OntoInfoG++: A Knowledge Fusion Semantic Approach for Infographics Recommendation

Gerard Deepak*, Adithya Vibakar, A. Santhanavijayan

Department of CSE, National Institute of Technology, Tiruchirappalli (India)

Received 8 April 2021 | Accepted 6 October 2021 | Published 22 December 2021

Abstract

As humans tend to improvise and learn on a constant basis, the need for visualizing and recommending knowledge is increasing. Since the World Wide Web is exploded with a lot of multimedia content and with a growing amount of research papers on the Web, there is a potential need for inferential multimedia like the infographics which can lead to an ultimate new level of learning from most viable information sources on the Web. The potential growth and future of technology have called for the need of a Web 3.0 compliant infographic recommendation system in order to be able to visualize, design and develop aesthetically. The trend of the Web has asked for better infographic recommendations in the attempt of technological exploration. This paper proposes the OntoInfoG++ which is a knowledge centric recommendation approach for Infographics that encompasses the amalgamation of metadata derived from multiple heterogenous sources and the crowd sourced ontologies to recommend infographics based on the topic of interest of the user. The user- clicks are taken into consideration along with an Ontology which is modeled using the titles and the keywords extracted from the dataset comprising of research papers. The approach models user topic of interest from the Query Words, Current User-Clicks, and from standard Knowledge Stores like the BibSonomy, DBpedia, Wikidata, LOD Cloud, and crowd sourced Ontologies. The semantic alignment is achieved using three distinct measures namely the Horn's index, EnAPMI measure and information entropy. The resultant infographic recommendation has been achieved by computing the semantic similarity between enriched topics of interest and infographic labels and arrange the recommended infographics in the increasing order of their semantic similarity to yield a chronological order for the meaningful arrangement of infographics. The OntoInfoG++ has achieved an overall F-measure of 97.27 % which is the best-in-class F-measure for an infographic recommendation system.

Keywords

Horn's Overlap Index, Infographics Recommendation, Knowledge Centric, Ontologies.

DOI: 10.9781/ijimai.2021.12.005

I. Introduction

In this contemporary Business world where people are running for money, time management has become a very important issue for everyone. There is lot of important information that people come across daily from posters on the road to product description on e-commerce websites. Remembering and processing extensive information is not only time consuming but requires techniques which are computationally expensive. To address this issue, the mode of communication or knowledge transfer has to be reformed where the amount of information communicated must be more with respect to the time consumed and the information must provide deep insights and must be easy to remember. There is so much information around us that it is impossible to understand and recall anything in a short span of time. However, using infographics to show data and statistics in the form of graphs, pictures, images, bulletins, etc., can help individuals absorb data in a much more efficient way.

Infographics are graphic portrayals of data, facts, information, or knowledge designed to easily and clearly display complex information. By incorporating graphic images, the perception of designs and patterns in the human visual system can be improved to an extent such that Infographics tell a tale as they help in organizing details and make it visually appealing and catchy, such that the audiences and other users can process, analyze, and interpret information quickly. Infographics display vast quantities of data and knowledge in the form of an information graph, flowchart or an image. They are used for many reasons such as they are fun to use and make learning enjoyable, eye-catching, succinct, and all the details they contain are easily absorbed by the reader, which makes them beneficial.

While the Web 2.0 is currently in use, a lot of additional technologies is continuously being added, which paves a way to the Web of data on the semantic standards of the Web continuous research, also referred to as Web 3.0. Web 2.0 was powered by Social Networks and cloud services while Web 3.0 is primarily based on newly developed technologies like Open Data Networks (ODN) and Semantic Intelligence. While Web 2.0 was powered by the emergence of smartphone, social and cloud services, Web 3.0 is based on three new levels of technical innovation: Ontology focused Computing, Knowledge based Computing and Semantic Inference. Semantic Web is an advancement of the existing Web which comprises of organized layers into a framework which are further modeled into Open Linked Data. The word Semantic means "processable information", and it is mandatory for the Semantic Web

* Corresponding author.
E-mail address: gerard.deepak.cse.nitt@gmail.com

to have a vocabulary with which both data and rules are articulated for Data Justification, which permits the export of entities for Knowledge Representation and Reasoning by information systems.

A unified data is always considered to be the uppermost priority for representation. A data or certain information which is not unified because of certain traits or features can cause clogging of thoughts while representing it. It is easy to understand and drive flexible solutions from a coalesced form of knowledge from heterogeneous sources. The key factor of a semantic paradigm is how the information is represented and reused. In the era of Web 3.0 where all entities are labelled, it is necessary to have some grounds where information is derived as useful auxiliary knowledge to be processed by information systems. In this era where there is limitless data, representation should be such that it allows machine to process the available information genuinely and provide accurate answers based on the queries imposed. Trending technological affairs drives towards a new approach of building solutions and developing Ontologies to represent information on the Semantic Web. The need for processable definitions and terms has now become a great requirement so that information extracted can be used and be manipulated according to the needs of the user and requirement of information systems. Knowledge Representation is a sub-basket of Artificial Intelligence dealing with interpreting, developing, and applying ways of expressing information on a machine such that programs can use this information for various purposes.

The semantically driven knowledge centric approach is entirely based on the paradigm of inferencing based on semantic similarity measures and diversity indices. Since the data on the World Wide Web is exponentially growing on everyday basis, it is almost impossible to learn from the contents on the World Wide Web. However, the Learning Based Approaches such as Machine Learning and Deep Learning Strategies, learn only from a sub-set of data, namely the dataset used in the approach. This transforms the problem as a closed domain problem, and specifically in Machine Learning, there is a need to choose or devise a feature selection approach which should perform well. However, in the Deep Learning paradigm, the entire approach is a Black Box where step by step computations are not visible. Moreover, owing to the large amount of Linked Open Data on the Web 3.0, it is highly complex to train a Learning Algorithm by accounting and preserving the Links between the entities. The semantic-based approach transforms the entire problem into an inferential scheme which is suitable for highly linked cohesive environment with a high data density like the Semantic Web.

OntoInfoG++ paves a way for eXplainable AI as it is based on semantic intelligence driven inferential paradigm. Learning Based Schemes do not promote eXplainable AI as eXplainable AI deals with breaking the Black Box involved in Machine Learning and Deep Learning Algorithms. Machine Learning Approaches however facilitate manual feature selection which can be configured separately in the algorithm and the strength of the algorithm can be improved. However, in Deep Learning even the feature selection is auto-handcrafted, and the computations for a specific set of data is a completely Black Box. eXplainable AI deals with solving a specific problem by formulating algorithms in which step by step computability can be reasoned out by human minds. OntoInfoG++ does not encompass Machine Learning or Deep Learning Algorithms, rather it makes use of Semantic Intelligence Driven Reasoning Schemes and transforms the problem into an inferential open domain problem. The entire computations which happen in the proposed approach is seen as a white-box and ensures human minds to reason out thus supporting eXplainable AI.

A. Motivation

Research on infographics over the past decade has been mainly focused on the role of the use graphical representation as an attention-grabbing strategy. Infographics have been a helpful tool in many areas of domain from education to advertisements. Nowadays for specific topics search engine yields a lot of infographics for the input topics, but how much of this is relevant? In order to solve this issue, a proper recommendation model has to be devised to furnish relevant content. Though infographics exists as images, traditional image processing or visual similarity-based recommender techniques cannot be used, as the queries are in the form of text and an annotation-based approach is the need of the hour. This can be achieved using a Semantically Driven approach, which would arrange the infographics relevant to that topic of user interest in a chronological order. The Semantically driven approach is responsible for giving practical and logical representations that can give more sensible solutions when compared to the conventional recommender systems that use only basic feature extraction techniques, where the real-world knowledge will not be taken care of. The traditional recommendation systems give solutions that are non-practical and may not generalize easily to all the topics of the same problem as learning the huge volumes of data from the web is infeasible. In a semantically driven approach, as the real-world knowledge is integrated from several heterogeneous sources, entities will be populated such that there will be more context terms with a high information density that will be added. The World Wide Web houses several Knowledge Sources which when incorporated increase the density of the knowledge, and thereby facilitate enriching the supplementary knowledge such that the synonymy, polysemy, cold-start, serendipity, context irrelevance, and cross domain data sparsity problems can be solved.

