Too Many Links in the Horizon; What is Next? Linked Views and Linked History.

Erietta Liarou and Stratos Idreos

CWI, Amsterdam, The Netherlands

Abstract. The trend for more online linked data becomes stronger. Foreseeing a future where "everything" will be online and linked, we ask the critical question; what is next? We envision that managing, querying and storying large amounts of links and data is far from yet another query processing task. We highlight two distinct and promising research directions towards managing and making sense of linked data. We introduce *linked views* to help focusing on specific link and data instances and *linked history* to help observe how links and data change over time.

1 Introduction

"Knowledge is Power"; this famous phrase was stated originally by the Persian poet Ferdowsi almost two thousand years ago. The World Wide Web is nowadays a characteristic example that reflects this quote; it has evolved to become a global pool of knowledge available to everyone with access to basic computing resources. In turn, this continuous and easy access to knowledge has transformed the way our society works, i.e., the way we do business, education, entertainment, etc.

The Future Web. The vision of the Web of the future [4, 6] expands even more to map the real life paradigm to the Web, i.e., everyone and everything is related to and affects each other [2]. Then, the need for defining relationships, i.e., links, between data becomes essential. By making available and by linking data, we will be able to extract more sophisticated knowledge, exploiting the way the diverse data is related to each other. The notion of linked data can be applied everywhere; some characteristic datasets include the DBpedia collection which is the linked data version of Wikipedia, the DBLP bibliography that provides bibliographic information on scientific publications, the FOAF vocabulary which is used to describe people and their relations, the LinkedGeoData that provides annotated geographical data, and many others that reflect the links of various aspects of our lives. The vision of the Web of Data [6] does not stop there; it aims to interlink all these diverse domains towards a *Huge Web Database*.

Too Many Links in the Horizon. There are clear signs that we are heading towards such a future where "everything" is online and linked, ranging from scientific data, government data, to personal data and media. The advent of social networks is a characteristic example where we see that people are ready to engage, create, tag and link data, given the proper technology. Crowdsourcing is a more recent technological and research path where again we see that people

are ready to create data online and participate even in solving complex problems given the proper incentive and technology. In the science domain it is indeed harder for people to "give" their data, but there are recent examples towards changing this paradigm. For example, with the SkyServer project [7] data and even query logs from astronomy are accessible publicly online nowadays. At the same time, there is a growing trend in making government data publicly available and online such as tax information, expenses of government institutions, etc.

So what is Next? The trends described in the previous paragraph are only expected to grow and the question we ask in this short paper is what will happen if the "everything connected" future we target for arrives. How are we going to deal with so much data and with so many links? Naturally, this is not a question to be answered in one paper or even by a single community as there are so many problems and challenges involved ranging from hardware issues, software and engineering issues all the way to social and economical issues. The particular question we are interested in is how are we going to best exploit and make sense of the huge collection of links and data?

Linked Views and History. We propose two distinct and novel concepts to enrich the linked data vision. The first one is *linked views* to enable quick and focused inspection of part of the linked information such that it complies with needs and requirements of both publishers and consumers. Second, we propose *linked history* as a means towards observing changes in linked information over time. Linked views and history will significantly increase the ability to turn knowledge into power in the Web of future.

Our ideas are inspired by the strong and ongoing research in several and broad areas in the information systems community. In particular, pioneering ideas such as the adaptive web [3] for personalization, web archiving [1], as well as recent trends for large data processing in the database community [5] have influenced us. Our proposal goes several steps further in envisioning ways specifically for the future web of linked data and the expected deluge of online links. In the rest of the paper, we will briefly touch upon further discussing the definition of linked views and linked history as well as their expected usefulness and the research challenges they bring.

2 Linked Views

We propose linked views as a new challenge and goal for the linked data future. Linked views will provide the means for isolating a part of the linked data and the web, given properties and requirements set by the people or applications that consume data or even by the data publishers.

A view contains only part of the web. In addition, it may contain data that have been inferred from the web. Linked views share the same technical interfaces as normal linked data, i.e., they should use URIs to identify data, they should be referenced via HTTP and return data in a standard format, while also linking to other related objects on the web. In other words, from a user point of view, accessing a linked view should be identical to accessing linked raw data.

