web is an enlargement of the World Wide Web in which the data is given with generally characterized significance. The thought of semantic web is leave the assignments and choices to machines, so that machines and individuals to work in co-operation. This is relevant by adding learning to the data accessible in the web by understanding which dialect the machine can comprehend and present the product executors that equipped to process the data.

The machines in the semantic based environment have a typical understanding from metadata tags and convey to one another. To speak with one another there is a need for archive which can characterize all the ideas. In semantic web, philosophy goes about as an imparted store. A metaphysics is an express, formal detail of an imparted conceptualization. This implies that a philosophy depicts the

Data in a machine reasonable manner. As it were, cosmology is to be recognized as a device which characterizes extra significant tags to site pages and makes them accessible to be utilized by programming executors and different provisions.

In the semantic based environment the assets are open by people as well as to mechanized methodologies. The computerization of undertakings hoists the status of the web from machine-decipherable to machine reasonable. The information on the web is unequivocally translated by the

product operators instead of verifiably deciphered by people. So as to understand this, there is a need to cohort metadata with the assets. One such component to copartner the metadata with the assets is annotation.

Whatever remains of the paper is sorted out as takes after ; area 2 depicts what is semantic annotation. Segment 3 talk about the getting to the web information utilizing semantic annotation. Segment

4 shows the annotation system. Area 5 depicts the conclusion and what's to come work.

## 2. SEMANTIC ANNOTATION

The procedure of adding notes or remarks to any asset is known as annotation. In the area of web, annotation methods including the data like notes, remarks or rundown to an existing asset like reports without changing their unique substance. These annotations are sharable over the system. The notes, remarks and synopses will be not sharable yet it will be helpful. Semantic annotation methods calculator justifiable metadata to assets. Essentially, there are three sorts of annotation, specifically: 1) casual annotations don't have formally characterized constituents and not machine-justifiable 2) formal annotations are machine-discernable yet don't utilize ontological terms 3) ontological annotations have formally characterized constituents and use ontological terms that will be socially acknowledged and caught on. Annotation is possible physically and naturally. Manual annotation might be finished effortlessly utilizing the writing apparatuses, for example, Semantic Word which gives the incorporated environment to at the same time composing and explaining the content. Manual annotation is an exorbitant process[3]. There are two purposes behind an ineffectualness of manual annotation in particular :

- Due to extensive sum of errands and assets it will be time devouring.

- Different sentiments can bring about conflicting information.

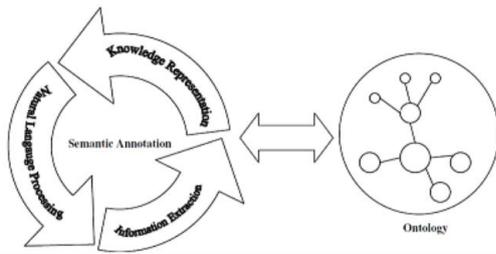


Figure 1: A review of semantic annotation and viable innovations

The key component of semantic annotation is metaphysics. Ontological structures give extra esteem to semantic annotation. Records are commented with idea cases from the Knowledge Base by making examples of the Annotation class, accommodated this reason. Annotation has two social properties specifically, occurrence and archive, by which ideas and archives are connected together.

### 3. ANNOTATION OF WEB PAGES

In this area we will see about the distinctive sorts of annotation of substance accessible in the web.

#### A. Semi-programmed Semantic Annotation

This self-loader semantic annoTo get the fundamental structure, it utilizes lightweight powerful parse and to deduce the source markup of essential certainties, the conversion principles use vocabulary and structural examples;

- To get the basic structure, it uses lightweight robust parse and to infer the source markup of basic facts, the transformation rules use vocabulary and structural patterns
- Externalization of realities to database for derivation;
- In request to construe the semantic markup of configuration certainties, change standards use deduction and structural examples, then the stamped up source will be prepared for outline mindful changes.

