Kleenks: collaborative links in the Web of Data

Razvan Dinu[1] and **Andrei-Adnan Ismail**[2] and **Tiberiu Stratulat**[3] and **Jacques Ferber**[4]

Abstract. Linked Data is an initiative towards publishing and connecting structured data on the Web, creating what is called the Web of Data. This allows the creation of new types of applications, such as mash-ups and semantic searches, which make use of and integrate data coming from multiple online repositories. However, the large amount of content produced by blogs, wikis and social networks, which have become de facto standards for publishing content in Web 2.0, suggests that the growth of Web of Data could also be supported by adding a social, unstructured, collaborative dimension to it. In this paper we introduce "kleenks", which are collaborative links that combine structured and unstructured data by allowing users to add unstructured content to links, in addition to the RDF predicate. The quality and importance of such links can be evaluated by the community through classical mechanisms such as ratings and comments. This approach stimulates the emergence of more complex and abstract relations between entities, allowing people to take part in the Linked Data process and actively contribute to the growth of the Web of Data. We discuss how kleenks can be modeled on top of RDF and RDFS, making them easy to implement, and identify the main challenges to be addressed by a platform implementing the kleenks model. Finally, we introduce an online platform that successfully applies kleenks in the research domain by allowing researchers to create kleenks between articles, books and any other type of online media.

1 Introduction

1.1 Context

"This is what Linked Data is all about: it's about people doing their bit to produce a little bit, and it's all connecting.,, - Tim Berners-Lee, TED 2009.

Linked data is a movement trying to expose the world data in a structured format and to link it all together in meaningful ways. This concept has been gaining traction as more and more organizations are starting to expose their data in a structured, computer-understandable format, besides the traditional website. Until recent, the habit was this: if an organization owned some data and it wanted to expose it to the public, it created a website allowing users to explore it. However, it soon became obvious that this was not enough; humans were not the only ones interested in working with this data, sometimes even computers or software agents delegated by humans should be able to manipulate it. In the dawn of this era, the web crawling and screen scraping concepts appeared. Programs that contained specific parsing code for extracting knowledge out of raw HTML emerged, and they were named crawlers or scrapers. Due to the technical difficulties of doing NLP (Natural Language Processing), these programs would use the underlying regularities in the HTML structure to parse the structured data. Soon, a war broke out between content owners who did not want to expose their data to machines and humans aiding the machines in extracting the data by continuously adapting the parsers to changes in HTML structure and to security additions aimed at differentiating humans from crawlers.

In the center of this war comes Berners-Lee's concept of Linked Data. Linked data is no longer data exposed by machines for machines, but it is data exposed by humans for their fellow machines. The Linked Open Data (LOD) project is leading this movement of encouraging people to expose their data to machines in a meaningful format. Most of the projects put forward by LOD are projects in which humans are in the center of the process of generating linked data. Big names in the internet industry such as Facebook agree with this vision, as confirmed by the launch of Facebook Open Graph v2 initiative at the F8 conference in 2011[5]. This announcement is about making the transition from the singular "like" action that users could perform on the social platform to a multitude of actions, even custom ones definable by application developers: read, watch, listen, cook, try out, and so on. Given the large amount of data continuously generated by users on their social networks, this step will finally expose all that data internally as structured data.

The DBpedia project is a community effort to extract structured information from Wikipedia and to make this information accessible on the Web [2]. This is actually an attempt at automatically parsing the Wikipedia infoboxes (the boxes with highlighted information usually in the right part of the page) into RDF triples. This database of triples is maintained in permanent synchronization with Wikipedia by using a subscription to the live modifications feed. In this case, people still play a central role in the generation of data, but their actions have the creation of linked data only as an indirect consequence.

Freebase [3] is an initiative supported by Google to apply the wiki concept to the world's knowledge. A user interface and a RESTful API are provided to users in order to be able to collaboratively edit a database of triples spanning more than 125 million triples, linked by over 4000 types of links, from domains as diverse as science, arts & entertainment, sports or time and space. One of the main focuses of this project is the user-created ontology, which is constantly evolving and cannot possibly be a set of fixed existing ontologies, no matter how complete they are, due to user friendliness reasons. There is actually one interesting conclusion arising from this fact: using a distributed workforce to crowdsource structured data requires a compromise between data precision and data quantity.

