Development of ontology from Indian agricultural e-governance data using IndoWordNet: a semantic web approach

Bhaskar Sinha, Somnath Chandra and Megha Garg



Bhaskar Sinha, Somnath Chandra and Megha Garg are all based at Department of Electronics & Information Technology, Ministry of Electronics & IT, Government of India, New Delhi, India.

Abstract

Purpose – *The purpose of this explorative research study is to focus on the implementation of semantic Web technology on agriculture domain of e-governance data. The study contributes to an understanding of problems and difficulties in implantations of unstructured and unformatted unique datasets of multilingual local language-based electronic dictionary (IndoWordnet).*

Design/methodology/approach – *An approach to an implementation in the perspective of conceptual logical concept to realization of agriculture-based terms and terminology extracted from linked multilingual IndoWordNet while maintaining the support and specification of the World Wide Web Consortium (W3C) standard of semantic Web technology to generate ontology and uniform unicode structured datasets.*

Findings – *The findings reveal the fact about partial support of extraction of terms, relations and concepts while linking to IndoWordNet, resulting in the form of SynSets, lexical relations of Words and relations between themselves. This helped in generation of ontology, hierarchical modeling and creation of structured metadata datasets.*

Research limitations/implications – *IndoWordNet has limitations, as it is not fully revised version due to diversified cultural base in India, and the new version is yet to be released in due time span. As mentioned in Section 5, implications of these ideas and experiments will have good impact in doing more exploration and better applications using such wordnet.*

Practical implications – *Language developer tools and frameworks have been used to get tagged annotated raw data processed and get intermediate results, which provides as a source for the generation of ontology and dynamic metadata.*

Social implications – *The results are expected to be applied for other e-governance applications. Better use of applications in social and government departments.*

Originality/value – *The authors have worked out experimental facts and raw information source datasets, revealing satisfactory results such as SynSets, sensecount, semantic and lexical relations, class concepts hierarchy and other related output, which helped in developing ontology of domain interest and, hence, creation of a dynamic metadata which can be globally used to facilitate various applications support.*

Keywords *Information technology, Semantic web, Information retrieval, Domain ontology, Artificial intelligence, Domain concept*

Paper type *Research paper*

Received 30 October 2014
Revised 6 November 2014
Accepted 7 November 2014

The authors' acknowledge Department of Electronics and Information Technology, Government of India, DEITY for: substantial infrastructural support and assistance for providing resource and utility to carry out this research work; taking initiative and placing discussions on the table with researchers and professors of IIT Delhi and IIT Kanpur as a necessary task of the nation.

1. Introduction

Agriculture is one of the most important domain of any country on which people's livelihood, food and basic needs depend. India is one of the countries whose dependency on agriculture and its related activity is mostly sought. In India, it contributes 16 per cent gross domestic product and provides employment to 52 per cent of the Indian population (Bhall and Singh, 2010). A country's population and agriculture is closely related, according to Food and Agriculture Organization of the United Nations (FAO) (Keita *et al.*, 2010), and every country should make their strategy accordingly to reduce the cost of generation of census data and manage the resources well. Because of India's vast, rich and varied agricultural produces and its allied products, distribution of products and service to access

"So, the approach for automatic extraction of metadata from web content entails ontologically linked and significantly domain-specific enriched with facts and information."

remains untouched in some parts of the country. In India, naturally six seasons fall in a year, but in terms of agricultural favor, two major agricultural seasons are known such as in *spring* (known as *Kharif*) and another in *autumn* (known as *Rabi*) (Bahuguna, 2012). Also, because of its traditional multicultural and multilingual states, where 22 languages are spoken officially in different parts of the country, India faces huge challenges in the agriculture sector in terms of resource utilization, weather conditions, specially information communications between the farming community and its stakeholders, supply and distribution of agricultural products and produces, etc., which subsequently generates a chain of problems.

To tackle these challenges, India is undertaking several initiatives and is specially focusing on this sector to minimize the gap of communication using technology and strategic planning. No doubt, there are a large number of complications in bringing all these to happen together, such as the language problem, agriculture-related records, geographically dispersed unconnected information and data and, above all, weather-related data that affects the crops and its allied products, which directly or indirectly affects the financial health and the social growth of the country. Appropriate access to information and service delivery is critical and of utmost important in this sector in terms of time-bound farm activities involved in all stages of the crop cycle (Kumar, 2010).

A big role of information communication and technology (ICT) is important; and to cater with its relevancy of information and its structure or more reluctantly reliable information and unique data format, definitely supports further exploration and innovation towards the progress of the society. Correlating with diverse subdomains and creating ontologies for agriculture domain is a challenging task. Moreover, our focus is on reducing the task of manual extraction of significant valuable agricultural information and data (spread across various URLs) so as to enhance the reusability and sharability among its subdomains and other dependent domains. Apart from that uniformity, is the objective of semantic Web architecture to make any type of facts and information data clubbed into a unique reusable and sharable format.

So, the approach for automatic extraction of metadata from web content entails ontologically linked and significantly domain-specific enriched with facts and information. The aim of the undergoing projects, Web internationalization standard and the World Wide Web Consortium (W3C) India initiative, is to incorporate Indian official languages in agricultural domain while linking with IndoWordNet as a source of information vocabulary (as indicated by W3C India, 2014) for the implementation of Indian languages unicode standard using *Semantic Web Technology* to get the benefits of linked object data in unique Resource Description Framework (RDF)/Web Ontology Language (OWL) format. The goal of these initiatives focuses on:

- wider access of resources through e-government services with localization support to reach all sections of the society; and
- enhance adoption of W3C standards in Web technology in e-government services.

Implementation of Semantic Web Technology in e-governance data is to enhance and upgrade the legacy system to make resource more web dynamic, reusable, sharable and easy available to its service users/stakeholders at anytime from anywhere. In this regard, the role of ICT using semantic Web technology is well suited to facilitate the

services and manage resource distribution through proper management of knowledge base. Semantic Web's support (as indicated by Ministry of Agriculture, India, 2006, for National Commission on Farmers) for linked data is one of the solutions to integrate all datasets and information spread across to different URL sites, assures the availability of information and data at any time any were in the Web. Semantic Web's metadata framework (Berners-Lee, 2007) supported data format helps in generating hierarchal ontology development for any domain of interest. Authors successive sections describe about crop subdomain ontology, which is based on this *semantic Web architecture* and because of its inherent powerful feature for machine readability, vocabulary description and dynamic linking (Hendler, 2008) with other information resource dataset is amazing over extensive markup language (XML)-based datasets. This paper explores and organized as follows: Section 2, discusses about related work. This section explores some work done using various languages based WordNet projects. Section 3 and its successive subsections focus on Indian agriculture domain-related scenario and its prospect with semantic Web technology. Section 4 and its subsections describe technological aspect of information extraction for Indian agriculture domain. Section 5 concludes with future work directions and its significance in every human sphere of life.

2. Review of relevant literature work

The use of WordNet and domain-specific vocabulary-based information and data has been implemented since quite some time. In recent years, electronic forms of dictionary data have been used to generate ontologies of desired domain of interest, which categorically helps in many intelligence-based applications. Some projects related to natural language processing (NLP)-based applications has been developed in European, Chinese, Portuguese, Indian and Arabian languages, which significantly eases the work activities in diverse area of social, political, media, entertainment, and so on.

In this regard, the Indian IndoWordNet is a linked structure of wordnets of major Indian languages from Indo-Aryan, Dravidian and Sino-Tibetan families. These wordnets have been created by following the expansion approach from Hindi WordNet (Bhattacharyya, 2010), which are made free for research in 2006. So far as ontology is concern, it is rightly said that an ontology is, therefore, a shared understanding of some domain of interest (Uschold and Gruninger, 1996).

Ontology provides structural format to express metadata in different hierarchal break-ups into terminological branches of tree-like structure and is accessible with reduced search path in lesser time. Later on, ontology-based WordNet attempted to develop a new domain for medical WordNet (Smith and Fellbaum, 2004), which got success in defining and maintaining the clinical records of patients in medical domain. In this regard, the authors have also reviewed some of the developed system and their functioning too. The Agricultural Information Institute (AII) of Chinese Academy of Agricultural Sciences (CAAS), cooperated with AIMS to map CAT to the AGROVOC project (started in 2005) to generate the Chinese version of agro vocabulary and also EUROVOC, a cross-lingual, based on the label string exact match mapped (by CAAS, 2011). Some related works in this area based on the machine-learning approach also have been developed (Steinberger and Camila,

"Implementation of Semantic Web Technology in e-governance data is to enhance and upgrade the legacy system to make resource more web dynamic, reusable, sharable and easy available to its service users/stakeholders at anytime from anywhere."

