Formation of Information Need as a Questioning: A Conceptual Representation for Semantic Web based QA System

Yongju Lee[1], Sungkwon Yang[1], Sueun Jang[1], Hong-Gee Kim[1]

[1] Biomedical Knowledge Engineering Laboratory,
Seoul National University, Seoul, Korea

{yongju_lee, sungkwon.yang, jchr119, hgkim}@snu.ac.kr

Abstract. Questioning is a formation of information need where an asker is aware of an anomalous state of knowledge. Such awareness can only emerge when a user encounters to certain contextual circumstance. Either user's surroundings in reality (location, back-ground music, etc.) or web (news article, blog, video, social network, etc.) can be a certain contextual circumstance. By examining a simple question "What triggers question?", we concluded that a certain contextual circumstance drives a question where an asker is aware of anomalous state of knowledge. We also identified a direct and indirect channels expressing contrastive types unfolding from a potential asker's contextual situation input to a formation of information need. Such channels could potentially serve as a medium of personalized model for predicting, adjusting, and retrieving user's need of information. Further application of "ambient" feature that readily captures a certain contextual circumstance, subjectively questions to QA system based on user's pattern and answer retrieval without a user's (potential asker's) query will envisage a step closer to AI-complete QA system.

Keywords: Semantic Web Service, Question Answering, Human-Centered Computing

1 Introduction

Conventional search engines provide a list of relevant documents, while the goal of question answering is to provide the information instantaneously to the user without crawling all documents. An exponential growth of data has led QA systems struggling with capacity, heterogeneity and authenticity of the basal knowledge [1].

The Question Answering over Linked Data (QALD)'s main challenge as follows. Given one or several RDF dataset(s) as well as additional knowledge sources and natural language questions (factoid questions that can be answered with simple facts expressed in short answers) or keywords, return the correct answers or a SPARQL query

that retrieves these answers. In 2017, QALD-7 challenges included following four including subjects: (1) Multilingual question answering over DBpedia by retrieve answers from an RDF data by given an information need expressed in a variety of natural languages. (2) Hybrid question answering by finding, processing, and combining information in multiple repositories. (3) Large-scale question answering and (4) Question answering over Wikidata.

All issues covered in above challenges represents a partial phase of human-centered question answering which is interactive and result in dialogues to get the most valuable knowledge to each individuals with a complex questions translated into a set of queries. Clark et al. [2] stated a three important direction for question answering. (1) Extending relation between question and corpus (i.e. Recognizing Textual Entailment), (2) Extending the range of answerable types of questions, (3) the answers for identical questions which is most likely the correct answer for the user.

In this paper, we ask a simple question, “What triggers question and how can we utilize them for intelligent question answering system?”. In section 2, a literature review regarding a formation of information need through motivations is followed. In section 3, as an answer to our primary question “What triggers question and how can we utilize them?”, a certain contextual circumstance is described in detail with some rationale. In section 4, we bring our concept to conventional framework for human-centered question answering application. In section 5, we map out our concept for implementation.

2 Formation of Information Need Through Motivations

This human-centered question answering generally initiates from user’s information need which can also be construe as a cause of information seeking or “a requirement that drives people into information seeking”. According to Belkin [3] and Taylor [4], when an asker is aware of his or her anomalous state of knowledge in a certain situation a formation of an information need(question) follows.

Choi and Shah [5] reported that while the question answering has impacted information-seeking behaviors, people's motive behind a question hasn't been extensively covered by the research community. With the assumption of which conceptualization of different contexts and situations of information needs can be rendered through cognizance of users' motivation, Choi and Shah [5] investigated a user's motivations by adopting typologies of motivations for media use introduced by Katz et al. [6].

Five different motivational variables were developed including cognitive, affective, personal integrative, social integrative and tension free needs that drive people to ask a question in web environment. Depending on asker's contexts and situations, the motives were varied followed by author's survey with "cognitive needs" identified as the most significant motivational variable.

However, a motive or motives themselves cannot be viewed as a sole driver regarding a formation of information need (question) since an asker's awareness of anomalous state of knowledge comes first. Moreover, an anomalous state of knowledge won't derive without a certain contextual circumstance which is confronted by an asker or user. In other words, our question starts with "What triggers a question?".