B. Contribution

A semantic approach for an Annotations Based Infographic recommendation has been proposed in this paper. OntoInfoG++ is a Knowledge-centric approach which uses real-world knowledge from various heterogenous sources. The OntoInfoG++ uses both user query and user clicks which are preprocessed and are formulated as a query word set and are collated to form a user initial set of user topic of interest. A Knowledge Graph is formulated by subjecting the query word set to topic enrichment using BibSonomy, DBpedia, Wikidata, LOD Cloud, and Crowd Sourced Ontologies. The titles and keywords extracted from the dataset are utilized to formulate Ontologies which facilitates in semantic concept alignment with the formulated knowledge graph, to yield the enriched topic of interest knowledge graph. The semantic similarity is computed with the help of EnAPMI, Horn's Index and information entropy between the enriched topic of interest knowledge graph and keywords extracted from infographics extracted from research papers. A story of infographics is created and it is recommended based on the scores obtained from the semantic similarity measures.

C. Organization

This paper is organized as follows. Section II addresses the relevant research work done related to this area. Section III depicts the Problem Definition and Assumptions. The proposed methodology is represented in Section IV. Section V composes the implementation details. The performance evaluation and results are depicted in Section VI. The paper is concluded in Section VII.

II. Related Work

Siricharoen et al. [1] have briefly explained how infographics were used in journalism, and also how infographics serves as better mode for effective communication in the digital age and have put forth the history, significance, and benefits of infographics and tools for making infographics more beneficial and effective. They have also addressed the suggestive guidelines of infographics creation. Sujia

Zhu et al. [2] have reviewed and classified automatic tools that cater to visual recommendations for visualizing storytelling, visualizations of graphs, visualization of annotations, and visualization of information networks in several varied perspectives. They have also posed many obstacles and directions for potential work in the field of automated infographics and visual recommendations. Wilkinson et al. [3] have performed content analysis obtained evaluating Diet-Related Infographics and have used it for Behavior Change Theories. This was implemented by pin creation that makes use of both pictures and textual descriptions to portray elements that convey information about nutrition.

Featherstone [4] has proposed how infographics is used as a primer and supported it with visual data and briefly explained about the tools used for the same. Mohd Noh et al. [5] have discussed how Infographics is implemented as a training tool to assist teachers in education and learning sessions to allow student and teachers understand and interpret concepts with ease. Siricharoen et al. [6] have addressed the critical aspects assessment approach for infographics, which is discussed briefly with questions. Murray et al. [7] have studied and discussed about some basic concepts for the design of effective infographics and have proposed some suggestions for the development of engaging infographics. Cifci et al. [8] have studied how Infographics affect students' achievements. The analysis is very significant which leads to designing instructional materials which can be used in classrooms. They have implemented the research as a quasi-experimental study, one of the quantitative methods of study. They have also inferred that the use of infographics in geography lesson improved academic performance.

Nuhoglu et al. [9] have researched how infographics can be generated as a scheme for visualizing content for interactive learning of scenarios by incorporating a design for infographics which caters to a collaborative technology based Bridge21 learning model which has been proposed to foster learning. Chen et al. [10] have proposed a deep learning driven strategy that retrieves a timeline template from the images which are quite magnanimous. They have adopted a deconstruction and reconstruction technique. Cui et al. [11] have put forth a strategic infographic generation technique and have built a system that synthesizes statements relevant to statistics to a potential infographic which is obtained from previous studies. However, there is an emphasis on the aspect of modeling infographics for statistics as a potential domain. Mackinlay et al. [12] presented an approach for designing graphical presentation on the basis of the perspective, that graphical representations are outcome based phrases resultant from graphics-based languages. They also introduced a model of a presentation method called APT with AI techniques focused on algebra and graphic design requirements.

Deepak et al. [13] have developed a Web 2.0 complaint RDF driven model that focuses on decreasing the irrelevance and promoting diversity in the results from semantic search. Indicator terms were yielded by computing dyadic RDF entities from a set of webpages for which an RDF polarization vector is derived from the inferencing of the modeled term-frequency matrix and term co-occurrence matrix. Middleton et al. [14] have built ontological models for profiling recommendation, namely, Quickstep and foxtrot. An ontological interference model has been employed for the improvement of the performance and also encompasses external ontological entities for achieving profile base bootstrapping. Furthermore, the visualization of user profiles to yield relevant feedback has been proposed. David Werner et al. [15] have designed an ontology based multi-layer recommendation system for economic articles based on a client's profile, that produces a magazine per customer composed of a set of daily produced articles. The main aim of the developed system was to reduce the overload of useless issues.

Peis et al. [16] have focused on reviewing semantic recommender systems based on classification criteria, ontological and conceptual diagrams, which have been proved to be effective for research and experimentation. Pazahr et al. [17] have designed an advertisement recommender system that has been semantically enhanced, and at the same time produces recommendations in a simplified manner. The proposed architecture uses semantic logic to showcase the recommended products and this in turn can differentiate between the recommender unit from the classical recommender methods. Prafulla et al. [18] have proposed an approach that makes uses of semantic clustering for task recommender systems to identify right personnel. The method utilizes a feature extraction scheme which is based on generation of synsets and a strategy of semantic clustering which is iterative is nature. The approach also cognitively maps synonyms which helps in yielding a better performance. The approach also solved the issue of scalability with reduced entropy. Huijsduijnen et al. [19] have designed a model, Bing-CSF-IDF+, a content-based RS for news which is semantically driven. They have compared the performance with a previously designed version, Bing-SF-IDF+, and found the former to outperform the latter by statistical means like F-measure and kappa. The approach uses concepts and relationships from domain ontology, synsets and synset relationship from semantic lexicon from WordNet. Tymchenko et al. [20] have proposed a multifactor selection scheme for the design of infographics. This is a hierarchical approach that uses a pair comparison model for the evaluation of higher levels of interconnection elements using factor comparisons between a pair of elements for infographics design.

Ontologies have played a pivotal role in improving the recommendations in web search systems. Owing to the reason that Ontologies provide a good amount of supplementary knowledge in improving the context and the scope of the query, they can be employed in recommending infographics. Ontologies have significantly increased the recommendation relevance in [21], [22], [23], [24], [25] and have solved the problem of polysemy, ambiguity, and context irrelevance. However, Ontologies, once modeled, need to be constantly monitored for quality and must be updated. Also, there can arise a scenario where the Ontology would not be able to dispatch the auxiliary knowledge for a specific query. In such cases, dynamically modeled metadata or dynamically extracted relevant entities to the query need to be supplied which is addressed in this paper by infusing entities from a wide array of knowledge sources and building a knowledge graph from the user query and the knowledge sources.

Berkani et al. [26] have proposed a semantically driven approach that depends on social representation of user profiles for recommendations of user's profiles and have also employed two categorization strategies in order to optimise the performance of the recommendations: using the K-mean algorithm (originally utilised for everyone) and K-Nearest Neighbors method (applied to newly added users). Javed et al. [27] have proposed a context-aware recommender system for filtering things related to the user interest, as well as a context-based recommender system for recommending those things. Their context-based recommender system extracts patterns from the World Wide Web based on the user's previous interactions and delivers recommendations for future news. Houari et al. [28] have proposed a domain specific tool for recommending experts using N. S., & PROMETHEE II and Negotiation in support of industrial maintenance.

Bobadilla et al. [29] have incorporated Deep Learning schemes for enhancing the quality of recommendations and in Recommender Systems. However, there is a need to arrive at techniques that either eliminate learning paradigms or break the Blackbox in machine learning and deep learning strategies in order to provision eXplainable AI.

Analysis of the Literature clearly points out that the infographics are quite useful in rendering knowledge and there is a need for

an infographics recommendation system. However, the existing infographics recommendations are either based on simple ontologies or semantic logics or even focus on very few parameters like the user-clicks and the user query or infuse learning algorithms from the Web. They either make the system complex when learning algorithms are infused or depict a lacuna as enough knowledge is not infused into the system. When clustering alone is the focus, then the approach results in high amounts of error rates. However, this can be solved when the right semantic similarity techniques are infused via semantic agents, and numerous entities are dynamically infused into the system.

III. Problem Definition and Assumptions

A. Problem Definition

Given a query which specifies at least one real-world entity, a dataset comprising research papers, an access to several real-world knowledge sources, and the user-click information, the first objective is to model the user topics of interest from the query words, user-click information and the entities from several real-world knowledge sources. The second objective is to model Ontologies using titles and keywords extracted from research papers from the dataset. The third objective is to achieve semantic concept alignment between the formulated topic of interest knowledge graph and the initially formulated Ontologies. The final objective is to furnish the infographics by computing the semantic similarity and arranging them in a chronological order in the increasing order of the semantic similarity and recommend to the user, until there are no further user-clicks recorded.