Nadzeya Kiyavitskaya et.al outlined a device that performs semantic annotation. The text based reports and a calculated Markup Mapping Expounded reports Database Archives produced utilizing the substances of the scheme.the workflow has two primary stages:

- first stage comprises of lightweight parsing and semantic markup of fundamental substances of an applied area and organizing the dialect.
- second stage will be the externalization of the realities to database, which can then be utilized via web search tool for questions.

#### B. Automatic Semantic Annotation

Fernando Gomez illustrated the programmed semantic translation of sentences. Semantic elucidation is performed for deciding the importance of the verb, or, verb predicate,its semantic parts, assistants, the connection of post verbal prepositional expressions, and likewise for a set number of normalizations. The faculties of things are additionally determined. Anyhow, compound things are not determined, unless the pair of things exist in the Wordnet vocabulary. They have given the semantic understanding to a portion of the deverbal nominalizations in the writings if the nominalization in our database is inadequate.

The semantic mediator does not resolve the structural uncertainty, however it recognizes the supplements of verbs including

Easy supplements. The errand of semantic translator is to append the Pps and any post verbal identifiers. So as to test the predicates, there is no need of parsing the complex sentences holding relativization, coordination and/or subordination, free relatives and so forth. At the same time the main thing required to go to the semantic mediator is all the supplements and postverbal modifiers in the proviso.

The yield of the semantic mediator is not difficult to peruse. The sentence is recorded first which is trailed by the progression of all super-predicates for the verb predicates in the sentence. The linguistic connection is recorded first which is trailed by the semantic part for the syntactic connection in the structure < (semantic – part) >. The calculation chooses the predicate that boosts semantic parts, or demonstrates the most linguistic relations. The project prints the predicate took after by the WN verbs faculties and the expression "backed by 'a-number' Srs," (where 'a-number' remains for a cardinal number), which demonstrates the amount of linguistic relations (Srs) that have acknowledged semantic parts. On the off chance that the yield for that sentence has more semantic parts than the number recorded in the expression "upheld by," those semantic parts are adjuncts[2].

#### C. Annotating the Web Database

Benjamin Donz and Dietmar Bruckner proposed a new answer for making web-database open for semantic executors. Instead of adding annotation to a web page or distributed the substance independently, the components of a site page are clarified and distributed in an outside record or even an outer site which might be utilized as direction manual to figure out how to translate the website pages and to access the Calculated model Semantically Interfaced reports Inquiry Motor databases. Since these annotations might be distributed independently from the first page, anybody can clarify any site and make it available for semantic executors.

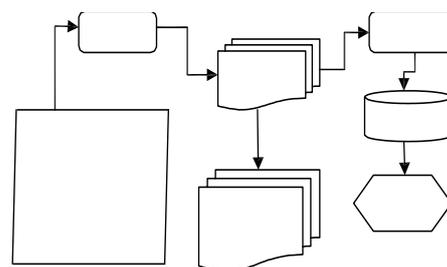


Figure 2: The workflow of semantic annotation

plan are the information of the framework. A piece of the existing area metaphysics will be the calculated plan. The tags are

- First phase consists of lightweight parsing and semantic markup of basic entities of a conceptual domain and structuring the language.
- Second phase is the externalization of the facts to database, which can then be used by search engine for queries.

### B. Automatic Semantic Annotation

Fernando Gomez explained the automatic semantic interpretation of sentences. Semantic interpretation is performed for determining the meaning of the verb, or, verb predicate, its semantic roles, adjuncts, the attachment of post verbal prepositional phrases, and also for a limited number of normalizations. The senses of nouns are also resolved. But, compound nouns are not resolved, unless the pair of nouns exist in the WordNet lexicon. They have provided the semantic interpretation for some of the deverbal nominalizations in the texts if the nominalization in our database is incomplete.

The semantic interpreter does not resolve the structural ambiguity, but it identifies the complements of verbs including casual complements. The task of semantic interpreter is to attach the PPs and any post verbal identifiers. In order to test the predicates, there is no need of parsing the complex sentences containing relativization, coordination and/or subordination, free relatives etc. But the only thing needed to pass to the semantic interpreter is all the complements and postverbal modifiers in the clause.