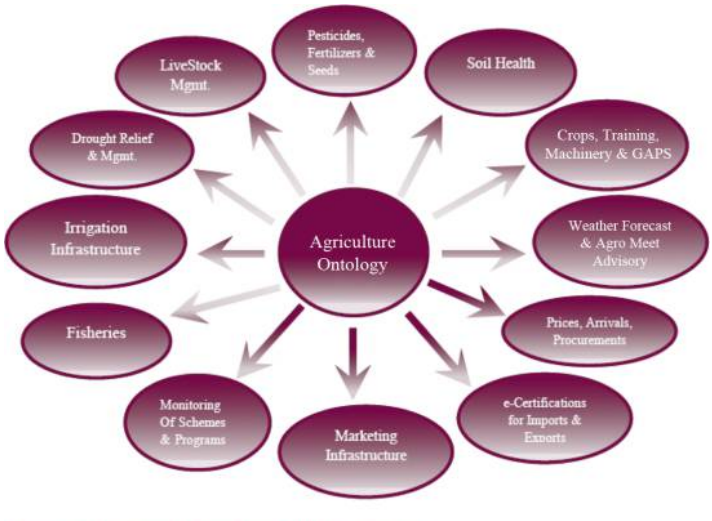
2003), which was based on category ranking classification. Later on for European Union official languages, an automatic *Eurovoc* thesaurus of parliamentary texts was developed (suggested by Steinberger, 2010). Some parallel works also have been done in this area, but there is no sufficiently automatic end-to-end solution available until now.

3. Indian agriculture domain at a glance and prospect for semantic Web technology

Indian agriculture has been broadly categorized into 12 core subdomains, which are shown as in Figure 1. Agriculture engineering and technological inputs have made significant contributions in an increasing production and productivity in agriculture sector through timely farm operations, accurate metering and better placement of seed and fertilizer, conserving soil and water resources. All these add irrigational potential and efficiencies, reducing loss of produce by providing improved storage structures and technologies with value addition. The marginal farmers, tribal farmers and landless agricultural labors are worst affected, as their coping capability is very limited. The biggest challenge is to make agriculture profitable, and this will be possible only by reducing the cost of cultivation (as indicated by the *Reports And Publications,* 2013 in National Food Security Mission) through mechanization and by higher returns to the farmers through value addition in production holdings and adopting loss prevention measures.

To facilitate services to these people, Ministry of Agriculture, Government of India, itself already has a number of Web sites and portals for different divisions. However, these Web sites do not share dynamic Web services among them and, hence, contents are static, non-integrated and challenges of extraction of data remains untouched. Most of the time citizens and other intended partners in the agricultural sector have to visit multiple Web sites to trace the desired piece of information, which is time-taking and cumbersome. The various existing un-matching technology standards and use of high resolution for graphical user interface (GUI's) and images, results in a lot of inconvenience to the user, which requires a lot of learning on their part to access the information and services. This resulted in mismatching of information, repeated efforts, outdated contents and multiple sources of information, giving rise to create confusions for service consumers.

Figure 1 Twelve core sectors of Indian agriculture domain which could be linked and share transparently through metadata among themselves



Source: Sinha and Chandra (2013a, 2013b)

3.1 Agriculture ontology design model

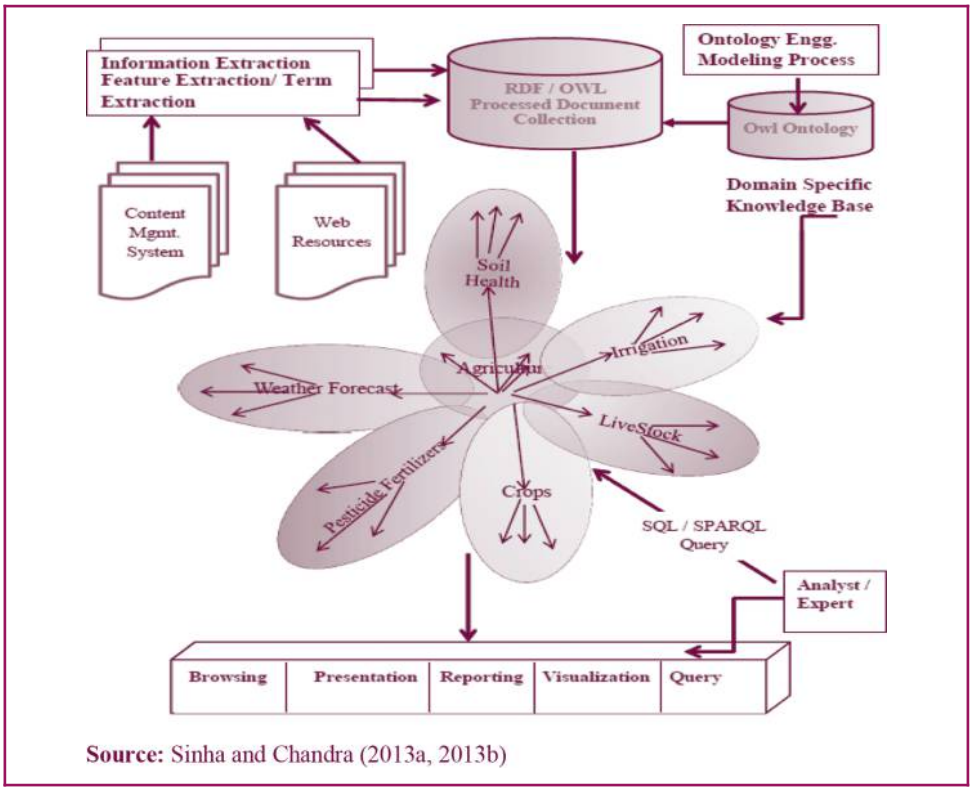
This model is an overall conceptual *agricultural ontology model*, and there are various associated components attached with this model (Sinha and Chandra, 2013a, 2013b). The authors have tried to generalize it by depicting its components. Other components are attached with its characteristics and functionality-based resource input/output and other concerning dependent objects. Content management system supports handling all features and terms available including local datasets, structured/unstructured and Web-related resources such as data available at URLs, which get processed to Resource Description Framework (RDF)/OWL format and stored into database and, hence, can be accessed by various components to provide services to farmers, stakeholders and end-users. Ontology engineering process helps to do proper analysis of its validity and strength of its demand and resources available, whereas reasoning/inferring deals with functionality of each component process guided through rules. Analyst/expert can avail the resources and do control and manage resources of interest.

In this model, in Figure 2, the authors have also tried to portray online Web documents extraction for the required piece of information, which may dynamically help in queries.

3.2 Indian crops ontology conceptualization – a model focused on crop

Conceptualization of real world's fact and information based on resource available can logically be designed and captured to give the shape of unique structural design. Conceptual crop ontology in Figure 4 is one of the major subdomain of *Indian agriculture domain*, whose significance is of inter-domain relationships and within its individual supporting domain relationship can be justified through the hierarchical concept, term, entity and relations. This is further classified using different properties they possess.

Figure 2 Model for overall Indian agriculture domain ontology. Each component has individual identity and role linked with each other



Source: Sinha and Chandra (2013a, 2013b)

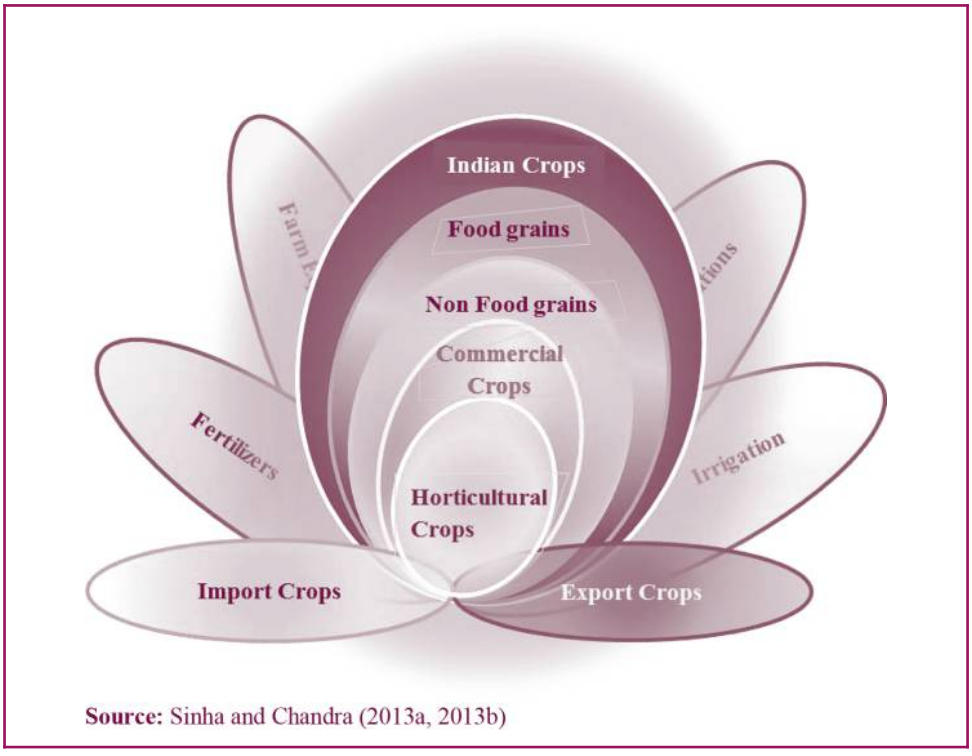
Authors extract these facts and information and classify them into various categories in the form of hierarchical inter-connection architecture.