1.2 Problem statement

As we have seen in the previous section, there is a growing need for exposing the world's data in a structured format, as confirmed

[1] University of Montpellier 2, LIRMM
[2] Politehnica University of Bucharest
[3] University of Montpellier 2, LIRMM
[4] University of Montpellier 2, LIRMM

[5] https://f8.facebook.com/

by industry giants and academia alike. There are a number of efforts trying to bridge this gap. Only to name a few:

- crowd-sourcing structured data from users; examples are Freebase and OpenStreetMap
- crowd-sourcing unstructured data from users, in a nearly-structured format; examples are Wikipedia and Facebook before the launching of Facebook Open Graph v2
- crawling / scraping data from unstructured data; this includes shopping, travel and news aggregators, just to give a few examples
- extracting entities and links from unstructured text using NLP (Natural Langauge Processing); one elocvent example of this is OpenCalais[6]

However, current efforts for structuring the web's data are mostly concentrated around describing entities and their properties, as shown in [2]. This is also the nature of the information usually found in web pages: in Wikipedia, each page is dedicated to one entity, and none to relations between entities. Also, most of the current approaches generate data through automated means, by parsing online data sources or exposing legacy databases in RDF format. This has two shortcomings: the only relations present in Linked Data are those detectable by a computer program (so only explicit relations can be detected), and also the decision of whether the data is correct or not is left to the computer. Moreover, the current quantity of available linked data in the largest such database was 4.7 billion RDF triples [2], compared to over 1 trillion of web pages in 2008 [7]. This tells us that the current approach of exposing the web's data in a structured form is not scalable enough when compared to the explosive growth of social content since the advent of Web 2.0 and the social web: tweets, statuses, blogs, wikis and forums are all very hard to understand for a computer program.

Therefore, it is our strongly held belief that general linked data would benefit from a social component, allowing its creation to be crowdsourced among enthusiasts, given that they are motivated correctly, without compromising data integrity. We envision that people should be able to easily create links between any two online entities identifiable by a unique URI and to associate extra information to these links. If this process of creating linked data is turned into a community process, the validity of the links can then be subjected to community scrutiny, a model that is not too scalable, but has proven to work given enough contributors in Wikipedia's case. The possibility of linking resources is already built in the HTML standard; however, the amount of extra information one can currently associate with a link is limited. Also, links in a webpage cannot generally be subjected to community examination for validity, and cannot be easily removed from the page.

A tool for editing links between entities and for visualising the most important links between contents is not available yet, and is a necessary step forward for communities to support the creation of linked data. However, this task of generating new linked data should be approached with care, since providing structured data requires a certain amount of rigor and time, whereas most people lack at least one of the two. This is why providing structured data for an ordinary user is still a challenge, as proved by the Freebase [8] project, and why currently linked data which is not generated automatically is created by experts or dedicated people.

[6] http://www.opencalais.com/
[7] http://googleblog.blogspot.com/2008/07/we-knew-web-was-big.html
[8] http:/www.freebase.com

1.3 Article outline

The remainder of the article is structured as follows. Section 2 presents relevant works that are related to the kleenk platform and how our platform relates to each of them. Section 3 introduces a scenario that will be used throughout the article to exemplify the utility of our proposed model and framework. Section 4 introduces the new type of link we propose, the kleenk, and discusses its formal definition and evaluation mechanisms. Section 5 presents major challenges that have to be overcome by an implementation of our proposed concepts. Section 6 discusses how kleenks can be modelled with existing theoretical frameworks, and why we have chosen RDF. In section 7 we present our current implementation of kleenks, a platform aimed at connecting scientific content through a crowd-sourcing mechanism. Finally, in the last section, we present our conclusions and future works.

2 Related works

Here, we have chosen a few relevant works that treat the same problems as mentioned previously: adding a social dimension to the web of data, using crowdsourcing to build up the web of data, or ways to open up linked data to the big public, which might be the only fighting chance of keeping up with the growth rate of online content.