B. Assumptions

The Ontologies modeled from the keywords and titles from the research papers must be strictly adherent to the papers present in the dataset and must be free of inconsistencies. The modeled Ontologies must be at least a strong representation of the upper-level ontologies. The user-clicks used in the approach must also be adherent to the domains in the dataset. The Queries must have at least one strong entity and must not exceed to more than 12 strong entities. It is a mandatory requirement that the dataset be categorical in nature and the Infographics must be strictly labelled or annotated.

IV. Proposed Methodology

The architecture of the proposed Semantically Driven Infographics Recommendation System, the OntoInfoG++, is depicted in Fig. 1 which is knowledge centric and constitutes a large amount of real-world knowledge from varied heterogenous sources. The OntoInfoG++ is also driven by the user query and imbibes the current user-clicks into the system. The user query is subject to initial pre-processing which constitutes of the Tokenization, Lemmatization, Stop Word Removal, and elimination of special characters. The preprocessed user queries are formulated as a query word set which is collated along with the current user clicks of the user to formulate an initial set of User-Topic of Interest.

The initial set of User-Topic of Interest is subject to topic enrichment by aggregation of auxiliary domain knowledge from real-world data stores, namely the BibSonomy, DBPedia, Wikidata, LOD Cloud, and Crowd Sourced Ontologies. Since the infographics recommendation comprises of recommending knowledge containing graphics and diagrams, there is a need for integration of knowledge from BibSonomy. The DBPedia and Wikidata further remove anomalies and integrate detailed auxiliary knowledge into the existing entities and facilitate linking of newer entities. The reason for including Crowd centric domain ontologies is to further enhance the density of information linked to the exiting entities and to yield a humanized perspective to the user query. The inclusion of domain centric ontologies helps in adding a broader human centric perspective to the query and also enhances the density of knowledge which enhances the diversity of results without much deviation from the user-interests and the query topic.

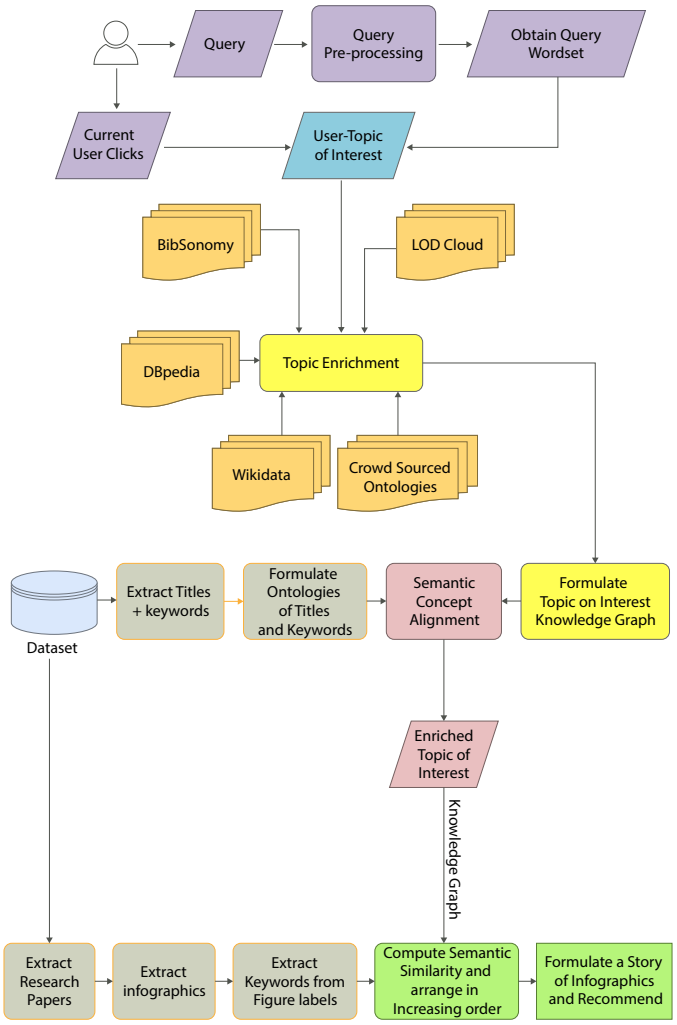

Fig. 1. Proposed System Architecture.

The Enriched Topic of Interests which is inclusive of cognitive real-world knowledge from several knowledge stores is modeled into Knowledge Graphs which is undirected and at least a single link is established between its constituent entities. The knowledge Graph simulates query relevant metadata which has been loaded from the Web 3.0. The User Topic of Interest Knowledge Graph serves as a liaison between the dataset and the user and integrates the categories in the dataset. OntoInfoG++ is a knowledge centric annotations-based infographics recommendation system where annotations play a vital role in formulating a sequential story comprising of the infographics and thereby recommend a collection of similar infographics which are relevant to the query and is suitable for satisfying the needs of the user. The titles and the keywords from the research papers are extracted from the dataset and an Ontology is formulated by creating links among the keywords in the research papers and meaningful words in the title of the paper.

A strategic Semantic Alignment of Concepts from the Ontology constituted from the Title and Keywords with that of the User Topic of Interest is carried out to further Enrich the amount of Knowledge and the Ontology- Knowledge Graph pair is associated with each

other by creating links between its concepts formulating a larger Knowledge Graph constituting Topic of Interest. Further, the research papers from the dataset are loaded, parsed, and figures are extracted based on an Agent which recognizes a specific figure is an infographic or not. There is also a term identification agent which eliminates performance graphs and keeps the infographics intact. System Architectures or diagrams describing a specific scenario or schematic or block diagrams are considered as infographics. Language Parsing Rules and matching with a set of terms is incorporated in the agent for extraction of infographics. The Semantic Similarity is computed between the Enriched Topic of Interest Knowledge Graph and the terms extracted from figures in the research paper. This is continued for all the infographics extracted. The terms are extracted in the increasing order of the Semantic Similarity and the infographics are also rearranged in the same order to formulate a story and arrange the infographics in a meaningful manner.

$$H_0 = \frac{\Sigma(p_{ij}+p_{ik})\log(p_{ij}+p_{ik}) - \Sigma p_{ij} log p_{ij} - \Sigma p_{ik} log p_{ik}}{2log2} \quad (1)$$

The Horn Index Ho as depicted in Eq. (1) is used to compute the Semantic Similarity or the Overlap of entities by computing the Ho between the instances in the Knowledge Graph and the keywords from the figure labels. When adapted into a semantic environment comprising of an Open Linked Data as knowledge graphs, the pij refers to the semantic similarity of terms i & j and pik refers to the semantic similarity between the terms i & k, such that 'i' corresponds to a term in label set of the infographic images while 'j' & 'k' correspond to the terms that are having a single link in the knowledge graph, which signifies they are adjacent to each other. Traditionally the Horn's Index used the proportion resource utilized by species of a type but when it is adapted in an Environment of the semantically driven information systems, it is substituted by the Semantic Similarity without having to modify the exact index. It must also be noted that any base of the logarithm can be used but it also must be ensured that a uniform logarithm base must be maintained throughout. The Horn's Index furnishes a value between 0 and 1, and the threshold is assumed as 0.5 for the Horn's Index for accepting and rejecting entities for recommendation.