Also, *automatic extraction* of these facts and information datasets reduces the repetitive task and can be reused as tool for similar tasks that enhances the resource utilization. These generated concepts, entities, relations and logically related facts and information can be tapped into real-world datasets in persistable form of RDF/OWL metadata format. These terms and concepts of authors c*rop ontology model* supports graphical model, commemorating these facts and allows various set operations onto it; such as *query*, *merging*, *addition*, *deletion*, etc. Here, in this novel model in Figure 3, the authors have tried to portray by mapping a conceptual concept of a crop model into a logical form, which helped in the creation of crop ontology, further supporting well-form format (w.f.f) structure of RDF/OWL of semantic Web technology. This machine-readable format supports query from any URL in the Web. Figure 4 shows each concept class that can have hierarchically one or more subclasses.

3.3 Crop subdomain and crop estimation methodology

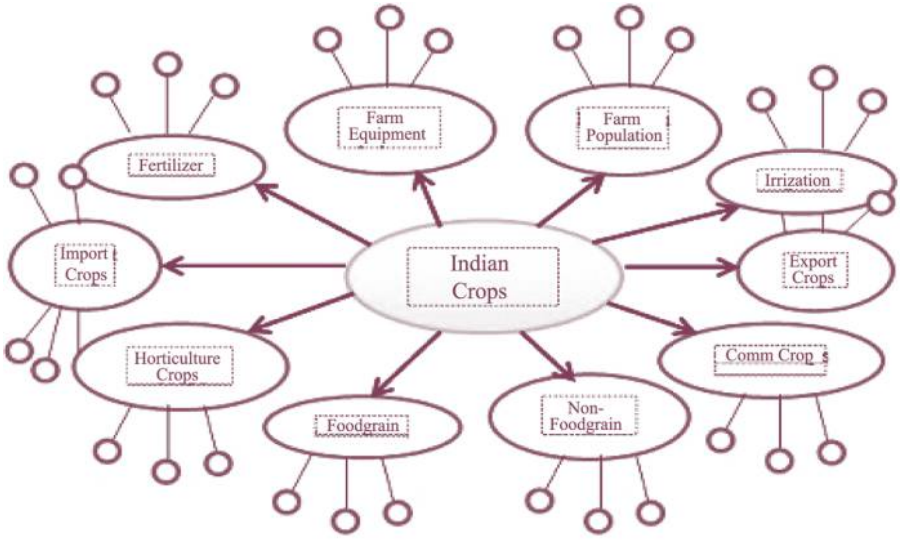
The National Policy for Farmers emphasizes the use of ICT at the village level for reaching out to the farmers with correct advisories and requisite information. ICT tools and space science applications are also being used for ensuring greater reliability of crop yield and production estimates, which will help in improving the process of planning and policy-making. Statistical data preparation for crop subdomain is done by the main governing body, Directorate of Economics and Statistics (DES) at the Center, which plays a major role in Ministry of Agriculture for agricultural statistics at all-India level. Some other principal agencies such as National Sample Survey Organization, Indian Agricultural Statistical Research Institute, State DES, etc. conduct methodical studies on agricultural statistics. Kharif and Rabi are two major seasonal crop period in India, which add to the *crop year cycle*. Principal crops of food grains, oilseeds, sugarcane, fibers and important

Figure 3 Concept to realization of Indian crop subdomain model reveals the fact that how taxonomically enriched real logical components can be transformed into ontology of domain interest



Source: Sinha and Chandra (2013a, 2013b)

Figure 4 Hierarchal model of Indian crops and its mutually linked subcomponent domain ontology

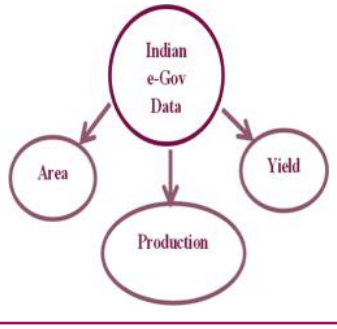


Source: Sinha and Chandra (2013a, 2013b)

commercial and horticulture crops are linked with the estimation of area production and yield, which DES releases. These crops are causes for the main agricultural output of 87 per cent of the total. *Multiplication of area estimates by corresponding yield estimates generates the estimated crop production* (as indicated by Ministry of Agriculture India, 2006). So, the estimated area and estimated yield play a vital role in agricultural statistics. The top level view can be seen in Figure 5.

1. *Area estimation statistics methodology*: So far, as collection of area estimation statistics is concern, all states in the country are divided into three broad categories:
   - *States and union territories (UTs)*: Area and land use statistics are assumed as part of Land Records, which are maintained by the revenue agencies and known as "Land Record States" or "Temporarily settled states". This system is followed by 17 major states and four UTs. These States/UTs cover 86 per cent of the reporting area and are covered in Timely Reporting Scheme, under which 20 per cent of the villages are selected at random for complete area of amount collection.
   - *Statistics are collected based on sample surveys*: Establishment of an Agency for Reporting of Agricultural Statistics had been established for reporting in

Figure 5 Area, production and yield estimation handle by DES



"The present research work is a proof-of-concept for partial automatic extraction and development of ontology from existing non-machine readable datasets of e-Government data to meet the standards of W3C specification."

these many states viz., Orissa and West Bengal, Kerala, and was later extended to Nagaland, Sikkim, Arunachal Pradesh and Tripura. It has a sufficiently large sample of 20 per cent villages or investigator zones, with an estimated reported area around 9 per cent.

- *Covers hilly districts*: It covers Assam and rest of the states in the north-eastern region (NER). But Goa, UTs of Andaman and Nicobar Islands, Daman and Diu and Lakshadweep have no reporting agency but entrusted with village headman. These areas/states responsible for 5 per cent of reporting area.
- *Area estimation top level*: Agricultural output of principal crops such as:
  - Eighty-seven per cent (food grains, oilseeds, sugarcane, fibers, important commercial and horticulture crops).

2. *Yield estimates*: This is obtained through analysis of Crop Cutting Experiments (CCE) conducted under General Crop Estimation Survey (GCES). At present, 95 per cent of the production of food grain is obtained through CCE, which is estimated on the basis of yield rates. A methodology comprised of experimental data, as well as mathematical calculation of sample design data. Sampling is done by various stratified multistage random sampling design at different above-mentioned divisions of crop estimation and sample design branch.
3. *Sampling design*: Layered multistage random sampling is done for carrying out GCES with Taluks/Revenue Inspector/Tehsils, etc. at first stage unit sampling, fields/survey numbers within each selected village as sampling unit at the second stage and experimental plot of a specified shape and size.

Sampling design for GCES
↓
Revenue village
↓
Survey number/field
↓
Experimental plots (specified size/shape)

So, the *Production Estimation* is calculated on the basis of the above facts and data collected and mathematical calculations give approximated statistical data. Procedures have been developed for acreage estimation and production forecast of major crops at various (national, state and district) levels, using *satellite remote sensing data, weather data and crop growth simulation models*. The technique is also useful in assessing the change in cropping pattern, the areas affected by flood (extent and duration) and change in crop vigor. This exercise has been undertaken under the Forecasting Agricultural Output using Space, Agro-meteorology and Land based Observations scheme (as indicated by Ministry of Agriculture India, 2006), which envisages multiple crop area and production forecasts of following *11 major crops* at national–state–district level depending on the status of

technology available. The unified modeling language (UML) design of the crop year cycle is expressed in the sequence diagram in Figure 6.

Continuing with crop subdomain and maintaining the machine-readable metadata approach for automatic extraction of facts and information in the form of concepts, sentence, terms and sense-based relation among these facts and/or data, here, in this collection of Hindi terms of Indian crop subdomain and matching Princeton WordNet based English word, correlates to concepts and terms in IndoWordNet of Hindi language with equivalent meaning (Table I).