A certain contextual circumstance could be an environment where someone asking a weather or a specific destination. It can be also being a news article about the Nobel prize in literature 2017 for someone asking "Who is Kazuo Ishiguro?" or "What are the titles from the laureate?". Also our question leads to if the question answering system discerns a certain contextual circumstance encountered by potential asker, how can we utilize such ambient feature in semantic web services.

3 Certain Contextual Circumstance

Questioning is a formation of information need where an asker is aware of an anomalous state of knowledge. Such awareness can only emerge when a user encounters to certain contextual circumstance. Either user's surroundings in reality (location, background music, etc.) or web (news article, blog, video, social network, etc.) can be a certain contextual circumstance.

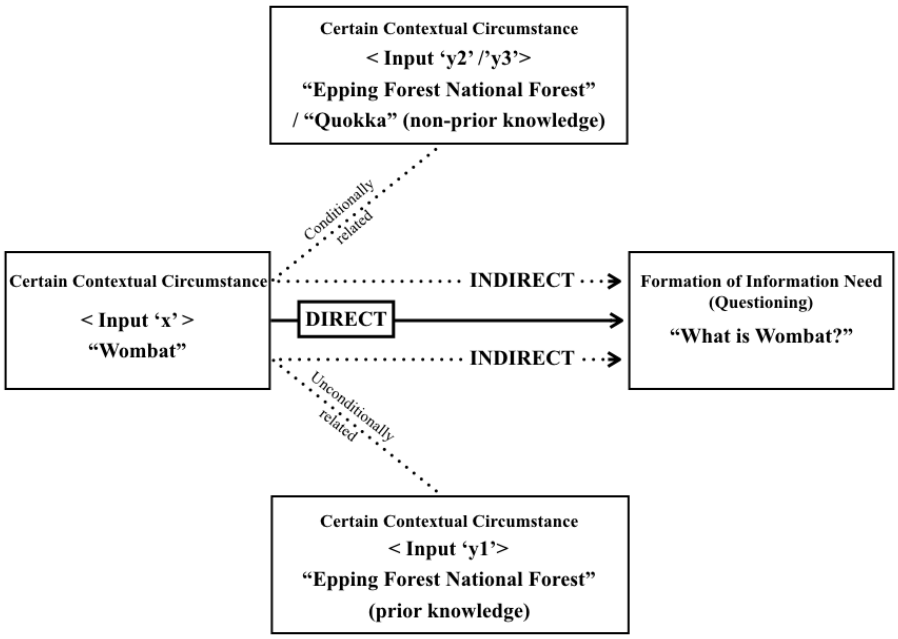

Fig. 1. A scenario when user (potential asker) confronts with a different types of certain contextual circumstances.

Fig.1 represents a scenario when user (potential asker) confronts with a different types of certain contextual circumstances. We will denote instance of certain contextual circumstance as an input 'x' which represents certain entity or relation. The direct and

indirect process of certain contextual circumstance to questioning (formation of information need) has shown in Fig.1. All questions initiate with input 'x' which is a certain contextual circumstance. We assume unless the potential asker "A" questions for input 'x' at least once, we count input 'x' as a non-prior knowledge.

The first edge called "Direct" represents input 'x' directly mediates a formation of information need(questioning). As an example, the potential asker "A" encounters with an entity "Wombat"(input 'x') as a text while reading a blog about Australian animals. To "A" a "Wombat" is non-prior knowledge. Unless the potential asker "A" disregards the text "Wombat", he or she may ask "What is a Wombat?" through motivation (i.e. Cognitive needs, tension free needs, etc.).

The second edge is "Indirect" which represents input 'x' indirectly mediates a formation of information need(questioning) via another input 'y'. We assume the potential asker "A" have already encountered with input 'x' and already questioned and retrieved the answer through system. This is because if input 'x' is a non-prior knowledge, the potential asker "A" cannot somehow relate 'x' to 'y'. For instance, the potential asker "A" encounters with an entity "Epping Forest National Park" as a text while browsing a map, which "A" is already encountered through the online conversation with the user "B" in above example. The potential asker "A" may ask "What is Wombat?" rather than "What is Epping Forest National Park?". If the potential asker "A" questioned "Epping Forest National Park"('y1') through system before (prior knowledge), then we assume that between "Wombat" and "Epping Forest National Park" is unconditionally related, specifically to "A". If "A" disregarded "Epping Forest National Park"('y2') through conversation (non-prior knowledge), yet he or she asks "What is Wombat?" via an entity "Epping Forest National Park", we assume that between "Wombat" and "Epping Forest National Park" is conditionally related, specifically to "A". This also includes a case when a whole new input comes up. Take an example which a whole new input 'y' is "Quokka"('y3'). The potential asker "A" encounters "Quokka" while reading a blog about Australian animals. "A" didn't asked about "Quokka", yet somehow asks "What is Wombat" rather than asking about "Quokka". The difference between 'y2' and 'y3' is that the 'y2' was actually encountered yet didn't progress to questioning, but for 'y3' it was introduced to "A" for the first time. Whatever the difference between two instances ('y2', 'y3') exists, we equally designate them as conditionally related with 'x'. Whether through direct channel or indirect channels a formation of information need is primarily on the basis of a certain contextual circumstances.