The Semantic Similarity within Horn's Index is computed using the Normalized Pointwise Mutual Information (NPMI) measure by considering only the magnitude of NPMI measure which lies between 0 and 1. If the NPMI furnishes a negative value, only the magnitude is considered while the negative number is ignored. The reason for choosing NPMI over EnAPMI within the Horn's Index is due to the fact that NPMI is computationally less expensive than the EnAPMI. Moreover, since the entities after being subjected to NPMI are further being passed into the Horn's Index, as a result the less stringent NPMI would be sufficient at this case. However, the threshold for NPMI is empirically considered as 0.75 to allow only the entities that are highly relevant through the NPMI measure into the Horn's Index. Furthermore, this is also the reason to keep the threshold of the Horn's Index to 0.5 as NPMI is much more stringent and Horn's index need not be as stringent as already relevant entities are passed into it for a further approval. Eq. (3) depicts the NPMI measure which is based on the Pointwise Mutual Information score (PMI). PMI is a knowledge analysis and statistics indicator used to measure the relationship between terms. The PMI score is based on the Eq. (2). The PMI Score is normalized such that the values occur between [-1, +1], resulting in -1 for uncooperative incidents, 0 for isolated events and +1 for co-occurrence and the equation of NPMI is as shown in the Eq. (3).

$$\text{PMI (X, Y)} = log \frac{P(x,y)}{P(x)P(y)} \quad (2)$$

$$\text{NPMI} = \frac{PMI}{\log P(x,y)} \quad (3)$$

$$\text{EnAPMI(m,n)} = \frac{Pmi(n,m)}{p(m)(n)} + y - \eta \quad (4)$$

$$y = \frac{1+\log[p(m,n)]}{p(n)\log[p(m)] - p(m)\log[p(n)]} \quad (5)$$

$$\eta = \frac{\log[p(m).P(n)]}{\log(p(m,n))} \quad (6)$$

$$\text{H(X)} = \sum_{i=1}^{n} P(x_i)\log P(x_i) \quad (7)$$

EnAPMI is an Enriched Adaptive Pointwise Mutual Information measure (EnAPMI), a novel semantic similarity, which enriches the Adaptive Pointwise Mutual Information (APMI) which is a model based on PMI for the measurement of semantic similarities based on the likelihood of the event and terms were suggested to be co-occurring with an adaptive coefficient. The EnAPMI is as shown in the Eq. (4) between a pair of terms m and n. The EnAPMI measure which belongs to a class of the variants of the PMI models and is associated with an adaptivity coefficient y and a drift indicator η which has been employed to estimate the semantic drift between a pair of terms. The EnAPMI measure is derived from the APMI measure by eliminating the drift indicator from the APMI measure and adding the adaptivity co-efficient to the existing APMI model. Eq. (5) includes the adaptivity co-efficient of a pair of terms m and n, which is coupled with a logarithmic quotient of the probability of co-occurrence of the pair of terms m and n in its numerator.

The product of probability of occurrence of a term with the logarithm of the standalone probability of individual occurrence with its pair term is computed and their difference is included in the denominator of the adaptivity co-efficient. The reason for coupling the adaptive co-efficient with the variant of the PMI is primarily because the semantic relatedness between a term pair can be computed with greater efficacy when the word pair co-occurrence probability, the probability of word occurrence, and when the measure of self-information has been taken into consideration. The drift indicator η as depicted in Eq. (6) can be described as the ratio of the logarithm between the individual occurrence probability of the term pairs to that of the ratio of the probability co-occurrence ratio of the term pair. The semantic gap between a pair of terms is quantified and measured by the drift indicator which is computed between the pair of terms and is eliminated from the EnAPMI measure. The EnAPMI is derived from the APMI measure and the EnAPMI model acts as a better performing semantic similarity model between the pair of terms when compared to the other PMI based conventional models in a highly cohesive semantic environment. Eq. (7) depicts the information entropy H(X) which depicts the average quantity of information, inherent at the interval of the potential outcomes of the variable. X is the discrete variable, with possible outcomes {x_1, x_2, ..., x_i} which occur with probability {{P(x_1), P(x_2)....., P(x_i)}} and this is formally defined as information entropy represented in Eq. (7). The Entropy depicted as H(X) in Eq. (7) is the product of the probability of occurrence of a term over a web corpus with that of the self-information in the term. The Information Entropy is also used as a standalone measure for computing the relevance of the entities as the degree of information associated with an individual term over a corpus, which serves as a potential indicator to estimate the extent up to which the presence of the term creates an impact in the specific corpus.

V. Implementation

The implementation was carried out using JAVA as the language of choice with Eclipse as the preferred IDE. The reason for using JAVA is the ease of integration with AgentSpeak and JADE which were used to model the agent to compute the Entropy, Semantic Similarity, and the Horn's Index. The experimentation was conducted on the RARD

II: Related Article Recommendation Dataset which can be accessed from the Mr. DLib (http://mr-dlib.org). The RARD II dataset comprises of 94m recommendations which covers an item space of 24m unique items. The unique terms in the RARD II dataset were linked with google scholar to obtain the relevant research papers in full text mode and was stored in the linked repository. The research papers which were available in full text mode only were used for experimentation via google scholar and institutional repository for full text research papers. The reason for choosing the RARD II Article Recommendation Dataset is primarily for the only reason that terms based on research topics are available in the RARD II dataset, and the topic linked infographics can be extracted from the research papers in a sequence. A Language Processing Agent is also modeled for parsing the figures and ensuring that it is an infographic or, if it is a graph, that is based on the terms which are used to label the figures and the textual description of the figures. The state of the agent is described to extract the infographic images along with the image labels from the Research Articles and creates a categorical state space comprising of infographic images, the keywords in the labels, and other associated annotations. This enables the OntoInfoG++ to yield infographic images that are being queried by the user to satisfy the information needs of the user pertaining to the topic of interest of the user.

The Crowd Sourced Ontologies are generated by picking up terms from the RARD II dataset based on the keywords from the articles and from those in the labels of infographic images, and subject them to the OntoCollab [30] framework which facilitates dynamic generation of OWL and RDF Ontologies which have been hierarchically arranged, axiomatized, and reasoned out. Apart from the Domain Ontologies which have been generated using the OntoCollab, user modeled ontologies were also included. Web Protégé was used for manual modeling of domain specific ontologies. The Crowd Sourced ontologies were also collected from various online research communities and were curated into a meaningful Ontology using OntoCollab. Moreover, the index terms from Semantic Wikis were considered, these are Crowd Sourced at a large scale, extracted from user-blogs, several research articles, and from portals where user-skills and technology are emphasized. A major portion of Crowd Sourced Ontology is automatically generated by OntoCollab by facilitating access to these user-centric sources. The auxiliary knowledge is supplied into the OntoInfoG++ framework from BibSonomy, Wikidata, DBpedia, LOD Cloud, and Crowd Sourced Domain Ontologies. The reason for using an array of knowledge stores or factual knowledge bases is because diversified entities can be included to increase the density of query relevant knowledge. SPARQL Endpoints designed using AgentSpeak are integrated into the environment of OntoInfoG++ which queries Entities from several real-world knowledge stores like BibSonomy, Wikidata, DBpedia, and LOD Cloud. The reason for combining different knowledge sources is to increase the variety and heterogeneity of entities to provide auxiliary knowledge into the proposed paradigm. Moreover, the incorporation of knowledge from varied sources increases the diversity of results. The OntoInfoG++ integrates entities that are relevant to the user topics of interest and helps in topic enrichment and solves the Serendipity problem by entity integration from distinguished knowledge sources. OntoInfoG++ individually harvests topic relevant entities from various knowledge sources and Crowd Sourced Ontologies and further integrates together to facilitate topic enrichment and accelerate diversification of results.

The entities are supplied into the OntoInfoG++ Framework as collective knowledge for enrichment of the Query Terms and the User-Clicks. Among several domains which were used for experimentations, Table I documents 12 distinct and standard domains and the number of concepts and individuals in the domain. It is indicative from Table I that the number of individuals is much higher than the number of concepts. However, these concepts which are depicted in Table I comprise of the core concepts, specialized concepts, upper ontologies, and the sub-concepts which are hierarchically arranged. The individuals are the implementations of the specialized concepts which are used in experimentation.

TABLE I. Details of 12 Domains Along With the Number of Concepts and Individuals Used for Experimentations

Domains	No. of Concepts	No. of individuals
Agriculture	1745	4452
Horticulture	1245	3845
Library Science	1223	4986
Information	1435	3754
Economics	2135	3121
Sociology	845	1259
Humanities	1121	2456
Cloud & Distributed Computing	895	3856
Robotics	969	3254
Urban Planning and Sustainability	1921	4512
Life Sciences	1895	3695
Chemical Engineering	2032	4875

Moreover, the diversification of entities would result in diversified and yet relevant infographics increasing the spectrum of visibility of the diversified infographics under the purview of the topic without major deviation. In order to create a benchmark query-set for the RARD II dataset, the metadata was harvested. 124 users were given broad area topics from the RARD II dataset and were asked to use Google Scholar, other research paper search engines, and the standard knowledge stores, and were asked to formulate queries and also yield the ground truth infographic keywords of image labels for the formulated query. There were 2457 queries with the ground truth which were collected by the user participants over the period of 168 days. The manually modeled Ontology and the Dynamically generated Ontology were merged into a single Crowd Sourced Ontology and were used for experimentation. An end-to-end SPARQL agent was encompassed to obtain the metadata from the individual knowledge stores.