4. An approach for implementation of semantic Web technology

Semantic Web technology is an advance form of current Web technology from its prior version, which implements the graph-based concept to express model of thoughts in

Figure 6 UML-based sequence diagram for crop cycle production estimation

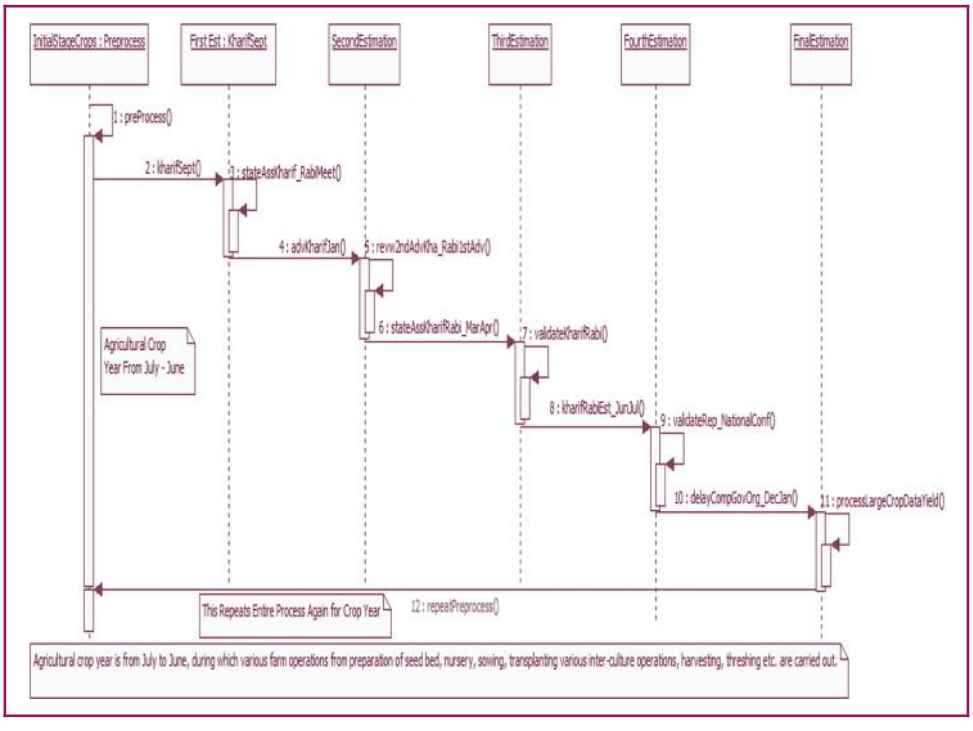


Table I Annotated resource data for Indian agriculture domain

Hindi Name	English Name
Dhan (धान) (for both season *Kharif* & *Rabi*)	Paddy (Rice)
Jowar (ज्वार) (for both season *Kharif* & *Rabi*)	Cholam (Great Millet)
Makka (मक्का)	Maize (Indian Corn)
Bajra (बाजरा)	Bara (Bulrush or Spiked Millet)
Jute (जूट)	Coarse Fibre
Mundua (मुंदुआ)	Ragi
Kapas (कपास)	Cotton
Ganna (गन्ना)	Sugarcane
Badam (बादाम) (*Kharif* & *Rabi*)	Groundnut
Sarso Beez (सरसों बीज) (Tael)	Rapeseed & Mustard Oil Seed
Gehun (गेहूँ)	Wheat

the form of computational logic. This trusted layered model can be represented in taxonomy of terms, concepts, relations of specific domain into *ontological* form of real world of interest. *Semantic Web technology* that conform W3C standard, supports several architectural constructs and frameworks of its own that are necessary for creation and generation of model envisioned for ontologies such as OWL, RDF (Berners-Lee, 2009), description logic and reasoning capabilities, etc. Metadata is simply machine-readable dynamic data that makes vocabulary for other resource data. Machine-readable (as indicated in Semantic Web, 2010 for *Setting Government Data Free*) facts and information-related metadata contributes to input some intelligence in our digitized world of applications.

As far as information extraction is concerned, the authors here are using NLP supported tools (GATE developer tools) to filter Web content of our interested domain resources so as to ease the creation of vocabulary of SynSets of regional languages. This also highlights the generation of uniform nature of datasets, which could be more useful in sharing, reuse of resource data and easy availability and bringing transparency within the system under consideration. Based on this trusted conceptual model, the authors have extracted terms, relations and concepts class information to generate ontology. Further, they generated statements of subject–predicate–object (S-P-O) by applying composition method, which categorically describes its terms, attributes and properties in the form of SynSets of specific category. Finally, the authors processed these facts and information and vocabularies using application framework (Protégé) to produce RDF/OWL format of Web-enabled dynamic machine-readable datasets (upon which various operations can be performed) for further inference and reasoning of resource metadata. Also, extracted terms, concepts and attributes of agriculture domain are later processed through model-based architecture framework (Jena) to enhance the feature of set operation for modeled data that adds the generation of voluminous metadata.

The model in Figure 7 shows matched terms from Wordnets (Princeton's English and IndoWordNet) of agriculture product term (concept) and its subclass-related entity attributes have been modeled to merge into single model of structured information. For which we have taken the English name and its attributed values of rice (*Oryza Sativa*, a botanical name of rice) merges to form a single model structure of a concerned type, saving time and space while reducing the complexity of search in Web content. Figure 7 depicts the merge process of terms and concept class to generate a uniform data format while maintaining the W3C standard of RDF/OWL metadata structure, which significantly helps in reducing the task of repetition work of conversion to unique format.

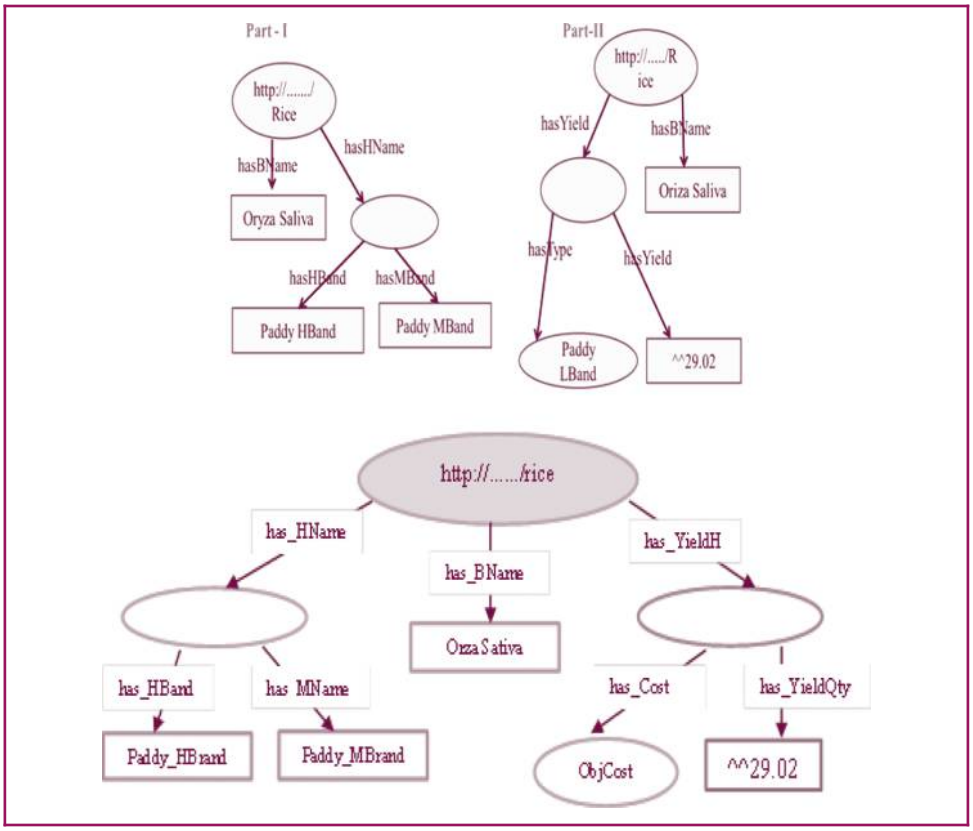
4.1 Terminology extraction

This subsection describes an approach for *extraction of terms* and *terminology*, which explores the process steps of extraction of terms, concepts and relations from input source (Web documents or raw files at URLs) corpora. Here, the authors have used reliable NLP-based tool (GATE Developer) in their experiment, which passes through such as tokenization, lexical parsing, parts of speech (POS) tagger filtration and machine learning-based NER to exploit the benefits of a supervised learning concept. Also, filtered and matched agricultural terms and other entities, facts and data are tapped to the local intermediate data files, maintaining the currently updated records for future use. During this progress, it also checks in Princeton English wordnet, IndoWordNet and agriculture-related terminology in the source vocabulary for multilingual match found for its equivalency. A sample of annotated filtered resource raw metadata of agriculture domain would be looked as below.

4.2 Semantic interpretation and linking to IndoWordNet

Once it finishes with a proper POS tagging of terms, and entity extraction, it enters for *semantic interpretation phase*. In this phase, tagged mark-ups and filtered terms are further

Figure 7 Merge model of concepts, terms into uniform structured model of metadata using Jena framework



looked for semantic process. Semantic morphological interpretation of concepts, terms, relations, etc. are matched with multilingual IndoWordNet consisting of SynSets schema structure which conforms to closer match with Princeton's English WordNet through Java API (Finlayson, 2014). Further, as it checks into the 16 relationships of IndoWordNet (Table III) as compared to Princeton's English WordNet to match and make sense of words, and relations between SynSets available in Hindi WordNet. Hindi WordNet relations are such as *hypernymy, meronymy, hyponymy, troponymy, antonymy*, etc. semantically related

Annotated resource raw metadata of agriculture domain (ref. Sinha and Chandra, 2013a, 2013b).