ConceptWiki[9] tries to apply the wiki concept to linked data. It contains a list of concepts as specific as "an unit of thought". Any person with an account on the website can edit the concepts and there are two main sections on the website right now: WikiProteins (which contains information about proteins) and WikiPeople (which contains information about authors in the PubMed database). The WikiPeople sections seems to be populated by extracting information from PubMed, an important technique in order to encourage user adoption that our proof-of-concept implementation,
kleenk.com, currently misses. Simply put, users tend to consider a website more reliable if it has more content on it. However, for ConceptWiki, this content is entity-oriented and is created automatically by machines instead of being created by humans (just like in DBpedia). Users can edit the existing content or add a new one, but the quantity of information needed to complete the page of a person can be quite daunting, which is why we suspect that ConceptWiki isn't still adopted on a large scale. We have derived one very important lesson from this project: using machines in order to generate enough data to bootstrap a community is a very good idea, as long as it is not too complicated for humans to emulate what machines are doing (or said differently, machines do now know the difference between user friendliness and otherwise).

Last but not least, Facebook Open Graph (v2) is a recent development of the social networking giant, allowing people to publish their social life as something very similar to RDF triples. People can now connect themselves to other entities by verbs like watch, read and listen, instead of the traditional like. Friends can afterwards rate and comment these actions, therefore this approach has also a very strong community evaluation component. However, this platform lacks in two respects: the first is generality, as it only connects people with entities, and through a pretty limited amount of actions (Facebook has to approve all new actions, giving it complete control over the ontology of predicates that appear); the second is aggregated visualisation capabilities, which is actually what makes the web of data interesting for the regular user: the ability to discover new content by navigating from content to content.

[9] http://conceptwiki.org/

The fact that there are a number of projects solving the same problem as us, some even approached by internet giants or the academia gives us the strength to believe that we are working on the right problem. However, our proposed solution is unique, in that it lets users easily create their own linked data, while giving them access to powerful visualisations, as we will shortly see in the next sections.

3 Working Scenario

We will use an academia-related working scenario in this article.

Rob is a PhD student in computer science and he is reading a lot of books and papers related to his subject, which is artificial intelligence. He is testing a lot of applications and algorithms to see how they perform in different scenarios. He would like to discuss his findings with other researchers to have their opinions and also make his results easily accessible. He is discussing with his friends and also he publishes multiple articles but he feels that the feedback is limited and delayed (at least a few months from an article submission to its publication). Rob also has some younger friends that study the same topic. Whenever they find a new interesting article or application they ask Rob about it: What's important about this article? How does this application relate to application Y? Rob could tell them to read his articles but that may take a lot more time and his friends may get confused and get lost in other information they might not need. He gives them the answer but he knows that there may be more students out there that would benefit from those answers. How can he structure this information, and where to put it, so that it can be easily found by all interested researchers?

4 Kleenks

In this section we will propose a solution to the problem stated in section 1.2. We start by considering a simplified model of the Web of Data which allows us to explain the role of our approach and how it fits in the existing landscape. We finish by identifying the main challenges for implementing our proposal.

4.1 Web of Data

We consider a simplified model for the Web of Data which consists of the following elements: *contents*, *entities*, *links*, *software agents*, *humans* and *ontologies*.

Contents represent any type of unstructured data such as text, images, sounds or videos and they may, or may not have, an URI that uniquely identifies them. Entities can represent anything such as places, people or articles and they are uniquely identified by URIs. Links connect two entities, have an associated type and they can represent any relation between entities. By software agent we understand any software application (desktop, web or mobile) that uses the Web of Data. Also, we consider that humans can access entities and links directly, making abstraction of the browser or any application in between. Finally, ontologies can be used by both humans and software agents to understand the links between entities.

4.2 A new perspective, a new type of links

Inspired by the explosion of content in Web 2.0, we believe that the Web of Data could also use an internal perspective in which links are created from the user's point of view. We believe that the Web of Data needs a new social, unstructured and collaborative dimension that would bring people, unstructured content, entities and links closer to each other (Figure 1).