OntoInfoG++ is a knowledge centric paradigm for recommendation of infographics and is semantically compliant. The OntoInfoG++ strategy formalizes Topic of Interest knowledge graph which is a constituent of the query, the current user-click, and auxiliary knowledge from real-world knowledge bases. Also, the semantic alignment has been encompassed into the system using three distinct measures namely the EnAPMI measure, the Horn's index, and the Entropy. The reason for encompassing three distinct and yet effective measures is to increase the relevance. Moreover, EnAPMI is based on the probability of the occurrence and co-occurrence of words over the data corpus. The entropy computes the information measure, thereby the most informative entity in correlation of the environment in which it is contained is preferentially selected. The usage of an Agent for the infographics extraction and the ranking of the infographics based on the computation of the semantic similarity between the labels of the infographics and the enriched Topic of Interest ensures that the infographics are arranged in a chronological order and enables the users to deduce inference and also a small story gets created as soon as the infographics are arranged in a logical order. The OntoInfoG++ algorithm is depicted in Table II.

TABLE II. ALGORITHM I

Algorithm 1: Proposed OntoInfoG++ Algorithm for Infographics Recommendation

Input: Multi-word Query, Current User-Clicks, Access to Real World Knowledge Bases, Crowd Sourced Ontologies, Categorical Dataset S.

Output: Recommendation of Infographics in a chronological interpretable order

Begin

Step 1: The query Q input is subject to pre-processing constituting tokenization, lemmatization, and stop word removal to yield query word set Qs.

Step 2: The current user-clicks recorded based on the user navigation are also pre-processed and are merged with Qs to yield User Topics of Interest ToI.

Step 3: while (ToI.next()!=NULL)

Set Le ⟵ Load Entities from Real World Knowledge Stores like BibSonomy, DBpedia, Wikidata, LOD Cloud, and Crowd Sourced Ontologies.

end while

Step 4 : for each entity in Le

Formulate a Topic of Interest Knowledge Graph ToI_KG by computing the semantic similarity between Instance Pairs and Rearranging them By At Least Having a Single Link among each of the instances in the KG

end for

Step 5: Extract Keywords and Titles from S and formulate Ontologies of Titles and Keywords as OKnow.

Step 6 : Semantically Align the concepts in OKnow and ToI_KG using the SemantoSim, Horn's's Overlap Index, and Entropy to yield Enriched ToI_KG as EnToI_KG.

Step 7: Extract Research papers from the S, and thereby parse the Infographics using an infographic recognition agent and Load the Infographics and Infographic Keywords as InfoG_Keywords.

Step 8: for each entitiy in the EnToI_KG

Compute SemantoSim (EnToI_KG.currentEntity(), InfoG_Keywords())

if (SemantoSim.curr()>0.75)

HashMap RecInfoG ⟵ (InfoG_Keywords, SemantoSim Measure)

end

Step 9: Arrange RecInfoG in the increasing order of SemantoSim, and recommend the corresponding infographic to formulate a chronological order and tell a story relevant to the query words.

Step 10: Record the current user-clicks and formulate the ToI and continue Steps 2 to 9 until there are no further user-clicks recorded.

VI. RESULTS AND PERFORMANCE EVALUATION

The Performance of the proposed OntoInfoG++ was evaluated using the Precision, Recall, Accuracy, F-Measure, False Discovery Rate (FDR), and the Normalized Discounted Cumulative Gain (nDCG) as the potential metrics. Precision, Recall, Accuracy, and F-Measure compute the relevance of results to the query as well as the user-interests. The FDR indicates the number of false positives recommended by the system. The nDCG measures the diversity of recommendation of results to quantitatively indicate the degree of diversity in the recommended infographics. Precision, Recall, Accuracy, and F-Measure are indicated by Eq. (8), Eq. (9), Eq. (10), and Eq. (11) respectively. Eq. (12), Eq. (13), and Eq. (14) represent the FDR, nDCG, and the Discounted Cumulative Gain respectively.

$$\text{Precision}=\frac{\text{Retrieved}\cap\text{Relevant}}{\text{Retrieved}} \tag{8}$$

$$\text{Recall}=\frac{\text{Retrieved}\cap\text{Relevant}}{\text{Relevant}} \tag{9}$$

$$\text{Accuracy}=\frac{\text{Proportion Corrects qualifying ground truth test}}{\textit{Total No.of Queries}} \tag{10}$$

$$\text{F-Measure}=\frac{2(\text{Precision}\times\text{Recall})}{(\text{Precision}+\text{Recall})} \tag{11}$$

$$\text{False Discovery Rate}= 1-\text{Positive Predictive Value} \tag{12}$$

$$\text{nDCG} = \frac{\text{DCG}\alpha}{\text{IDCG}\alpha} \tag{13}$$

$$DCG = \sum_{i=1}^{\alpha}\frac{Rel_i}{log(i+1)} \tag{14}$$

From Fig. 2, it is easily inferable that OntoInfoG++ framework has yielded an overall Precision of 98.12%, an overall Recall of 96.4%, an overall Accuracy of 97.21%, and an overall F-Measure of 97.27 % . It can be easily interpreted from Fig. 3, that the proposed OntoInfoG++ framework furnishes a FDR of 0.02 with an nDCG of 0.95.

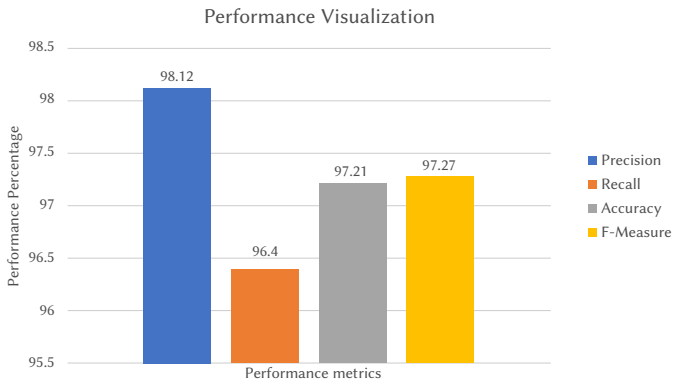

Fig. 2. Performance of the Proposed OntoInfoG++ for Infographics Recommendation.

The reason for the high values of Precision, Recall, Accuracy, and F-Measure is owing to the reason that the OntoInfoG++ is a knowledge centric paradigm for recommendation of infographics and is semantically compliant. The OntoInfoG++ strategy formalizes Topic of Interest knowledge graph which is a constituent of the query, the current user-click, and auxiliary knowledge from real-world knowledge bases. Moreover, OntoInfoG++, is a infographics recommendation approach that takes into consideration both the user-query and the current user-clicks of the user. It amalgamates topics of user interest from varied sources from which the entities are sourced. The sources of user interests include contents from BibSonomy, DBpedia, Wikidata, LOD Cloud, and Crowd Sourced Ontologies pertaining to varied domains which are being considered for experimentation. The domains are chosen in a way such that it is recurrent in the dataset. The approach also considers the elements from the dataset which include the titles from the dataset and the keywords. Furthermore, the approach specifically formulates the Ontologies inclusive of titles and keywords which are subject to semantic concept alignment from the knowledge base which was formulated initially from the topic of interest. The proposed OntoInfoG++ enhances the relevance of results predominantly considering the fact that it uses Horn's Index, EnAPMI measure to compute the semantic similarity, and the Information Entropy separately for computing the semantic relevance of results.

The use of a system of three different and yet comprehensive measures ensures that relevance of results is maintained and is non-deviated with respect to the user query, query-clicks, and the topic of interest of the user is adhered strictly.

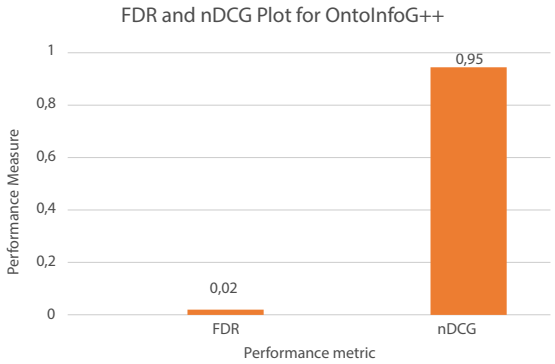

Fig.3. FDR and nDCG Plot.