The term <*Rabi: CROPTYPE*> means "<*spring: SEASON*>" in <*Arabic: MISC*>, and the <*rabi: CROPTYPE*> <crops: *THING*> are grown between the months mid <*November: MONTH*> to <*April: MONTH*>. The <water: *THING*> that has percolated in the <ground: *NATURALTHING*> during the <rains: *NATURALTHING*> is main source of <water: *THING*> for these crops. <*Rabi: CROPTYPE*> crops require irrigation. So a good or bountiful <rain: *NATURALTHING*> may tend to spoil the <*Kharif: CROPTYPE*> <crops: *THING*> but it is good for <*Rabi: CROPTYPE*> <crops: *THING*>. These <crops: *THING*> are taken after the departure of <*monsoon*: *SEASON*> <rains: *NATURALTHING*>. The <seeds: *THING*> are <sown: *VERB*> after the <rains: *NATURALTHING*> have gone and <harvesting: *VERB*> begins in <*April: MONTH* > /<*May: MONTH* >. Major <*Rabi: CROPTYPE*> <crop: *THING*> is <*wheat: CROPS*> in <*India: COUNTRY*> followed by <*barley: CROPS*>, <*mustard: CROPS*>, <*sesame: CROPS*> and <*peas: CROPS*>. (They are <harvested: *VERB*> early as they are ready early). So <*Indian*: *COUNTRY*> <*markets: THING*> are <flooded *NATURALTHING*>with < *Green Peas: CROPS*> from <*January: MONTH*> to <*March: MONTH*> (Peak is <*Feb: MONTH*>.)

between SynSets, and lexical relations are between words and these serve as organized lexical knowledgebase. Figure 8 shows hierarchy of English WordNet.

In this experiment, the authors took some extracted Hindi terms matching with Princeton's English WordNet, which further matches with its corresponding concept and terms of agricultural Hindi words in IndoWordNet to locate its matching SynSets, relations, ontological relations of these terms and concepts, offsets, sense count, etc. for each SynSets in the IndoWordNet. An algorithm is created to extract information from IndoWordNet (Figure 12). This resulted for helpful estimation to reduce the task of manual translation and matching of words in the electronic data dictionary. Using this extracted information, authors approached closer to generate ontology of local language and, hence, into RDF/OWL knowledgebase, which is reusable with persist option and uploadable to a quad server in the Web. Figure 9 below depicts a conceptual logical model for input–process–output of experimental findings in which authors categorically used an input source, operations and the desired output as to demonstrate the overall facts as proof of concept.

Table II shows the sample matched words (in Princeton WordNet and IndoWordNet) taken as input source. Table III shows Hindi WordNet relations of different SynSets and Words.

4.3 Semantic interconnection of concepts and terms

Semantic interconnection of concepts and terms relates to SynSets, WordSense and their relationships among themselves. The authors have categorically linked multilingual IndoWordNet to development framework (JAVA) while supplying source corpora of words

Figure 8 Structure of Princeton's English WordNet

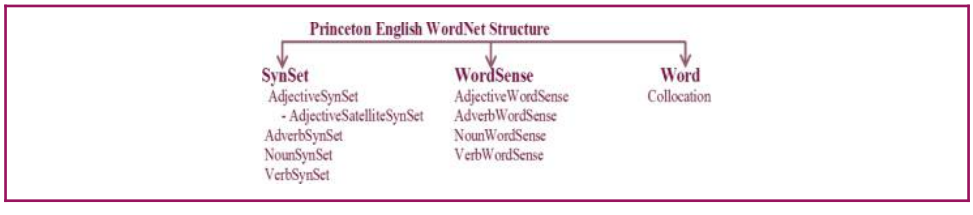


Figure 9 IndoWordNet Linking and the process of ontology generation. This model depicts input source match with IndoWordNet then does pre-process to generate ontology

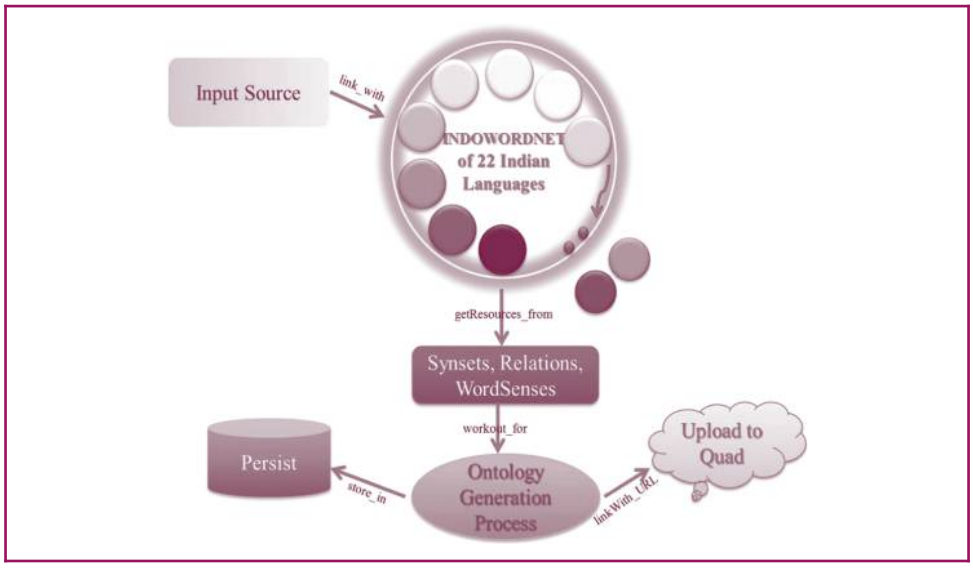


Table II Agriculture Terms used for sample data matching with English WordNet as well as to IndoWordNet

अनाज	Grain
चना	Gram
भोजन	Food
फल	Fruit
आम	Mango
धान	Rice
गेहूँ	Wheat
खाद	Edible
कपास	Cotton
पेड़	Tree

Table III Semantic and lexical relationships between SynSets and words of IndoWordNet

Hindi WordNet Relations

Semantic Relations (exits between synsets)	*Lexical Relations* (exits between words)
Hyponymy and Hypernymy (kind of relation) **Meronymy and Holonymy** (Part-whole relation) **Ability Link** (specifies inherited features of a nominal concept) **Capability Link** (specifies acquired features of a nominal concept) **Function Link** (specifies function of a nominal concept) **Causative** (link causative verb and base verb) **Attribute** (relation between noun and adjective)	**Antonymy** (represent opposite words) **Gradation** (represent intermediate concept between two opposite concepts) **Entailment** (relation between two verbs) **Troponymy** (specifies elaboration of another verb) **Derived From** (specifies the root from which word is derived) **Modifies Noun** (Adjectives that modify nouns) **Modifies Verb** (Adverbs that modify verbs)

of matching lexical terms, concepts to automatize the extraction of corresponding terms, concepts and relationships of agricultural domain. Further, composition rule was applied over output-extracted words, terms, etc. to get unique words of SynSets. This can be explained through in successive examples.

Let $t = w_n \ldots\ldots w_3, w_2, w_1$ having multiword term belonging to lexically created tree T (Velardi *et al.*, 2007). The Word Sense of t can compositionally expressed:

$WS(t) = [WS_i \mid WS_i \in SynSets(_{Wi}), {}_{Wi} \in t]$ where $SynSet(w_i)$ is a set of WordSense from WordNet for word w_i such as:

$$WS(\text{“खाद्य_पदार्थ”}) = [\{\text{अनाज, अन्न, गल्ला, खाद्यान्न, धान्य}\},\{\text{ज्वार, जवार, जुआर, जुवार,रक्तजूर्ण }\}]$$

Below, the generated sentence is composed of SynSets, and there is a relationship of *kind_of*, which denotes that super-set and subset SynSets are linked. A case of *Hypernymy* and *Hyponymy* relationship is established. Also, sentence is assumed to be linked using the S-P-O relationship of RDF/OWL standard, which adds to the vocabulary from extracted terms of interest of a SynSet, which inherently forms a piece of hierarchical ontology. In this way, the authors have generated an ontology of concepts, terms, relations of agriculture domain (could be any domain of interest) while adding attributes and properties to create vocabularies for sentence of agriculture domain. Figure 10 depicts the concept of creation of sentences.

In this effort to achieve automatic extraction of facts and information about a concept and creation of knowledgebase of uniform structure using semantic Web technology, the

authors have linked Princeton English WordNet and IndoWordNet to automatized extraction of concepts, terms, relations, etc. to map into a uniform structure of well-form formatted unique metadata datasets in the form of RDF/OWL of W3C standard. Although it is tricky and substantial effort is required to establish hierarchical relationships among Hindi verb-based terms, phrases and their semantic relations with SynSets and words because of its complexities and cross-lingual local semantic meaning that *differ* among SynSets and between words in IndoWordNet, which is still in the phase of change for its full and final version.