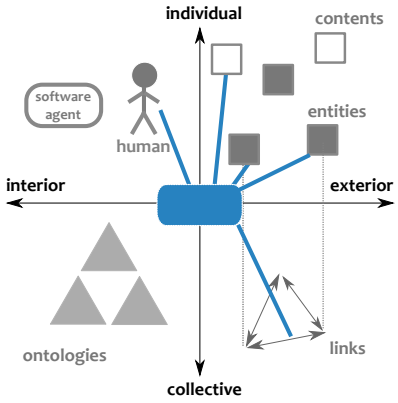

Figure 1. Social, Unstructured, Collaborative dimension to the Web of Data

We argue that this can be achieved through a new type of links, that we call *kleenks* (pronounced "clinks"), which are collaborative links created, evaluated and consumed by the users of the Web of Data. A kleenk (Figure 2) is a directed connection and consists of the following (below the words "entity", "content" and "link" have the meaning considered in the simplified model of the Web of Data from the beginning of this section):

1. **Source.** The source of a kleenk is an entity.
2. **Target.** The target of a kleenk is another entity.
3. **Type.** The type is a verb or expression that summarizes the link from the source to the target.
4. **Contents.** The contents represents the most important elements of a kleenk and they can have different roles:
   - *Description*. Descriptive contents can be simple text paragraphs, other media contents such as images and videos or even domain specific. They provide more details about the connection and they are added by the creator of a kleenk.
   - *Feedback*. As with descriptive contents, feedback contents can take any form but they are added by other participants to the kleenk (other people or software agents).
   - *Evaluation*. Evaluation contents must provide means to obtain quantitative data about the quality of a kleenk and they can take the form of ratings, like or thumb up/down buttons etc.

Kleenks are collaborative links because new content can be added to a kleenk at any time by its creator or by other participants. Kleenks have un unstructured dimension because the content added to a kleenk is in an unstructured form. Finally, a kleenk is social because it provides a mechanism for users to express their position (like, agree, disagree, etc.) with respect to it.

The term "kleenk" is actually a short version for collaborative link with a slightly different spelling since the term "clink" has been used in other works such as Project Xanadu[10] and we wanted to avoid confusion.

Let's take an example. Rob, from our first working scenario, reads a paper X that talks about an efficient implementation of an algorithm described in another paper Y. He will create a kleenk from the article

[10] http://www.xanadu.com

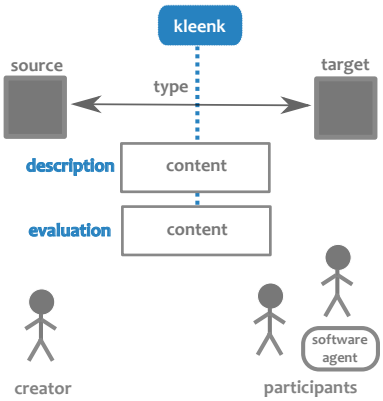

Figure 2. Elements of a kleenk

X to article Y with the type "efficient implementation of". Also, if the implementation is accessible on the internet he can also create a second kleenk from X to the implementation with the type "implemented here". As a description of the first kleenk he will provide a few details about what exactly makes the implementation efficient. Other researchers can express their opinion about the implementation directly on the kleenk, and comment for instance that the performance improvement is visible only on a particular class of input data. Other implementations can be kleenked to the same article X and the implementations can also be kleenked between them. Now, whenever an younger friend of Rob finds paper X he will quickly see the most important implementations of the algorithm and the relations with other important papers and they can continue their research without interruption.

4.3 Benefits and quality of kleenks

One main feature of kleenks is the ability to add unstructured content, in any form, to structured links. This has multiple benefits for both the user and the Web of Data. First because kleenks are richer in content than simple links, this makes them important on their own. Up until now, in the Web of Data, it is rare that links are very important on their own but rather in sets that describe an entity or a topic. We believe that making each link important on its own will engage people more in creating meaningful links.

Second, allowing people to create links with content will also facilitate the apparition of new links of high abstraction level that otherwise would have been impossible to extract automatically.

Allowing people to contribute to existing kleenks with new content is meant to make kleenks become more accurate and complete. However, as it has been seen in many projects such as Wikipedia and StackOverflow, an explicit evaluation system for user contributed content is necessary. The design of rating systems has been widely studied in computer science [7]. An overview of techniques that can be used to heuristically assess the relevance, quality and trustworthiness of data is given in [1].