4 Commercial QA and Certain Contextual Circumstance

In previous section we proposed a certain contextual circumstance triggers a formation of information need either through direct or indirect channels. Since almost every major conventional question answering system is first initialized with arbitrary question, a concept "formation of information need" is a baffling subject to discuss.

Take Google Assistant as an instance. Like "Siri" from Apple, "Google Assistant", a virtual personal assistant application enforced by Google Knowledge Graph was intro-

duced at its developer conference in 2016. The Google Knowledge Graph was developed 4 years earlier as a means of semantic search enhancement for Google search engine. Erhlinger and Wöβ [7] stated "A knowledge graph acquires and integrates information into an ontology and applies a reasoner to derive new knowledge." With two-way conversations, users using Google Assistant can query by natural language and the answers can be retrieved either from the knowledge graph or search engine.

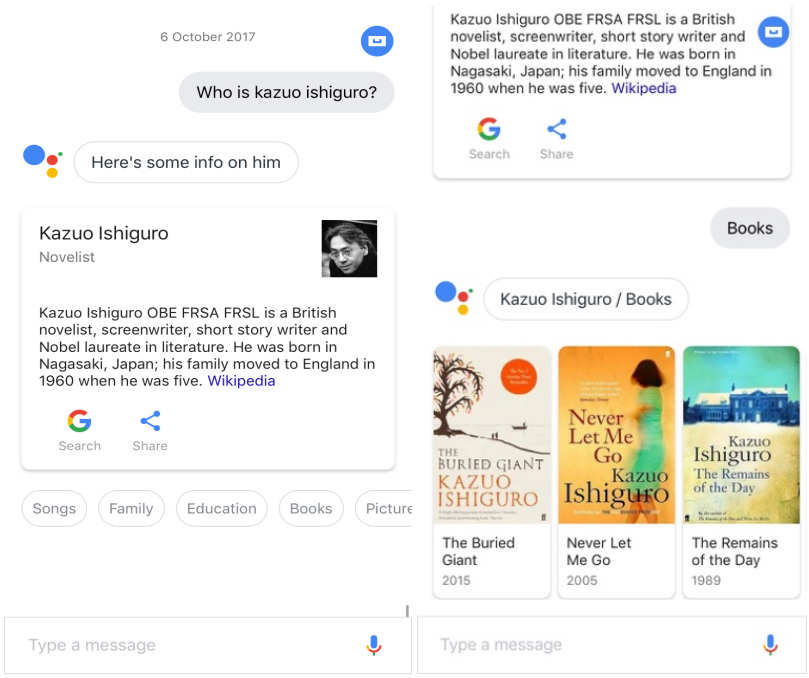

Fig. 2. The Google Assistant retrieved answer based on query "Who is Kazuo Ishiguro?"(Left). One of the suggestions has been selected from left image, in this case "Books" (Right).

As an example shown in Fig. 2., the Google Assistant retrieved answer based on query "Who is Kazuo Ishiguro?". It also displayed a suggested questions related to entity(attributes) including Songs, Family, Education, Books and Pictures. Right side of the Fig. 2. shows when one of the suggestions has been selected, in this case "Books". It requires a query with target object, in order to get a snippets of suggestions which can be eluded as user's potential question.