The performance of the OntoInfoG++ is compared with the base line approaches as shown in Table III. Since infographics recommendation is quite new, there are not many baseline models available, except for the PCMSI. However, the other famous models in image recommendation were considered and implemented for the infographics and were taken as baseline models, namely the Collaborative Filtering, Fuzzy c-means clustering, and the CNN-K-Means Clustering, respectively. The PCMSI used pairwise comparisons and graphs for selection of infographics where the hierarchical evaluation of elements has been realized for recommendation of infographics. To evaluate the PCMSI in the proposed environment, the model has been used to recommend infographics in the exact same environment of the OntoInfoG++. PCMSI is quite fascinating in its approach, however the lack of metadata into the approach makes it linger in its performance when comparison to the OntoInfoG++. The CNN K-Means clustering when implemented in the environment of OntoInfoG++ makes it computationally expensive and learning the annotations makes it lag. The K-Means when coupled with CNN does not do wonders to the performance. The combination of Collaborative Filtering with Fuzzy c-Means clustering also could not perform well as it exhibited cold start problem and there was sparsity in the recommendation results.

The low value of the FDR is a clear indication that the proposed OntoInfoG++ performs well with a high degree of efficacy. The primary reason for the low FDR value is mainly due to the incorporation of auxiliary knowledge by aggregating entities from various knowledge bases and fact stores which are multi-faceted. Also, inclusion of entities from Wikidata, DBpedia, BibSonomy, LOD Cloud, and Crowd Sourced Domain Ontologies enhances the density of knowledge and thereby solves the serendipity problem in infographics recommendation. The inclusion of EnAPMI, Horn's Index, and Entropy together facilitates the integration of entities that are relevant in all aspects with respect to the user query and the user preferences in terms of query click. The reason why three different strategies with different perspectives are used is mainly for the reason to filter out and eliminate the false positives to a large extent such that diverse and yet highly relevant entities to the query and the user-preferences are to be retained. The Precision, Recall, Accuracy, F-Measure vs the number of recommendations are plotted in Fig. 4 (a), (b), (c), (d) respectively. From the Fig.4 (a) it is very clear that the precision of the proposed OntoInfoG++ is 10.44% higher that the PCMSI [23] when compared to the Collaborative Filtering with Fuzzy c-means clustering. The precision of OntoInfoG++ is 13.41% higher than that of Collaborative Filtering with Fuzzy c-means clustering. When compared with that of CNN-K-Means Clustering the precision is 89.43% higher. From Fig. 4 (b) it is inferable that the recall of the proposed OntoInfoG++ is 11.48%, 16.22%, 16.57 % higher than that of the PCMSI [23], Collaborative Filtering with Fuzzy c-means clustering, and CNN-K-Means clustering respectively. Seeing the plot in Fig. 4 (c) it is inferable that the accuracy of OntoInfoG++ is 10.93%, 14.83%, 15.98% higher than PCMSI [23], Collaborative Filtering with Fuzzy c-means clustering, and CNN-K-Means clustering respectively.

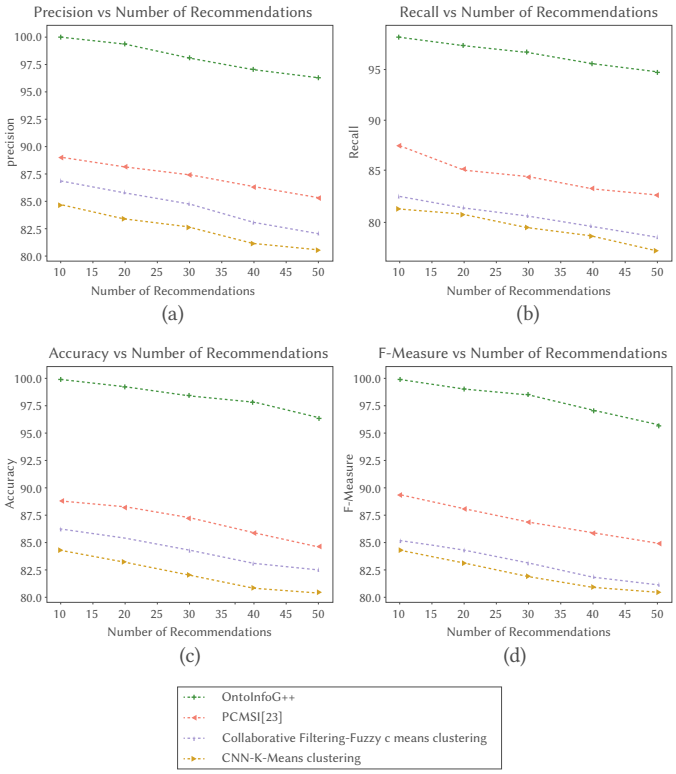

Fig.4. Performance metrics vs number of recommendations.

From the Fig. 4 (d) it is inferable that the F-Measure of the proposed OntoInfoG++ is 10.99%, 14.89%, 16.04% higher than that of the PCMSI [23], Collaborative Filtering with Fuzzy c-means Clustering, and CNN-K-Means clustering, respectively. As the number of recommendations increases, the curve for precision, recall, accuracy and F-Measure vs Number of recommendations also tend to decrease naturally. The reason why it decreases is because, as the number of recommendations increases, the irrelevance in the recommendations increases. However, the relative performance of the proposed OntoInfoG++ is higher when compared to the baseline models and benchmark approaches irrespective of the number of recommendations.

TABLE III. Comparison of Performance of the Proposed OntoInfoG++ With Other Approaches

Search Technique	Average Precision %	Average Recall %	Accuracy %	F-Measure	FDR
PCMSI [23]	87.68	84.92	85.87	86.28	0.13
Collaborative Filtering-Fuzzy c means clustering	84.71	80.18	82.45	82.38	0.16
CNN-K-Means Clustering	82.69	79.83	80.46	81.23	0.18
OntoInfoG++	98.12	96.4	97.21	97.27	0.02

The reason for this superiority of performance of the proposed approach is because the PCMSI [23] is a graphical approach that uses spectrum based comparison, which requires a proper graph to be modeled and pairwise relations need to be computed. The deviations occurring while calculating the pairwise relations results in a fair amount of un-correlated associations, and in Collaborative Filtering with Fuzzy c-means clustering approach there will be cold start problem and data sparsity problem and also, as it is using Fuzzy c-means Clustering with it, it makes the approach more computationally complex and, as the ratings can definitely differ, it is not a feasible technique. In the CNN-K-Means clustering, CNN is a learning algorithm which takes in hand crafted features, thereby increases the learning load of the model. All these three approaches do not use any form of real-World Knowledge sources to learn entities from the World Wide Web and there is a very low amount of data and very low amount of information and therefore all the three approaches lack in diversity. The proposed OntoInfoG++ is a lightweight inferential paradigm as there is dynamic computation of semantic relatedness using software agents, which make it quite efficient.

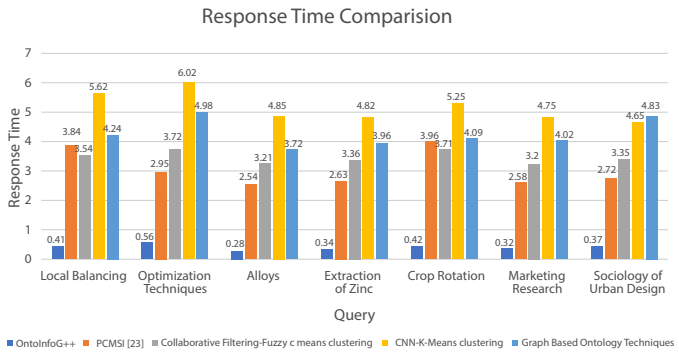

Fig. 5. Response Time Comparison of the Proposed OntoInfoG++ for Infographics Recommendation.