Also, allowing social validation through mechanisms such as likes, agree/disagree or ratings allows important kleenks to step ahead of the less important ones guiding the users through what is important and what is less important. Of course the best way of validating a content can differ from domain to domain and each platforms that uses kleenks is free to choose the method that is more suitable.

5 Challenges

In the previous section we have introduced a new way of creating links in the Web of Data, at the conceptual level. This new type of links are called "kleenks" and they are collaborative links which contain unstructured content in addition to the typical RDF predicate. We believe that this approach will engage everyday users to take a more active part in building collaboratively the Web of Data and bring it to its full potential. However, implementing a system based on kleenks, be it targeted to a specific domain or as a general platform, raises a few challenges that must be properly addressed in order to be successfully used.

5.1 Access to entities

A kleenk, as an RDF triple, is a link that connects two entities and in addition it adds more content to the link. Letting regular users create such kleenks raises an important question: *"How will a user quickly select the entities he's interested in kleenking?"*.

The answer to this question depends on the type of platform: domain specific or general. In case of a domain specific platform it means that the user will kleenk entities he's working with. Usually these entities are already gathered in some databases and the kleenk platform only needs to integrate with these databases to provide quick search of the entities the user wants to kleenk.

On the other hand, a general platform is faced with a much more difficult question due to inherent ambiguities. If a user wants to use "Boston" as the source of a kleenk the platform has to decide whether it's about the city, the band or the basketball team. In this context we believe that semantic searches and large open databases such as DBpedia and Freebase will help in the disambiguation process.

Also, the user might want to kleenk things that don't yet have an URI and the platform must be able to create such URI's on the fly.

5.2 The ontologies for kleenks

Even though kleenks contain unstructured content, their type, as with RDF links, will still be a predicate in an ontology, allowing computers to have at least a basic understanding of what a kleenk means and use them in new ways. However, allowing users to create any type of links between entities means that it is very hard to develop a comprehensive ontology from the start. A kleenks platform would have to provide a mechanism that would allow users to define ontologies, such as in Freebase, or it must integrate with platforms that allow users to build ontologies such as MyOntology.

5.3 Visualization and privacy

Allowing users to create kleenks between any two entities has the potential of creating a very big number of kleenks. Users must be able to handle a big number of kleenks related to the entities that are of interest to them. Since kleenks form a graph structure, we can use visualisation techniques for graphs and create interactive ways of navigating the kleenks. We believe that since kleenks contain more content on the "edges" between the nodes, than just a simple predicate, more interactive and engaging visualizations can be built.

Since kleenks contain more content than simple RDF links and since most of this content will be based on the user's experience, the problem of the visibility of a kleenk must not be neglected. A user might want to create a kleenk between two entities and allow only a limited number of persons to see it. Also, kleenks can be used

to collaboratively build some data (i.e. state of the art on a topic) which might, at least on its early stages, be visible only to a limited number of people. So, a kleenk platform must also provide proper mechanisms for kleenks' visibility.

6 Modeling kleenks

In this section we will look at the theoretical and technical aspects of modeling kleenks using existing techniques in semantic web. We will first analyze different alternatives and motivate our chose for one of them. Finally, we will give an example of what a kleenk might look like.

6.1 Theoretical model

Basically, the kleenk model could be seen as an extension of the RDF model with support for unstructured data. In the semantic web many extensions of the RDF model have been proposed during the last years. There are extensions dealing with temporal aspects [5], with imprecise information [8], provenance of data [4] or trust [9]. In [10] a general model based on annotation is proposed which generalizes most of the previous models.

All the above mentioned techniques are based on the named graph data model, a well known technique in semantic web to attach meta-information to a set of RDF triples. Even though these techniques could be applied to model kleenks, that would require that each kleenk has its own named graph (with its own URI), in order to associate the unstructured content with it.

A different technique, known under the name of RDF Reification, is described in the RDF specification [6]. This technique has well known limitations and weak points such as triple bloat and the fact that SPARQL queries need to be modified in order to work with reified statements. However, we believe that this techniques is the most suitable for modeling kleenks because a kleenk needs many different types of meta-information associated with it: creator, description content, feedback content (i.e. comments), evaluation content (i.e. ratings) and possibly other domain specific data.