The output of the semantic interpreter is easy to read. The sentence is listed first which is followed by the hierarchy of all super-predicates for the verb predicates in the sentence. The grammatical relation is listed first which is followed by the semantic role for the grammatical relation in the form

< (semantic – role) >. The algorithm selects the predicate that maximizes semantic roles, or explains the most grammatical relations. The program prints the predicate followed by the WN verbs senses and the phrase “supported by ‘a-number’ SRs,” (where ‘a-number’ stands for a cardinal number), which indicates the number of grammatical relations (SRs) that have realized semantic roles. If the output for that sentence has more semantic roles than the number listed in the phrase “supported by,” those semantic roles are adjuncts[2].

### C. Annotating the Web Database

Benjamin Donz and Dietmar Bruckner proposed a new solution for making web-database accessible for semantic agents. Instead of adding annotation to a web page or publishing the content separately, the elements of a web page are annotated and published in an external file or even an external website which can be used as instruction manual to find out how to interpret the web pages and to access the

- Conceptual model
- Semantically

Linked documents Query engine databases. Since these annotations can be published separately from the original page, anyone can annotate any website and make it accessible for semantic agents.

The outside annotations can essentially hold the same data as inline annotations, yet need to incorporate a reference to a html component in the first document. In the event that the page itself can't be changed (e.g. when it has a place with somebody else), this makes it troublesome to comment distinct words in a section of content, however still permits to explain structural components like tables, fields, joins, catches and others. rather than most existing requisitions for the Semantic Web, the errand of concentrating data from web databases proposed for human utilize includes deciphering information, as well as calling capacities on the page, for example, submitting an inquiry to begin the pursuit or calling a capacity to recover the following 10 effects on an effect page. The principal fundamental refinement is hence a detachment of capacities and information. For capacities, the page depiction must hold the purpose of the capacity, the component to use for executing the capacity (catch or connection) and a rundown of partnered parameters including the fields where they ought to be entered[1].

### D. Clarifying Srrs from Web Database

An expansive parcel of the web is database based (ie), information returned in the list items are encoded from the underlying organized databases. This sort of internet searchers are known as Web Databases (WDB). An ordinary item came back from WDB holds numerous Srrs. Every SRR holds numerous information unit , where every information unit speaks to a true element.

Information Unit is a bit of content that semantically speaks to one idea of a substance. It varies from a content hub which is a succession of content encompassed by a couple of HTML tags. Tragically, the information unit in the outcome page is not given semantic marks. Yiyao Lu et.al., proposed how to naturally dole out names to the information units inside the Search Result Records came back from Wdbs. The programmed annotation methodology comprises of three stages to be specific : 1) Alignment Phase , 2) Annotation Phase and 3) Annotation Wrapper Generati

#### 1)Arrangement Phase

In this stage, first recognize all the information units in the Srrs and afterward compose those information units into distinctive gatherings with each one gathering relating to an alternate idea. As an aftereffect of this stage with every segment holds information units of the same idea over all Srrs. Gathering information units of the same semantic can help distinguish the basic examples and characteristics around these information units. These normal characteristics are the support of annotators.

The comparability between two information units d1 and d2 is dictated by the weighted aggregate of the five similitudes between them. The five similitudes incorporates the

accompanying : Data Content Similarity (Simc), Presentation Style Similarity (Simp), Data Type Similarity (Simd), Tag Path Similarity (Simt) and Adjacency Similarity (Sima).

This arrangement strategy additionally needs the closeness between two information units bunches G1 and G2 where every bunch will be an accumulation of information units. The similitude between gatherings G1 and G2 is to be the normal of the similitudes between every information unit in G1 and each information unit in G2.

The arrangement calculation comprises of four steps : 1) Merge content hub 2) Align content hub 3) Split (Composite) content hub and 4) Align information units.