In the next successive findings and results that depict authors effort and approach towards the mapping of conceptual hypothesis to a real-world ontology generation. Using Java

Figure 10 Semantic graph that forms Sentence as of concept extraction

Figure 11 An extracted sample output of Rice (Dhaan-धान) class concept shows the results of extraction (using Java program) of terms, SynSets, Sensecount, Hierarchical Relations (parent–child) of nodes etc. from IndoWordNet

धान

Sense Count is 2
Offsets[0] 6302
Offsets[1] 6303
Synset [0] 6302 - NOUN - [धान, धान्य, शालि, धान्यक, हैमन]
Synset POS: NOUN
Synset Num Pointers:11
HYPERNYM : null
HYPONYM : 7443 - NOUN - [जड़हन, शालि, जड़हन_धान]
HYPONYM : 7882 - NOUN - [साठी, साठी_धान, गर्भपाकी, वृही, मुकंद, मुकन्द, मुकंदक, मुकन्दक, शतपुष्प, महाव्रीही, हैमना]
HYPONYM : 14468 - NOUN - [तिन्नी, तिन्ना, पसही, पसाई, ननोई, तीनी, पसारी]
HYPONYM : 14857 - NOUN - [अगहनी, अगहनिया]
HYPONYM : 16988 - NOUN - [बासमती]
HYPONYM : 23891 - NOUN - [बेलन]
HYPONYM : 32286 - NOUN - [असरा]
HYPONYM : 37507 - NOUN - [देव-धान्य, देवधान्य]
MERO_COMPONENT_OBJECT : 6303 - NOUN - [धान, धान्य, धान्यक, शालि, धान्योत्तम, हैमन]
ONTO_NODES : वनस्पति (Flora) (FLORA उदाहरण:- शैवाल,लता,वृक्ष इत्यादि); सजीव (Animate)
(ANIMT उदाहरण:- मानव,जानवर,वृक्ष इत्यादि); संज्ञा (Noun) (N उदाहरण :- गाय,दूध,मिठाई इत्यादि); TOP (Top Level Node)
Synset [1] 6303 - NOUN - [धान, धान्य, धान्यक, शालि, धान्योत्तम, हैमन]
Synset POS: NOUN
Synset Num Pointers:14
HOLO_COMPONENT_OBJECT : 6302 - NOUN - [धान, धान्य, शालि, धान्यक, हैमन]
HYPERNYM : 19 - NOUN - [अनाज, अन्न, गल्ला, खाद्यान्न, धान्य, शस्य, वाज, इड़, इरा]
HYPONYM : 7347 - NOUN - [बोरो]
HYPONYM : 14469 - NOUN - [तिन्नी, तिन्ना, पसही, पसाई, ननोई, तीनी]
HYPONYM : 15969 - NOUN - [अनंदी, अनन्दी]
HYPONYM : 16319 - NOUN - [कनकजीरा]
HYPONYM : 16989 - NOUN - [बासमती]
HYPONYM : 21260 - NOUN - [बादामी, बदामी]
HYPONYM : 29210 - NOUN - [अमाघौत]
HYPONYM : 33026 - NOUN - [तिलवासिनी, राजप्रिया]
MERO_COMPONENT_OBJECT : 5207 - NOUN - [चावल, चाँवल, तंडुल, तंदुल, धान्यसार]
ONTO_NODES : प्राकृतिक वस्तु (Natural Object) (NAT-OBJCT उदाहरण:- पर्वत,लकड़ी,जल इत्यादि); वस्तु (Object)
(OBJCT उदाहरण:- पुस्तक,छाता,पत्थर इत्यादि); निर्जीव (Inanimate) (INANI उदाहरण:- पुस्तक,घर,धूप इत्यादि);
संज्ञा (Noun) (N उदाहरण :- गाय,दूध,मिठाई इत्यादि); TOP (Top Level Node)
ONTO_NODES : खाद्य (Edible) (EDBL उदाहरण:- आम,मिठाई,दही इत्यादि); वस्तु (Object) (OBJCT उदाहरण:- पुस्तक,छाता,पत्थर इत्यादि);
निर्जीव (Inanimate) (INANI उदाहरण:- पुस्तक,घर,धूप इत्यादि); संज्ञा (Noun) (N उदाहरण :- गाय,दूध,मिठाई इत्यादि); TOP (Top Level Node)
ONTO_NODES : भाग (Part of) (POF उदाहरण :- पंख, टहनी, पैंच इत्यादि); संज्ञा (Noun) (N उदाहरण :- गाय,दूध,मिठाई इत्यादि); TOP (Top Level Node)

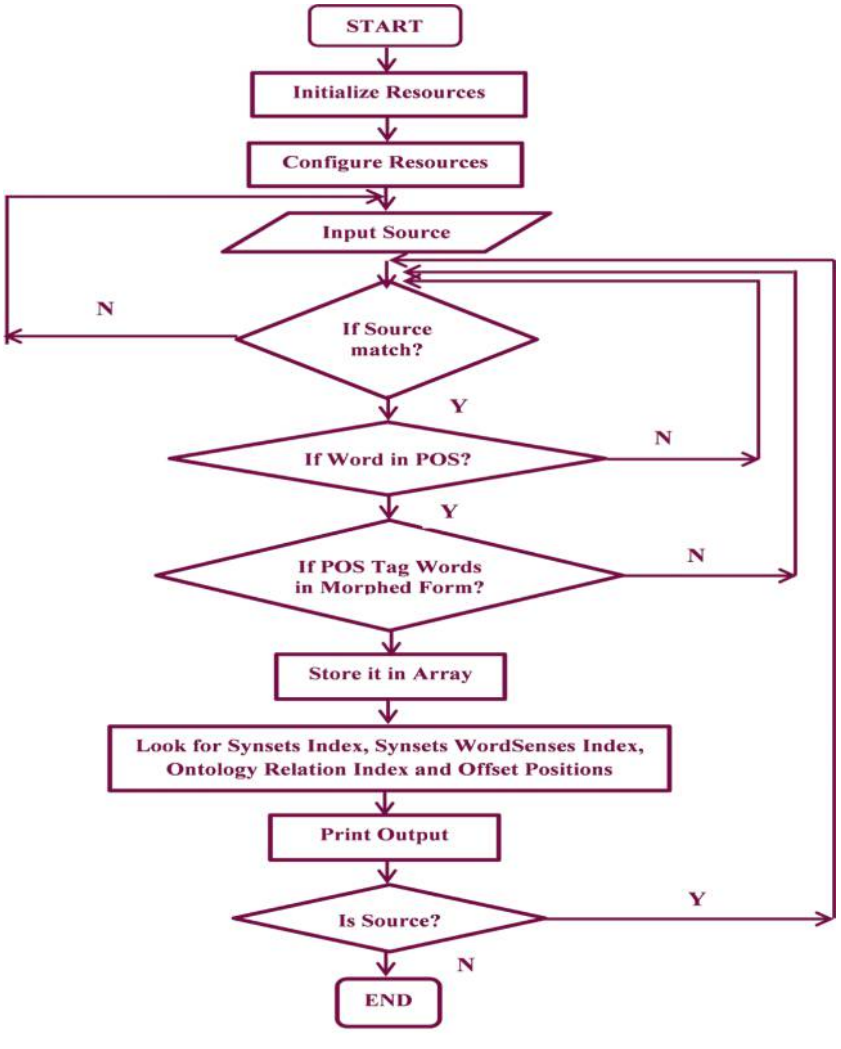
Source: Garg *et al.* (2014)

development environment and frameworks-supporting Java feature, authors have created an algorithm to extract SynSets, Sensecount, lexical relations and others by linking IndoWordNet to it. This conceptual semantic linking helped in extracting significant information, which was further processed to generate ontology and metadata of a uniform structure.

Figure 11 shows the output of sample term Rice (*Dhaan*-धान) (Figure 12; Table IV).

Finally, ontology of Hindi terms is generated from the extracted concepts, terms and their natural semantic and lexical relations in the form of semantically well-form hierarchical architecture of loosely coupled and highly related metadata. This metadata dataset has been created using the Protégé development framework, which is persistable and uploadable to a quad Web server so as to make knowledgebase more dynamic, sharable, reusable and easy accessible through query interface and set operations (Sinha and Chandra, 2013a, 2013b) at anytime from anywhere in optimized and transparent manner. Sample ontology for rice (Dhaan -धान) has been generated, as shown in Figure 13.

Figure 12 A flow-chart diagram for extraction algorithm for concept, term, relations, sense count, etc. of linked IndoWordNet



Source: Garg *et al.* (2014)

5. Conclusion and summary of findings, limitations implications with future work directions

In conclusion, authors' present research work is a proof-of concept for partial automatic extraction and development of ontology from existing non-machine readable datasets of e-Government data to meet the standards of W3C specification. Authors have linked multilingual IndoWordNet support for Indian languages as a source to compare with natural language and extract information and data to generate ontology of agriculture domain in the uniform structured data (RDF/OWL) format. The unicode supported data helps to dynamically link open data in the Web, thereby making accessing information easier and sharable among various related domains and, hence, adds to the reusability of resources.