The comparison of response time of the OntoInfoG++ framework for a set of 7 distinct queries for recommendation of infographics is depicted in Fig. 5. Although, the overall evaluation of average response time computation is done for all the 2457 queries, the comparisons are tabulated only for 7 queries each of them from a distinct domain, by including the minimum and the maximum response time in the table. It is clear that OntoInfoG++ has a range of response time between 0.32 ms and 0.56 ms, while the PCMSI has recoded the response time between the range of 2.54 ms and 0.39 ms. However, the Collaborative Filtering with Fuzzy c-means clustering has recorded a response time in the range between 3.2 ms and 3.54 ms. The CNN with K-means clustering has recorded a response time in the range between 4.65 ms and 6.02 ms. The Graph Based Ontology Technique has recorded a response time in the range between 3.72 ms and 4.98 ms. It is quite evident and clear that from the tabulations of response time yielded by several approaches, the OntoInfoG++ has recorded the lowest average response time of 0.38 ms for 2457 queries. The PCMSI, Collaborative Filtering with Fuzzy c- means clustering has recorded an average response time of 3.05ms and 3.45 ms respectively for 2457 queries. The CNN with K-Means Clustering and Graph Based Ontology Techniques have recorded an average response time of 5.14 ms and 4.26 ms respectively for 2457 queries.

The reason why OntoInfoG++ has the lowest application response time is mainly due to the fact that it is an Agent Centered Approach and does not use a learning algorithm or a learning scheme for recommendation. Instead, the OntoInfoG++ is built on an inferential mechanism which uses three tactical approaches namely the EnAPMI Semantic Similarity Measure, the Horn's Index, and the Information Entropy. The incorporation of three distinct strategies for computing the most relevant items to the user-query ensures that the recommendation items are most relevant to the query without any deviations. The response time of OntoInfoG++ is mainly because of the Agents which are modeled for computing the EnAPMI, Horn's Index, and Information Entropy at a single step. However, the PCMSI is a graphical model where spectrum-based comparison is done by computing the pairwise relations each time the relevance between the elements in the graph has to be computed. Moreover, the PCMSI does not follow an inferential paradigm and the absence of specialized agents increases the response time of the model.

In case of Collaborative Filtering with Fuzzy c-means clustering, the two techniques are carried out in series one after the other and there is no parallel processing involved. Moreover, Collaborative Filtering is based on User-Item Matrix and the Ratings which result in complex computation and cold start problem which increases the processing time of the framework. CNN with K-Means clustering has the highest average response time primarily due to the reason that the CNN is based on Convolutions, and training the Neural Network and testing consume a lot of GPU cycles which increases the processing time of the queries. Moreover, the CNN when coupled with K-means clustering tends to make the application bulky increasing the overall response time. Finally, the Graph Based Ontology Techniques also consume a lot of CPU cycles mainly due to the fact that large Ontological Graph is traversed using BFS or DFS and path-based computations are done which tend to increase the overall complexity of the application, thereby increasing the overall response time.

The case in OntoInfoG++ is much different as the Ontologies and Auxiliary Knowledge are fed into the framework as Knowledge Embeddings representing the most distinct relations. The entities are populated and fed from several sources using SPARQL Endpoint which co-operate with the actual recommendation application framework but does not consume its burst time. Similarly, the agents modeled using JADE and AgentSpeak also co-operate together in parallel to compute the semantic relatedness and load Ontologies and Entities Dynamically. In OntoInfoG++ there is no concept of path-based traversal or learning from the dataset. Instead, the OntoInfoG++ infers from the knowledge embeddings which are fed as auxiliary knowledge from various concrete sources. The knowledge is already reasoned out, modeled, and accepted by a community and the OntoInfoG++ infers from the knowledge embeddings by inferencing through agents which makes OntoInfoG++ quite light weight in nature and has the least response time of 0.38 ms.

The qualitative evaluation results for the query "Chemistry in Everyday Life" for infographics recommendation is depicted for the proposed OntoInfoG++ and the baseline models which were used for comparison. Fig. 6 depicts the top 10 infographic recommendations collaged as a single image for the OntoInfoG++. It is quite clear that the individual infographics are quite relevant to the query and the essence of the query is clearly visible by yielding infographics, which are not only the best fit to the query but also yield infographics that are quite informative and create a chronology between the individual infographics yielded. Fig. 7 furnishes the infographics which are yielded by the PCMSI model. The infographics yielded focuses mainly on the term "Chemistry" in general and the essence of the query "Chemistry in Everyday Life" is not brought out by the top 10 infographics yielded by the PCMSI model as it is a graphical model which factors priorities among elements and a spectrum based comparison is followed where each time pairwise relations has to be computed. The infographics furnished by the Collaborative Filtering with Fuzzy c-means clustering sandwich model is depicted by Fig. 8 where again the term "Chemistry" is given more weightage than the query "Chemistry in Everyday Life". However, the CNN with K-means

clustering furnishes the results in Fig. 9 where the query is learnt, and the essence of the query "Chemistry in Everyday Life" is brought out in a few recommendations while most of the recommendations cater to the generic query term "Chemistry". It is very clear from the qualitative analysis that the proposed OntoInfoG++ furnishes results that are quite comprehensive to the query term and ensures that the essence of the query terms is preserved as a whole. The relevance of results in OntoInfoG++ is comprehended mainly because of the usage of three distinct measures for computing the semantic relatedness, namely the EnAPMI measure, Horn's Index, and the Information Entropy. Apart from this, the encompassment of Ontology Alignment, usage of several cognitive real-world Knowledge Sources, namely the BibSonomy, DBpedia, Wikidata, LOD Cloud, and Crowd-Sourced Ontologies, supplies auxiliary knowledge and populates entities which increase both the diversity and the relevance of results.

Fig. 6. Qualitative Results of OntoInfoG++ for the Query "Chemistry in Everyday Life".

Fig. 7. Qualitative Results of PCMSI for the Query "Chemistry in Everyday Life".

Fig. 8. Qualitative Results of Collaborative Filtering with Fuzzy c-means clustering for the Query "Chemistry in Everyday Life".

Fig. 9. Qualitative Results of CNN-K-Means Clustering for the query "Chemistry in Everyday Life".

VII. Conclusion

Infographics is a very effective tool to represent lots of information in a single picture which can be easily understood and memorized and recommending the relevant and appealing infographics for the query will make it efficient and will enable the users in learning. A semantically driven knowledge fusion approach, OntoInfoG++ has been proposed to recommend infographics based upon the user queries and user clicks. The OntoInfoG++ achieves topic enrichment by integrating entities from real-world knowledge sources like the BibSonomy, DBpedia, Wikidata, LOD Cloud, and Crowd Sourced ontologies. The semantic alignment is achieved by computing the semantic similarity between the knowledge graph formulated from the enriched topic of interest and the Ontology formulated from the titles and keywords of research papers from the dataset. The semantic similarity computation has been realized with three distinct measures namely the EnAPMI, Horns's index, and the information entropy amalgamated through an agent. The OntoInfoG++ has achieved an overall accuracy of 97.21% with a very low FDR of 0.02 with a very low response time of 0.39 ms for the experimentations conducted on RARD II dataset which makes OntoInfoG++, the best in class approach for recommendation of infographics from research papers. The high value of nDCG furnished by the proposed OntoInfoG++ indicates that OntoInfoG++ has solved the serendipity problem by improving the diversification of recommended results.

Acknowledgment

The authors thank the Ministry of Human Resources Development, India and the National Institute of Technology, Tiruchirappalli for funding this research by timely release of HTRA Research Fellowship. The authors thank God the Eternal Father and Lord Jesus Christ for providing the required knowledge and insights for carrying our this work.

References

[1] Siricharoen, Waralak, "Infographics: The New Communication Tools in Digital Age," *Proceedings of The International Conference on E-Technologies and Business on the Web, Bangkok, Thailand* , pp.169-174, 2013.

[2] Sujia Zhu, Guodao Sun, Qi Jiang, Meng Zha, Ronghua Liang, "A Survey on Automatic Infographics and Visualization Recommendations," *Visual Informatics*, vol. 4, no. 3, pp. 24-40, 2020.

[3] Wilkinson JL, Strickling K, Payne HE, Jensen KC, West JH, "Evaluation of Diet-Related Infographics on Pinterest for Use of Behavior Change Theories: A Content Analysis," *JMIR Mhealth Uhealth*, vol.4, no.4, pp.1-11, 2016.

[4] Featherstone, Robin, "Visual Research Data: an Infographics Primer," *Journal of the Canadian Health Library Association*, vol. 35, no.4, pp. 147-150, 2014.

[5] Mohd Noh, Mohd Amin & Shamsudin, Wan Nur & Amin Nudin, Anith & Narimah, Nik & Harun, Mohd, "The Use of Infographics as a Tool for Facilitating Learning," *Proceedings of The International Colloquium of Art and Design Education Research, Malaysia, Springer*, pp.559-567, 2014.