2)Annotation Phase:

The annotation stage comprises of different essential annotators with every abusing one sort of characteristics. Every essential annotator is utilized to process a mark for the information units inside their gathering and a likelihood model is received to focus the most suitable name for each one gathering. At the end of this stage, a semantic mark Lj is allocated to every segment. The six essential annotators are utilized to name the information units, with each of them recognizing a unique kind of examples/characteristics. The annotators are :

- Table Annotator (TA)
- Query-Based Annotator (QA)
- Schema Value Annotator (SA)
- Frequency-Based Annotator (FA)
- In-Text Prefix/Suffix Annotator (IA)
- Common Knowledge Annotator (CA)

A solitary annotator is not equipped for completely marking all the information units on diverse effect pages. The relevance of an annotator is the rate of the ascribes to which the annotator could be connected. The consequence of all fundamental annotators could be consolidated so as to accomplish higher rate of information units.

3)Annotation Wrapper Generation Phase

Once the information units in the effect page have been explained, these clarified information units are utilized to build an annotation wrapper for WDB. At that point when the same Srrs are recovered from the same WDB could be explained utilizing this wrapper rapidly without reapplying the entire annotation process.

Each one clarified gathering of information units compares to a quality in the Srrs. The annotation wrapper is a portrayal of the annotation tenets for all the qualities on the consequence page.

After the information unit gatherings are clarified, they are sorted out dependent upon the request of its information units in the first Srrs. Think about the ith bunch Gi. Examine every information unit in the gathering Gi in both forward and retrogressive to get the prefix and postfix of the information unit. The examining of information unit is ceased when it experiences a legitimate information unit relegated with a serious name. At that point think about all the prefix of the information units in Gi, so that a typical prefix is gotten which

might be imparted by all these information units in Gi. Also look at all the addition of the information units in Gi, so that a typical postfix is gotten for all these information units.

To utilize the wrapper to expound another outcome page, for every information unit in a SRR, the annotation principles are connected on it one by one dependent upon the request they show up in the wrapper. On the off chance that this information unit has the same prefix and postfix as specified in the guideline, the standard is matched and the unit is named with the given name in the principle. In the event that the separators are specified, they are utilized to part the unit, and labeli is allotted to the unit at the position unitindexi.

E. Semantic Techniques to Access Web Data

These days, individuals use diverse magic word based web search tool to recover the important data from the web. Since the words are polysemous recovering the correct data will be troublesome. So, Raquel Trillo et.al. utilized a semantic method to find the significant information[5]. The building design of the framework comprises of two principle steps.

Step 1: Discovering the semantics of client decisive words – The principle point of this methodology is to discover all the conceivable implications of the client question. The client question holds more than one catchphrases in the inquiry. For every catchphrase in the inquiry, it discovers all the conceivable magic word faculties are positioned and explained those essential words by utilizing the connection of the catchphrases. The score of the watchword is a quantitative procedure and this reaches from 0 to 1. The conceivable feeling of the catchphrase with the most elevated score is the particular case that is proposed by the client. The framework need to discover all the conceivable implications for every pivotal word ki (called conceivable

pivotal word faculties and signified as {si , Si ,... , Si }}, so semantic scores acquired in the Disambiguation of User Keywords stage (2) the rate of hits arranged in the classification, and (3) the position of the first hit in that class in the positioning gave by the web crawler.

4. USING THE TEMPLATE

Quratulain Rajput et.al., displayed a semantic annotation skeleton which is equipped for concentrating pertinent data from unstructured, ungrammatical and confused information sources. BNOSA (Bayesian Network and Ontology based

1 2 u depictions will be needed. The semantic portrayals are gave by diverse lexical sources such as thesauri, lexicons, ontologies, and so forth.

Step 2: Semantics-guided information recovery – The objective of this step is to gathering the Srrs returned by a customary web search tools like Google or Yahoo. The classifications are positioned as stated by the enthusiasm of the clients. There are four steps included in this procedure. They are memory of hits, lexical annotation of hits, order of hits and exhibiting the outcome to the client.