Table IV Sample Output results of Hindi language extracted trans relations and various lexical & semantic relations of a few Agricultural terms and its synsets & their sensecount from IndoWordNet

Hindi Words	अनाज	भोजन	धान	गेहूँ	खाद	कपास
English Match	Grain	Food	Paddy	Wheat	Fertilizer	Cotton
Sense Count	1	4	2	2	1	2
Synset POS	NOUN	NOUN	NOUN	NOUN	NOUN	NOUN
Synset[i]- e.g. Synset[0], Synset[1], Synset[2] etc. →	[अनाज, अन्न, गल्ला, खाद्यान्न, धान्य, शस्य, वाज, इड, इरा]	**Synset[0]** -[भोजन, खाना, आहार, अन्न, आहर, रोटी, डाइट, ज्योनार, जेवन]; **Synset [1]**- [भोजन, भोजन_कर्म, अन्न_ग्रहण, अशन]; **Synset [2]**- NOUN - [शिकार, भक्ष्य_पदार्थ, आहार, भोजन]; **Synset [3]** - [भोजन, खाना, आहार, डाइट]	**Synset [0]**- 6302 - NOUN - [धान, धान्य, शालि, धान्यक, हैमन]; **Synset [1]**- 6303 - NOUN - [धान, धान्य, धान्यक, शालि, धान्योतम, हैमन]	**Synset [0]**:- 4045 - NOUN - [गेहूँ, कनक, गोधूम, गंदुम, गन्दुम, सुमन, शुक्रद, बहुदुग्ध, गेहूं]; **Synset [1]**:- 11734 - NOUN - [गेहूँ, गोधूम, गंदुम, गन्दुम, सुमन, गेहूं]	**Synset [0]:** - 4648 - NOUN - [खाद, फर्टिलाइजर]	**Synset [0]**:- 2178 - NOUN - [कपास, अपूरणी, वादरा, तुंडकेरिका, तुण्डकेरिका, तुंडकेरी, तुण्डकेरी, पाटद, बदर, स्थूला]; **Synset [1]**:- 12345 - NOUN - [कपास, बाँगा]
HYPERNYM → ↓	NOUN - [खाद्य_वस्तु, खाद्य_पदार्थ, खाद्यवस्तु, आहार, खाद्य, भोज्य_पदार्थ, खाद्य_सामग्री, खाद्य-सामग्री, आहार_पदार्थ, अर्क, खाना, इरा, फूड, रसद]	**Synset[0]**- [खाद्य_वस्तु, खाद्य_पदार्थ, खाद्यवस्तु, खाद्यपदार्थ, आहार, खाद्य, भोज्य_पदार्थ, खाद्य_सामग्री, खाद्य-सामग्री, आहार_पदार्थ, अन्न, अर्क, आहर, खाना, इरा, फूड, रसद]	**Synset[0]**- 19 - NOUN - [अनाज, अन्न, गल्ला, खाद्यान्न, धान्य, शस्य, वाज, इड, इरा]; **Synset[1]** - NULL ;		**Synset[0]:** 744 - NOUN - [पदार्थ, वस्तु, चीज, चीज़, द्रव्य]	**Synset[0]:** 4578 - NOUN - [झाड़, झाड़ी, क्षुप, गुच्छ, गुच्छक]; Synset [1]: 12345 - NOUN - [कपास, बाँगा]
		Synset[1] - [काम, कार्य, कर्म, करम, करनी, कृत्य, कृति, आमाल]				
HOLO_MEMBER_COLLECTION; **HOLO_COMPONE-NT_OBJECT;** **HOLO_STUFF_OBJECT** → ↓	7251 - NOUN - [बाल, बाली, रास, राशि, धनधान्य, धन-धान्य, अन्नकूट, अन्न-कूट]			**Synset[0]:** 14930 - NOUN - [गोजई, गोजरा]		HOLO_COMPONENT_OBJECT : **Synset[1]:** 2178 – NOUN [कपास, वादरा, अपूरणी, तुंडकेरी तुंडकेरिका,तुण्डकेरिका]

(*continued*)

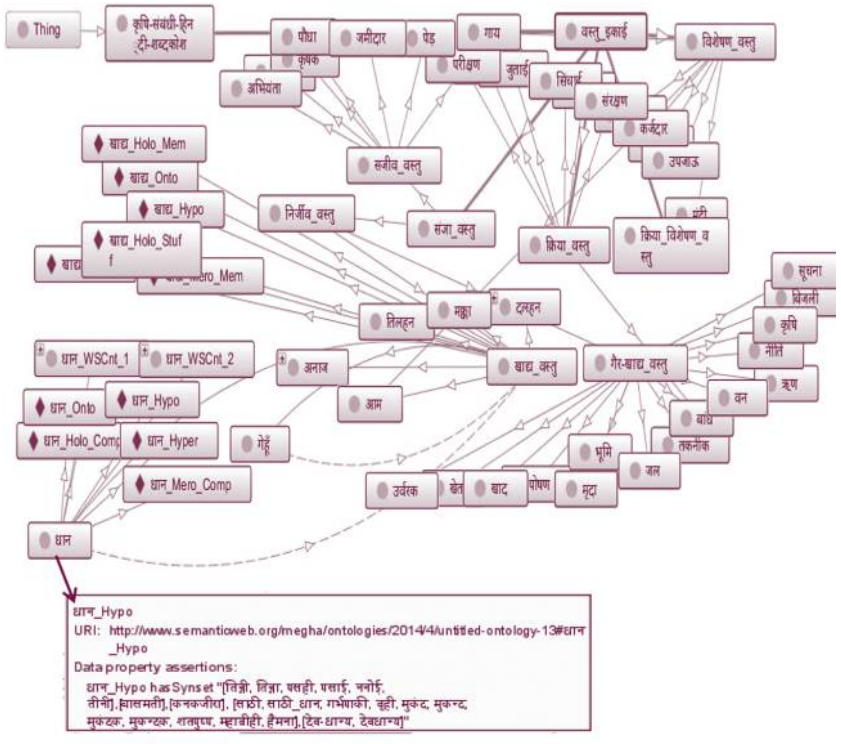
Table IV

HYPONYM : → ↓	1967-NOUN - [दलहन]	**Synset[0]:** - 441- NOUN - [डोसा, दोसा]	Synset[0]: 7443 - NOUN - [जड़हन, शालि, जड़हन_धान]	**Synset[0] :** 16443 - NOUN - [पम्मन, कठिआ_गेहूँ]		**Synset[1]** : 15997 - NOUN - [रढ़िया]
	1394-NOUN - [मक्का, मकई, मक्की, इक्षुपात्रा]	**Synset[0]:** -5006- NOUN - [कौर, निवाला, ग्रास, कवल, गस्सा]	Synset[0]: 7882 - NOUN - [साठी, साठी_धान, गर्भपाकी, वृही, मुकंद, मुकन्द, मुकंदक, मुकन्दक, शतपुष्प, महाव्रीही, हैमना]	**Synset[0] :** 17801 - NOUN - [दुरुम]	58 - NOUN - [कंपोस्त, कंपोस्त_खाद, कंपोस्ट, कंपोस्ट_खाद, कम्पोस्त, कम्पोस्त_खाद, कम्पोस्ट, कम्पोस्ट_खाद]	**Synset[1]** : 14618 - NOUN - [नरमा, देवकपास, देव-कपास, रामकपास, मनवा]
	12066-NOUN - [ज्वार, जवार, जुआर, जुवार, इक्षुपात्रा, रक्तजूर्ण, जुन्हरी, जुंडी]		Synset[0]: 37507 - NOUN - [देव-धान्य, देवधान्य]			
	33944-NOUN - [ऊँज्जारी, ऊँजरी, ऊजारी, अगऊँ]	**Synset[1]:**- 8199 - NOUN - [प्रीतिभोज, दावत, पार्टी, ज्योनार]				
MERO_MEMBER_COLLECTION: , **MERO_COMPONENT_OBJECT :** →			**Synset[0]:** 6303 - NOUN - [धान, धान्य, धान्यक, शालि, हैमन]; **Synset[1]:** MERO_COMPONENT_OBJECT : 5207 - NOUN - [चावल, तंदुल, धान्यसार]	**Synset[1]:** MERO_COMPONENT_OBJECT : 4045 - NOUN - [गेहूँ, कनक, गोधूम, गंदुम, गन्दुम, सुमन, शुक्रद, बहुदुग्ध, गेहूं]		
ONTO_NODES : →	खाद्य(Edible),प्राकृतिक वस्तु(Natural Object), निर्जीव(Inanimate), भाग(Part of)	मानवकृति (Artifact), शारीरिक कार्य(Physical), कार्य(Action), वस्तु(Object), जन्तु(Fauna), सजीव(Animate),निर्जीव(Inanimate), संज्ञा (Noun), TOP (Top Level Node)	**Synset[0]:** वनस्पति (Flora) ; सजीव (Animate) ; संज्ञा (Noun) ; TOP (Top Level Node); Synset[1]: प्राकृतिक वस्तु (Natural Object); वस्तु (Object); निर्जीव (Inanimate) ;खाद्य (Edible) ; भाग (Part of)	झाड़ी (Shrub) ; खाद्य (Edible); वनस्पति (Flora); सजीव (Animate) ; संज्ञा(Noun) ; प्राकृतिक वस्तु (Natural Object); वस्तु निर्जीव (Inanimate) ; TOP (Top Level Node)	वस्तु (Object) ; निर्जीव (Inanimate) ; संज्ञा (Noun) ; TOP (Top Level Node)	**Synset[0]**: वनस्पति (Flora) ; सजीव (Animate) ; संज्ञा (Noun) ; Synset[1]: प्राकृतिक वस्तु (Natural Object); वस्तु (Object) ; निर्जीव (Inanimate) ; भाग (Part of) TOP (Top Level Node)

In this succession of findings, a lot of exploration of mapping and matching of natural language suitability and acceptability in real-life use of agricultural terms, concepts and their relations within the domain have been performed to insure the validity for automatic extraction of agricultural terms, concepts (SynSets) and relations (lexical and semantic) from IndoWordNet. No doubt, it is a tricky job to extract information and generate ontology from it and to thereby create a vocabulary of metadata using various modeling approaches, and technological support of tools and development interfaces.