Dynamic and personalized linked views can enable more sophisticated and focused search as well as they can enable new applications and even boost search performance. Knowledge hides better among big piles of data; the more data, the harder it becomes to infer knowledge. By isolating and even refining part of the data, it allows for better access especially for user communities or application groups with similar interests and needs. For example, a view may be created by simply choosing part of the data and their links. Alternatively a view may be an interpretation of part of the web, i.e., by inducing new links or removing others. In addition, a view may contain data that are calculated out of the raw linked data. One of the first visions for the web was for scientists to share documents; now it is data. Imagine scientists creating views targeting a very specific problem; these would contain relevant data worldwide along with the scientist's observations inferred as new data. Then, other scientists may later jump in this view to continue refine and expand the search or simply browse various views to get inspiration by other scientist's explorations.

There are several semantic and technical challenges. If one considers a single data set in a single web server, then the solutions are "trivial" in the sense that we can reuse existing and established technology and ideas, e.g., from view technology in database systems. Once we consider, however, the massively distributed web, the highly linked data sets and the dynamic, often unpredictable, nature of links this shapes to be a completely new challenge.

Some of the research questions are the following. How do we store views? How do we keep track of views? Can anyone create and drop a view? How do we query views? Naturally full materialization is impractical and thus careful relinking is required to establish that a data entry and its links belong to a given view. In this case, keeping track of the current view identity is important when browsing and following links in a given view, i.e., a user should automatically see only the links and data that belong in his/her active view only. Anyone should be able to create and drop their own views but naturally large institutions and organizations may have more resources in supporting the storage and search over their views while a plain user may have to rely purely on non materialized views. Alternatively large organizations may provide view support, e.g., for small businesses, very much as it happens today with cloud infrastructures.

3 Linked History

Linked views can show a glimpse of the web. Linked history goes a step further to show the way the web, i.e., the data and their links, changes over time. Observing change implies identifying several properties, e.g., how the values of a specific data set change over time, who changed those values and when, where an object is hosted or replicated over time, how data sets grow or shrink, how links develop over time on a specific data entry or data set, etc. Being able to access this kind of information can significantly aid in interpreting and exploiting the web.

One natural use of linked history would be to see how the web or part of the web was at given point in time. This is the simplest case though. The linked

history vision is to provide even more flexibility. For example, if one has the option to see who, when and how updates links on a linked data set, then they may interpret this data set differently or use it differently. Links that have been updated recently and by a large number of trusted peers may be of more importance to a particular user or application than links that have been updated in the past. Linked history should be accessible via HTTP much like a linked view. Then users can search through time across multiple of these views; imagine a web service similar to today's web search engines.

Overall there is a plethora of open problems that range from pure engineering to novel research issues. Linked history immediately implies more computation and storage needs than a typical linked view. In order to be able to access history information the delta changes should be archived and then there should be enough compute power to service requests. Requests may be local or they may require combination of linked data, views and history from multiple points of the web. This immediately creates a challenging query processing and network problem. Indexing, caching, ranking in linked history becomes a challenge due to the diverse dimensions and the depth of the history. Adaptive solutions can be of significant importance here where the history is built and maintained adaptively as more users request access to specific history time-lines.

4 Summary

We are soon embarking on a massive web of linked data. We propose linked views and linked history as two research paths and functionalities that will help making sense of the numerous linked data sets and the way they evolve. With the web being dynamic, diverse and constantly growing it is critical to allow for focusing on specific web parts with linked views as well as identifying and understanding changes with linked history. We expect linked views and history to play a critical role in the future web as well as to trigger several research challenges towards identifying how to design and implement this kind of functionality.

References

1. A. Anand, S. J. Bedathur, K. Berberich, R. Schenkel, and C. Tryfonopoulos. Everlast: a distributed architecture for preserving the web. In *JCDL*, 2009.
2. N. Christakis and J. Fowler. *Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives.* Little, Brown and Company, 2009.
3. P. Dolog, N. Henze, W. Nejdl, and M. Sintek. Towards the adaptive semantic web. In *1st Workshop on Principles and Practice of Semantic Web Reasoning*, 2003.
4. J. Hendler and T. Berners-Lee. From the semantic web to social machines: A research challenge for ai on the world wide web. *Artif. Intell.*, 174(2):156–161, 2010.
5. M. Kersten, S. Idreos, S. Manegold, and E. Liarou. The researcher's guide to the data deluge: Querying a scientific database in just a few seconds. In *PVLDB Challenges and Visions*, 2011.
6. LinkedData. http://linkeddata.org/.
7. Sloan Digital Sky Survey (SkyServer). http://cas.sdss.org/.