At the outset, the hits returned by the conventional web crawler are shown dependent upon the positioning calculation

utilized. The request of the hits (i.e., the positioning of the outcomes given by the internet searcher utilized) relies upon the particular procedures utilized by the specific web search tool for that errand and its different inside parameters. Then, the hits returned by the seek

motor are given as information to the following stage incrementally, in

squares of hits of a certain size. Thusly, number of new hits could be recovered while the first pieces are continuously transformed.

In this second stage, each one hit acquired from the past stage (holds a title, a URL and a piece) is consequently explained lexically. Accordingly, firstly, each one hit  $H_j$  experiences a purging procedure where stopwords are separated out (making the sifted hit  $H'_j$ ). After that, the applicable expressions of the title and the piece of each one sifted hit  $H'_j$  will be recognized to perform a lexical annotation of the hit. A lexical annotation is a bit of data added to a term that alludes to a semantic information asset, for example, a lexicon, a thesaurus, a semantic organize, or any other asset which communicates, either certainly or expressly, a general metaphysics of the world or a particular area. In this third stage, the hits (effectively expounded as a aftereffect of the past procedure) are assembled into classes by acknowledging their lexical annotations. Firstly, the framework characterizes the conceivable classifications that will be set to be acknowledged. At that point, squares of hits are ordered. The potential classes are characterized by acknowledging all the conceivable combos of conceivable magic word faculties of the information catchphrases (i.e., the cartesian result of the conceivable sense sets of the client watchwords).

At long last, the outcomes of the Categorization of Hits stage are exhibited to the client. The framework shows, in diverse tabs or class interfaces, the classifications acknowledged that hold hits. The request of the tabs or class connections relies upon the likelihood that the comparing classification speaks to the semantics that the client had at the top of the priority list when he/she composed his/her inquiry. Along these lines, to rank the classes, three components are viewed as: (1) the Semantic Annotation) structure utilizes philosophy to conceptualize an issue space and to concentrate information from the given corpora, and BNOSA is utilized to resolution the clashes and to foresee the missing information. For every single class and for its properties BNOSA characterizes the semantics in the annotation area of a philosophy. This sort of information is alluded to as connection words and the information sorts of the traits/classes are utilized to recover the significant data from a site page. Throughout the procedure of recovering data, if there are more than one worth is concentrated for a quality or if the worth is not discovered then the Bayesian systems are utilized to resolve the clashes and missing quality forecast. BNOSA can interface the metaphysics with the comparing Bayesian arrange powerfully keeping in mind the end goal to concentrate information from a specific space makes it exceedingly extensible.

BNOSA performs the semantic annotation process, which holds three steps namely[6] :

- For information representation and Bayesian system taking in creating a cosmology
- With the help of metaphysics and Bayesian system separate the applicable data
- For the removed data semantic tagging is carried out

Information representation and BN taking in BNOSA concentrates on a specific issue area in the type of a philosophy which will store the data about the information sorts of all the applicable classes and additionally their setting data. Physically the philosophy could be developed and utilized within the methodology of data extraction, which alluded to as extraction metaphysics.

There are two issues to be tended to when removing the data from unstructured, ungrammatical and indistinguishable information sources. a) the area of applicable information on the website page ought to be distinguished and b) characterize the examples for concentrating such information. In request to tackle this issue, BNOSA utilizes setting words connected with every idea in the remark segment of a metaphysics. By utilizing the information sorts characterized for all the traits, BNOSA characterizes the examples or manages as customary outflows.

#### A. Data Extraction

There are two stages in the data extraction. 1) employments metaphysics to concentrate the data and 2) utilization Bayesian system to discover the most proper quality for an ascribe or to foresee the absent for a property.