Basically, these AI-driven personal assistant apps envisaged an AI-complete a step closer by exploiting machine learning and natural language processing. Nevertheless, in commercial perspective Verto Analytics reported [8] that the personal assistant apps sector has seen modest growth, yet also with decrease of user engagement and stagnation in major platform such as Siri from Apple. Between May 2016 and May 2017 Siri lost 7.3 million monthly users, which is dropped by nearly 15% of its total. During May 2017, Verto also reports that with just 44% of all smart smartphones in the U.S. had a

personal assistant app that was used at least once and compared to 369 million hours of total time spent on search websites through smartphones, only 14.6 million hours were spent by U.S. adults over 18. While the less appealing consumer behavior cannot directly assess the question answering system research and development, it may reflect the deficits in the pursuit of human-centered essence.

Recently, Google announced the new feature on their new hardware "driven by that move is the option to have the device listen out for music and identify the song currently playing in an environment" [9]. Further application of "ambient" feature that readily captures a certain contextual circumstance, subjectively questions to QA system based on user's pattern and answer retrieval without a user's (potential asker's) query will envisage a step closer to AI-complete QA system.

5 Formation of Information Need Through Motivations

Our next question is what is required to discern a certain contextual circumstance encountered by a potential asker in question answering system, and how can we utilize such ambient feature in semantic web services.

By its nature an implementation of a certain contextual circumstance conflicts with two major barriers across the web. (1) A potential asker's potential certain contextual circumstance is predominantly originating from unstructured with exponential growth.

User's medium is majorly consisting a text, image, and medium. Without a "large networks of entities, their semantic types, properties, and relationships between entities", an 'input' aware question answering is an impracticable task [10]. An ideal __modus operandi__ would be generating a rdf model what user is browsing, then mapping to central interlinking hub such as DBpedia. (2) An apparatus for learning user specific certain contextual circumstance and pairing with questions.

Human-centered question answering demands consistent pattern learning which captures human behavior in sequential order. It also has to concern with efficacy for mass calculation as a role of mediator between an input and source for each bilateral communication.

6 Future Work and Conclusion

By examining a simple question "What triggers question?", we concluded a certain contextual circumstance drives a question where an asker is aware of anomalous state of knowledge. We also identified a direct and indirect channels expressing contrastive types unfolding from a potential asker's contextual situation input to a formation of information need. Such channels could potentially serve as a medium of personalized model for predicting, adjusting, and retrieving user's need of information. Currently, we are drafting theoretical framework for cognitive research experimentation and algebraic representation based on basic category theory.

As Clark et al. [2] mentioned, it is crucial that question answering system to tailor the answer for each and every user. To our best knowledge, research concerning a certain contextual circumstance based questioning, nor pre-question behavior consideration

hasn't been made. We believe our concept is empiric enough to be an annex to the framework of human-centered question answering.

Acknowledgments. This work was supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIP) (No. 2013-0-00109, WiseKB: Big data based self-evolving knowledge base and reasoning platform).

References

1. Jurafsky, D., Martin, J.H.: Speech and language processing : an introduction to natural language processing, computational linguistics, and speech recognition. Pearson Prentice Hall, Upper Saddle River, NJ, USA (2009)
2. Clark, A., Fox, C., Lappin, S.: The handbook of computational linguistics and natural language processing. Wiley-Blackwell, Chichester, West Sussex ; Malden, MA, USA (2010)
3. Belkin, N.J.: Anomalous states of knowledge as a basis for information retrieval. Canadian Journal of Information Science. Toronto, ON, Canada (1980)
4. Taylor, R.S.: Question-negotiation and information seeking in libraries. College and Research Libraries, Chicago, IL, USA (1968)
5. Choi, E., Shah, C.: User motivations for asking questions in online Q& A services. Journal of the Association for Information Science and Technology. Medford, MA, USA (2015)
6. Katz, E., Guervitch, M., & Haas, H.: On the use of the mass media for important things. American Sociological Review, Notre Dame, IN, USA (1973)
7. Erhlinger, L., Wöβ, W.: Towards a Definition of Knowledge Graphs. SeMANTICS 2016. Leipzig, Germany (2016)
8. Verkasalo. H.: Rise of the Machines: How AI-Driven Personal Assistant Apps Are Shaping Digital Consumer Habits. Verto Analytics. New York City, NY, USA (2017)
9. Gibbs, S.: Google's Pixel 2 and Pixel 2 XL: an AI-infused challenge to the iPhone. The Guardian. Guardian News and Media Limited, The Guardian, Kings Place, 90 York Way London N1 9GU, United Kingdom (2017)
10. Kroetsch. M., Weikum. G.: Special Issue on Knowledge Graphs. Journal of Web Semantics. Karlsruhe, Germany (2016)