[6] Siricharoen, Waralak & Siricharoen, Nattanun, "How Infographic Should be Evaluated?," *Proceedings of The 7th International Conference on Information Technology, Amman, Jordan*, pp.558-564, 2015.

[7] Murray, Iain & Murray, A. & Wordie, Sarah & Oliver, Chris & Murray, A. & Simpson, Hamish, "Maximising the Impact of Your Work Using Infographics," *Bone and Joint Journal*, vol.6, no.11, pp.619-620, 2017.

[8] Cifçi, Taner, "Effects of Infographics on Students Achievement and Attitude towards Geography Lessons," *Journal of Education and Learning*, vol.5, no.1, pp.154-166, 2016.

[9] Nuhoglu Kibar, Pınar & Sullivan, Kevin & Akkoyunlu, Buket, "Creatıng Infographics Based on the Bridge21 Model for Team-based and Technology-mediated learning," *Journal of Information Technology*

Education: Innovations in Practice, vol.18, pp.87-111, 2019.

[10] Chen, Z., Wang, Y., Wang, Q., Wang, Y., & Qu, H, "Towards Automated Infographic Design: Deep Learning-based Auto-Extraction of Extensible Timeline," *IEEE Transactions on Visualization and Computer Graphics*, vol.26, no.1, pp.917-926, 2020.

[11] Cui, W., Zhang, X., Wang, Y., Huang, H., Chen, B., Fang, L., Zhang, H., Lou, J., & Zhang, D, "Text-to-Viz: Automatic Generation of Infographics from Proportion-Related Natural Language Statements," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no.1, pp. 906-916, 2020.

[12] Jock Mackinlay, "Automating the Design of Graphical Presentations of Relational Information," *ACM Transactions on Graphics*. vol.5, no.2, pp.110–141, 1986.

[13] Deepak, Gerard, and A. Santhanavijayan, "OntoBestFit: A Best-Fit Occurrence Estimation strategy for RDF driven faceted semantic search," *Computer Communications, Elsevier*, vol.160, pp.284-298, 2020.

[14] Middleton S.E., De Roure D, "Shadbolt N.R. Ontology-based Recommender Systems," *Staab S., Studer R. (eds) Handbook on Ontologies, International Handbooks on Information Systems. Springer, Berlin, Heidelberg*, pp.477-498, 2004.

[15] Werner, David & Cruz, Christophe & Nicolle, Christophe, "Ontology-based Recommender System of Economic," *ArXiv*, vol.1, pp.1-4, 2012.

[16] Peis, E. & Morales-del-Castillo, José & Delgado-López, J, "Semantic Recommender Systems. Analysis of the state of the topic," *Hipertext.net; Edició en anglès*, vol.6, pp.1-10, 2008.

[17] Pazahr, Ali & Samper Zapater, J. Javier & Garcia-Sanchez, Francisco & Botella, Carmen & Martínez, Rafael, "Semantically-enhanced Advertisement Recommender Systems in Social Networks," *Proceedings of the 18th International Conference on Information Integration and Web-based Applications and Services, Singapore*, pp.179-189, 2016.

[18] Bafna, Prafulla & Shirwaikar, Shailaja & Pramod, Dhanya, "Task recommender system using semantic clustering to identify the right personnel," *VINE Journal of Information and Knowledge Management Systems*, vol.49, no.4, pp.1-28, 2019.

[19] Huijsduijnen, Lies & Hoogmoed, Thom & Keulers, Geertje & Langendoen, Edmar & Langendoen, Sanne & Vos, Tim & Hogenboom, Frederik & Frasincar, Flavius & Robal, Tarmo, "Bing-CSF-IDF+: A Semantics-Driven Recommender System for News," *New Trends in Databases and Information Systems*, 2020.

[20] O. Tymchenko, S. Vasiuta, O. Khamula, O. Sosnovska and M. Dudzik, "Using the method of pairwise comparisons for the multifactor selection of infographics design alternatives," *20th International Conference on Research and Education in Mechatronics, Wels, Austria*, pp. 1-6, 2019.

[21] Deepak, Gerard, and Dheera Kasaraneni, "OntoCommerce: An Ontology Focused Semantic Framework for Personalised Product Recommendation for User Targeted E-commerce," *International Journal of Computer Aided Engineering and Technology*, vol.11, no. 4-5, pp.449-466, 2019.

[22] Deepak, G., Teja, V., & Santhanavijayan, A, "A Novel Firefly Driven Scheme for Resume Parsing and Matching Based on Entity Linking Paradigm," *Journal of Discrete Mathematical Sciences and Cryptography*, vol.23, no.1, pp.157-165, 2020.

[23] 1S. Haribabu, P. S. Sai Kumar, S. Padhy, G. Deepak, A. Santhanavijayan and N. Kumar D., "A Novel Approach for Ontology Focused Inter-Domain Personalized Search based on Semantic Set Expansion," *Fifteenth International Conference on Information Processing, Bengaluru, India, IEEE*, pp. 1-5, 2019.

[24] Deepak, Gerard, Naresh Kumar, G. VSN Sai Yashaswea Bharadwaj, and A. Santhanavijayan, "OntoQuest: An Ontological Strategy for Automatic Question Generation for e-assessment using Static and Dynamic Knowledge," *Proceedings of Fifteenth International Conference on Information Processing, Bengaluru, India, IEEE*, pp.1-6, 2019.

[25] Deepak, G., & Priyadarshini, J. S. "Personalized and Enhanced Hybridized Semantic Algorithm for Web Image Retrieval Incorporating Ontology Classification, Strategic Query Expansion, and Content-Based Analysis," *Computers & Electrical Engineering, Elsevier,* vol.72, pp.14-25,2018.

[26] Berkani, Lamia & Belkacem, Sami & Ouafi, Mounira & Guessoum, Ahmed, "Recommendation of Users in Social Networks: A Semantic and Social Based Classification Approach," *Expert Systems*, vol.38, no.2, 2020.

[27] Javed, Umair & Shaukat Dar, Kamran & Hameed, Ibrahim & Iqbal, Farhat & Mahboob Alam, Talha & Luo, Suhuai, "A Review of Content-Based and Context-Based Recommendation Systems," *International Journal of Emerging Technologies in Learning (iJET)*. vol.16, no.3, pp.274-306, 2021.

[28] Houari, N. S., & Taghezout, N, "An Effective Tool for the Experts' Recommendation Based on PROMETHEE II and Negotiation: Application to the Industrial Maintenance," *International Journal of Interactive Multimedia and Artificial Intelligence*, no.6, pp. 67-77, 2021.

[29] Bobadilla, J., Lara-Cabrera, R., González-Prieto, Á., & Ortega, F, "DeepFair: Deep Learning for Improving Fairness in Recommender Systems," *International Journal of Interactive Multimedia and Artificial Intelligence*, no.6, pp.86-94, 2021.

[30] Pushpa, C. N., Gerard Deepak, J. Thriveni, and K. R. Venugopal, "Onto Collab: Strategic Review Oriented Collaborative Knowledge Modeling using Ontologies," *Proceedings of The Seventh International Conference on Advanced Computing, Chennai, India, IEEE*, pp. 1-7, 2015.

Gerard Deepak

Gerard Deepak holds a Masters in Engineering degree in Computer Science and Engineering from UVCE, Bangalore University. He is a University Level First Rank Holder for his masters and has qualified the KSET examination. Currently he is pursuing ahis PhD from National Institute of Technology, Tiruchirappalli. He has a h-index of 17 and has received the Budding researcher award from NITT twice successively. He also has received the best paper award 7 times in several international conferences in India and Abroad. He has 50 articles to his credit to date in both Journals and Conferences of repute. His areas of interests include Semantic Web, eXplainable AI, Semantic Web Mining, and Ontology Engineering.

V. Adithya

V. Adithya is a student pursuing his final year Undergraduate degree in Computer Science and Engineering. He is an AI and Machine Learning enthusiast who has also published and presented papers in many international conferences and has also received one best paper award. His other areas of interest include Software Engineering, Deep learning and eXplainable AI.

Dr. Santhanavijayan A

Dr. Santhanavijayan A holds a PhD from National Institute of Technology Tiruchirappalli and a Masters in Computer Engineering from Anna University, Chennai. He has received the Best Performing Faculty Award under his cadre from National Institute of Technology Tiruchirappalli. His research interests are Semantic Web, Natural Language Processing, Deep Learning, and Data Science.