Stage I : Extraction by means of cosmology – In this stage, BNOSA recovers the data joins from unequivocally gave Urls. In an iterative way, the substance of each one connection are utilized to concentrate the information. BNOSA likewise remove the data utilizing Knowledge put away within a metaphysics and create tenets dependent upon the information sorts of an trait. If there will be a match then it recommends that the relating information worth is available in the area of this essential word and it could be looked utilizing predefined examples/standards.

Stage II : Conflict determination and missing quality forecast In this stage, all the missing and clashing worth issues are unraveled utilizing Bayesian systems. On the off chance that a quality has different values then all the non-missing/non-clashing are utilized within discovering the back minor likelihood of the property. The worth which has the most noteworthy back quality is chosen as an

proper worth for that trait. So also for missing quality

of a trait all the non-missing and non-clashing qualities are utilized to find the back minimal likelihood. The trait with the most astounding back likelihood swaps the missing quality.

#### Semantic tagging

After effectively finishing the above two steps, BNOSA populates the extracted information in an ontology. Then, by using RDF/OWL adding the semantics to the extracted information, so that the information become machine-understandable.

#### 5. CONCLUSION

The seek results returned by the conventional web search tools are clarified by utilizing lexical assets, for example, thesaurus and cosmology. Semantic annotation naturally distinguishes the information things that are connected and likewise recognizes the relationship between the information. The expounded hits could be assembled by the semantic of the information which has a browsable synopsis portrays what information the sorted hit holds. By envisioning that outline the clients can view the hits holding the applicable data of the given pivotal words.

To enhance the indexed lists effectiveness, a set of semantic systems to gathering the outcomes acquired from the conventional

web search tools. Need to think about the alluring characteristics distinguished in the writing for this sort of frameworks: pertinence, browsable rundowns, cover, scrap tolerance, speed and incremental. Additionally, the proposed structure could be gathered utilizing distinctive parts for some of the proposed undertakings, for example, diverse disambiguation calculations or sense finding methodologies that utilize diverse sorts of lexical information sources.

Subsequently, have adjusted our work in different ranges, for example, philosophy matching and probabilistic word sense disambiguation for information mix frameworks to the connection of web data recovery. Additionally, current methodologies don't manage compound things, (for example, "creditcard") and formal people, places or things, (for example, "Towerbridge"), that could show up in the title and/or the scrap of the recovered hits. In this way, systems that help the disambiguation of compound things and named substance distinguishment methods will be embraced. Specifically, we will think about the likelihood to apply or adjust the methodology for compound things translation that we proposed in the setting of pattern matching..

#### REFERENCES

- [1] Benjamin Donz., Dietmar Bruckner (2012), "External Semantic Annotation of Web-Databases", IEEE Digital Library, pp 841-845.
- [2] Fernando Gomez (2006), "Automatic semantic annotation of texts", University of Central Florida, Orlando, FL 32816.
- [3] Hamed Hassanzadeh and Mohammad Reza Keyvanpour (2011), "A machine learning based analytical framework for semantic annotation requirements", International Journal of Web and Semantic Technology, Vol.No.2, pp 27-38.
- [4] Nadzeya Kiyavitskaya., Nicola Zeni., James R.Cordy., Luisa Mich and John Mylopoulos (2006), "Semi-Automatic Semantic Annotation for WebDocuments".
- [5] Raquel Trillo., Laura Po., Sergio Ilarri., Sonia Bergamaschi and Eduardo Mena (2011), "Using semantic techniques to access web data ", Elsevier Information Systems, Vol.No.36, pp 117-133.
- [6] Quratulain Rajput and Sajjad Haider (2011), "BNOSA: A Bayesian network and ontology based semantic annotation framework", Elsevier Web Semantics: Science, Services and Agents on the World Wide Web, Vol.No.9, pp 99-112.
- [7] Yiyao Lu., Hai He., Hongkun Zhao., Weiyi Meng and Clement Yu (2013), "Annotating search results from web databases ", IEEE Transaction on Knowledge and Data Engineering, Vol.No.25, pp 514-527.