Diversification, authenticity and natural acceptability of languages in the society and their practical implementation are some of the limitations in generation of hierarchal ontology of domain of interest, which deviates from the desired result and leads to the wrong conclusion of concept under investigation. Since India is rich in this regard and lot of applicability and implementation of applications are based on these features, the eventual

Figure 13 Graph view of generated ontology of agricultural terms, entity and their relationships with the respective level of ontological hierarchy



Note: Showing trans-relation of object entity, lexical and semantic relations (hypernymy, hyponym, meronymy, holonymy, etc.) along with their object property. Also depicting SynSets and data property of rice

Source: Garg *et al.* (2014)

litmus test for such diversification is inevitable. Apart from that, very limited authenticated source of information related to individual domain is available in recorded electronic form to make use of it. So, IndoWordNet (which links all Indian regional languages) is one of the interfaces and resource that can be used for bringing it together by linking diversified Indian regional language-based society for its reform and development. However, it is in process of revision and the next version is yet to come with better usability.

Once a common interface for interpretation of diversified language knowledgebase is realized, a lot of implications and applicability of domain of interest can be put into practice which governs the social value systems, as well as good, reliable, and transparent e-governance systems. For example, as shown in Figure 1,, all peer domain and subdomain of agriculture is expected to be linked with uniform object metadata, thereby applicability, usability, transparency and well-to-do systems availability is automatically assumed without any barrier of language or domain interface constraint. In this process of development and pre-processing outcome of intermediate results of ontology generation using IndoWordNet, the authors have focused to work out more with this knowledgebase to enrich IndoWordNet, as well to bring these into reality and practicable because without this knowledgebase, options remain limited and the applicability of structured metadata is reduced.

Hence, the scope for development and implication of applications can be envisioned as a future scope of work directions. A significant work can be possible to link cross-multilingual local language support to IndoWordNet and thereby to create agro-vocabulary of Indian

languages (CAT: AII, CAAS, 2011) to link with FAO's Linked Open Data, which supports cross multilingual metadata for agriculture domain globally. A lot of applicability and a wide range of such applications support and reuse of such metadata for future scope of work justify the construction of uniform structured knowledgebase specially related to agriculture domain as well as others, for which the *semantic Web technology* needs to succeed globally.

References

Bahuguna, A. (2012), *Agriculture Census 2010-11*, Department of Agriculture & Cooperation Ministry of Agriculture Government of India, Krishi Bhawan, New Delhi.

Berners-Lee, T. (2007), "Linked data", available at: www.w3.org/DesignIssues/LinkedData.html/ (accessed 12 August 2013).

Berners-Lee, T. (2009), "Putting government data online", available at: www.w3.org/DesignIssues/GovData.htm/ (accessed 18 September 2013).

Bhall, G.S. and Singh, G. (2010), "Final report on planning commission project growth of Indian agriculture: a district level study", Centre for the Study of Regional Development Jawaharlal Nehru University, New Delhi.

Bhattacharyya, P., Fellbaum, C. and Vossen, P. (Eds) (2010), "Principles, Construction and Applications of Multilingual Wordnets, 5th Global Wordnet Conference Proceedings", Narosa Publishing House, Mumbai.

CAT: AII, CAAS (2011), "Linked open data of Chinese agricultural thesaurus", CAT: AII, CAAS, China, A web link project, available at: www.aims.fao.org/community/agrovoc/blogs/chinese-agricultural-thesaururs-published-linked-open-data# (accessed 14 March 2014).

Finlayson, A.M. (2014), "Java libraries for accessing the princeton wordnet: comparison and evaluation", *GWC 2014, Proceedings of the Seventh Global WordNet Conference*, *Tartu*, pp. 78-85.

Garg, M., Sinha, B. and Chandra, S. (2014), "Semi-automatic RDFization of Hindi agricultural words using indowordnet", *IEEE International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, *Greater Noida*, pp. 7269-7274.

Hendler, J. (2008), "Web 3.0: Chicken farms on the semantic web", *Computer, IEEE Computer Society*, Vol. 41 No. 1, pp. 106-108.

Keita, N., Srivastava, M., Ouedraogo, E. and Kabore, M. (2010), "Collecting agricultural data from population census: overview of FAO recommendations and experiences of Burkina Faso and other countries", *Fifth International Conference for Agricultural Statistics*, *Kampala*, pp. 2-5.

Kumar, R. (2010), *Ailing Agricultural Productivity in Economically Fragile Region of India: An Analysis of Synergy between Public Investment and Farmers' Capacity*, A research report published in Indian Council of Agricultural Research, New Delhi, available at: http://iiss.nic.in/publication/AP%20Cess%20Report.pdf

Ministry of Agriculture, India, (2006), "National commission on farmers: serving farmers and saving farming", available at: http://agricoop.nic.in/Agristatistics.htm/ (accessed 20 February 2014).

Reports and publications (2013), "National food security mission", available at: www.agricoop.nic.in/ (accessed 21 May 2013).

Semantic Web (2010), "Setting government data free", available at: www.semanticweb.com/ (accessed 18 September 2013).

Sinha, B. and Chandra, S. (2013a), "Semantic web query on e-governance data for crop ontology model of Indian agriculture domain", in Dutta, B. and Madalli, D.P. (Eds), *International Conference on Knowledge Modelling and Knowledge Management (ICKM)*, Bangalore (Bengaluru), pp. 56-66.

Sinha, B. and Chandra, S. (2013b), "Semantic web ontology model for Indian agriculture domain", in Dutta, B. and Madalli, D.P. (Eds), *International Conference on Knowledge Modelling and Knowledge Management (ICKM)*, Bangalore (Bengaluru), pp. 101-111.

Smith, B. and Fellbaum, C. (2004), "Medical wordNet: a new methodology for the construction and validation of information resources for consumer health", *Proceedings of Coling: The 20th International Conference on Computational Linguistics, Geneva*, 23/27 August, pp. 31-38.

Steinberger, R. (2010), "Automatic Eurovoc indexing of parliamentary texts in all official EU languages", European Commission, Luxembourg.

Steinberger, R. and Camila, I., *et al.* (2003), "Automatic annotation of multilingual text collections with a conceptual thesaurus", in *Proceedings of the Workshop Ontologies and Information Extraction at the Summer School The Semantic Web and Language Technology – Its Potential and Practicalities*, (EUROLAN'2003). Bucharest, Romania, 28 July – 8 August 2003.

Uschold, M. and Gruninger, M. (1996), "Ontologies: principles, methods and applications", *The Knowledge Engineering Review*, Vol. 11 No. 2, pp. 93-136.

Velardi, P., Cucchiarelli, A. and Petit, M. (2007), "A taxonomy learning method and its application to characterize a scientific web community", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 19 No. 2, pp. 180-191.

W3C INDIA, (2014), "National e-governance plan- better government through better use of the web", available at: www.w3cindia.in/ (accessed 23 March 2014).

About the authors

Bhaskar Sinha has completed his MCA, MTech (Gautam Buddha University, Greater Noida, India) with specialization in Intelligent System and Robotics. Completed his MTech thesis work in "Semantic Web Technology" and presently attached with research work on NICSI and TDIL based project using semantic Web technology. Earlier, he had worked within the IT industry and is presently associated with research work on semantic Web at W3C India, New Delhi, as Research Consultant. Bhaskar Sinha is the corresponding author and can be contacted at: bhaskar_sindel@hotmail.com

Dr Somnath Chandra a Scientist and is currently working as Dy. Country Manger W3C India, CGO complex, New Delhi, India. He has done his BTech, MTech from IIT Kharagpur, India, and PhD from IIT Delhi, India. He has extensively worked on Optical Fiber and WebTechnology. Key role in establishing W3C India office, TDIL India and leading the activity of Internationalization requirements in W3C standards.

Megha Garg has completed her MSc, MTech (Gautam Buddha University, Greater Noida, India) with specialization in Intelligent System and Robotics and is preparing for a PhD program. She did her MTech thesis in "Nature-Inspired Metaheuristics". She is currently associated with research work on semantic Web at W3C India, New Delhi, as a Project Associate.