



June 26, 2018

Subject: Letter of Support for Nomination of Kalpa Gunaratna's Dissertation "Semantics-based Summarization of Entities in Knowledge Graphs" for SWSA Distinguished Dissertation Award

Dear Members of the Award Committee:

I am delighted to nominate Kalpa Gunaratna's dissertation for the SWSA Distinguished Dissertation Award as the co-chair of his dissertation committee and co-advisor (primary) of his dissertation research. I confirm that Kalpa successfully defended his dissertation on April 19, 2017, and was awarded the degree of doctor of philosophy (Ph.D.) on April 29, 2017. Kalpa's dissertation is framed within the Semantic Web area and is about how to capture the essence of large entity descriptions in knowledge graphs for efficient human/machine consumption. The interesting findings of his dissertation research are published in top-tier highly competitive conferences such as AAAI, ESWC, and IJCAI. I briefly point out some highlights of his dissertation research below.

Building knowledge graphs and making use of them in applications to enable a greater level of intelligence has gained significant attention in the recent past. Knowledge graphs and knowledge-based support for semantic computing impact our now considered normal tasks like Web search and interactions with personal assistants. What makes this dissertation particularly distinguished is that it makes prominent theoretical contributions and aligns them with practical applications. It deals with issues that I have held dear for some time and had also worked on nearly 20 years ago when I had founded/led Taalee/Voquette/Semagix that developed a semantic application development platform used for a variety of applications ranging from search to business intelligence (we had developed the first large knowledge graph based commercial semantic search engine and received the first patent in Semantic Web area: <http://bit.ly/15yrSemS>). In those systems, the creation of snapshots (i.e., summaries) was guided through some defined rules and in Kalpa's work, he proposed approaches that captured the underlying semantics of the knowledge graph data to construct high-quality summaries. Further, Kalpa's work on entity summarization came into the research community when it was in its early stages of development and could be considered a few of the early attempts along the direction. Not restricted to traditional publication pipelines, he initiated a workshop series (SUMPRE) on entity summarization at ESWC conference (2015 and 2016) to support and build the community around these important research directions. The workshop series gained attention and was among the top workshops in terms of community engagement and paper submissions for the respective ESWC conference.

Kalpa's dissertation covered three main topics for entity summarization extending the state-of-the-art results. He addressed and brought in the successful use of diversity in creating single



entity summaries and took the challenge of creating multi-entity summarization, which is proven computationally np-hard.

Kalpa's first contribution is the FACeted Entity Summarization (FACES) approach that makes use of conceptual grouping of facts to create the summaries. Existing approach for entity summarization at the time focused entirely on ranking mechanisms whereas FACES supports the automatic identification of different groups conforming to conceptual (i.e., latent) similarity of facts (both property and value) and then select facts to make summaries both concise and comprehensive. The approach implements an incremental conceptual clustering method that could further serve as an independent incremental grouping mechanism for RDF triples. The evaluation results for FACES significantly outperformed other existing state-of-the-art approaches and the findings are published in AAAI2015.

Kalpa's second contribution is to use both object and datatype property triples to summarize single entity descriptions in the FACES framework. The conceptual grouping method requires a semantic understanding of property values. Even though datatype properties cover a significant proportion of the encoded knowledge in knowledge graphs like DBpedia, their wide-scale use is hindered because the property values have no semantic meaning (no semantic type information) associated with them. FACES approach is first extended to compute appropriate semantic types for literals in triples and then use the type-computed literals to compute entity summaries, increasing the fact selection space. This is the first effort in the literature to compute a semantic type for literals in order to make them more useful in applications. The results are published in ESWC2016.

From a scientific point of view, the third contribution of the dissertation is more computationally challenging as it looks to summarize multiple entities at once. Trying to compute summaries for a collection of entities at once is challenging because a selection of a fact for an entity can effectively invalidate facts already selected for other entities. In this proposed work Relatedness based Multi-Entity Summarization (REMES), all the facts of a given entity collection are processed to maximize intra-entity summary diversity and importance and inter-entity summary relatedness. Since this is a multiple objective optimization problem, a variation of knapsack problem-solving technique (np-hard) is adapted. The results are published in IJCAI2017. When I compare with the semantic search and browsing for the companies I founded/managed, a multi-entity processing approach like REMES presents a significant advance. In my view, this is the kind of progress that pushes the boundaries in a given field and makes a Ph.D. dissertation noteworthy.

With the nature of top-tier publications associated with this dissertation, Kalpa has taken Semantic Web contributions to a broad audience in Artificial Intelligence in addition to the core Semantic Web community. Further, Kalpa's dissertation research has also influenced other events like inventing patents and impact in industrial/external research labs (Insight Center, NLM, and IBM), organizing workshops to promote research along the dissertation (ESWC 2015 and 2016, and extending it to a broader community in CIKM 2018), and serving in top-tier conference PC committees (WWW and ISWC) while being a graduate student. I believe Kalpa's



achievements with his dissertation research are quite impressive and surely be most Ph.D. students' dreams.

In summary, I am extremely happy with the comprehensive contributions that this dissertation work has made to the Semantic Web research on a number of aspects – breadth, depth, quality, novelty, timeliness, and significance. The quality, novelty, significance, and timeliness of the research in the dissertation is justified by the top-tier highly competitive conference publications and continued adaptation of them as state-of-the-art baselines. Further, the dissertation discusses both single and multi-entity summarization approaches covering the breadth and the algorithms implemented are theoretically well-grounded signifying the depth of the work. A dissertation demonstrating aforementioned qualities deserves recognition and I wish Kalpa's dissertation gets its due credits. Consequently, I enthusiastically and strongly recommend his dissertation for the SWSA Distinguished Dissertation Award.

Please feel free to contact me if you should require any further information.

Sincerely,

Amit Sheth

LexisNexis Ohio Eminent Scholar/Professor, Computer Science & Engineering  
Executive Director, Ohio Center of Excellence in Knowledge-enabled Computing (Kno.e.sis)  
Email: [amit.sheth@wright.edu](mailto:amit.sheth@wright.edu) <http://knoesis.org/amit>

Prof. Amit Sheth is an Educator, Researcher and Entrepreneur. He is an Ohio Eminent Scholar, executive director of Kno.e.sis- the Ohio Center of Excellence in Knowledge-enabled Computing & BioHealth Innovations, an IEEE Fellow (2006), and an AAAI Fellow (2018). He is a well cited computer scientist (h-index = 100), especially in Semantic (<https://aminer.org/search?t=b&q=Semantic%20Web>), Distributed Workflows and WWW. He has founded three companies by licensing his university research outcomes, including the first Semantic Web company in 1999 that pioneered technology similar to what is found today in Google Semantic Search and Knowledge Graph. Several commercial products and deployed systems have resulted from his research. He is particularly proud about his students' exceptionally success in academia, industry research labs and as entrepreneurs. His current interests encompass smart data and personalized AI (extracting insights and actionable information from IoT, sensor, health big data using knowledge-enhanced ML and NLP techniques), augmented personalized health and computing for human experience.