7 kleenk.com

7.1 Description

kleenk.com [11] is an online collaborative platform for linking scientific content. The project's motto is: "Smart-connecting scientific content". It is allows users to link scientific contents, revealing other relations than citations, such as:

- paper P1 implements the algorithm in paper P2 (relation: "implements algorithm in")
- diagram D1 is an explanation for the theory in paper P2 (relation: "explains the theory in")
- algorithms A1 and A2 solve the same problem (relation: "solves the same problem as")

This kind of relation is not easy to extract neither by an automated program, and nor by humans that are just starting their research in a certain area. In Europe, the first year of a PhD program is usually dedicated to researching the state of the art, which consists of reading many scientific contributions by other authors and creating mental links like those mentioned previously. Given the exploding number of scientific works, conferences and journals it is hard to keep up-to-date even for a scientific advisor, which makes the work of a starting researcher even harder. Kleenk actually solves this problem by allowing the community to create and visualise kleenks between the contents.

This platform is aimed at the following groups of persons:

- PhD students which need community guidance in order to read the most relevant and up-to-date materials related to their subject
- professional researchers who need to stay in touch with the vibrant scientific community's developments
- other people interested in quickly gaining an overview of a scientific domain

The platform allows the easy selection of content to kleenk from a number of sources by manually adding it, importing it from web pages (such as ACM or IEEE public pages of articles) and even by importing BibTeX bibliography files. Once all the content a user wants to kleenk is available in the platform, the user can start creating kleenks by selecting a source and destination content.

After they are created, kleenks can be shared with research fellows or made public, and grouped around meaningful ideas using tags. Every time a new content is created or updated, the interested users are notified using their personal news feed. Therefore, changes to a kleenk or any comment reach out across the entire community instantly.

Authors have the chance to kleenk their own papers to existing ones, and by subjecting these kleenks to the community scrutiny, the platform makes it possible for them to obtain early feedback for their ideas. In today's society, when the internet allows information to be propagated from one end of the world to another in seconds, the traditional peer review system is becoming more and more criticized due to the number of months passed from submitting the work to actual post-publication feedback from the scientific community. Our service aims to complement the quality and thoroughness of the peer review system with the opinion of the crowd. One important observation is that the opinion of the crowd is not necessarily misinformed, as proven lately by the tremendously successful service for programmers StackOverflow [12]. This website is a collaborative question answering system, with world renown experts easily connecting and answering each others' questions. We think that the scientific community would benefit from a low-latency alternative to obtaining feedback for a piece of work.

7.2 Implementation of the theoretical framework

Having earlier detailed the kleenk model and characteristics, we will now underline which instantiation of the general principles was used in order to implement this knowledge sharing platform. First of all, in our particular case, the kleenk has the following elements:

- **the source, destination and type** - these are also present in the general model
- **the description** - this is specific to this pair of content, and represents a more detailed explanation of the type. It should be used in order to motivate the choice of type and to give more relevant results
- **comments** - since each kleenk has its own set of comments, these can be used in order to discuss the relevance of the link and to give extra information by anyone who can see it. These are similar to

[11] http://app.kleenk.com

[12] http://www.stackoverflow.com

Wikipedia's talk pages, which are used by contributors to clarify informations in the main page

- **ratings** - together with ratings, these allow the community to evaluate the quality of a kleenk. In the visualisation, kleenks with better community score (which is computed from the ratings, number of comments, number of views and a number of other metrics) are displayed with a thicker connecting line, signifying a greater importance. Ideally, an user who is interesting in exploring the web of scientific articles will first navigate the most important kleenks.
- **privacy level** - as already mentioned in the general model, there should be a privacy setting associated with each kleenk. This allows users to first try out their own ideas in a personal incubator before promoting them to the whole community. In our implementation, there are 3 privacy levels: private (visible only to the owner), public (visible to anyone) and shared (visible to research fellows, which can be added through a dedicated page, given that they also agree).
- **tags** - each kleenk can be part of one or more tags. This is actually a mechanism for grouping tags related to the same idea or topic under a single name. For example, when writing this article, the authors created a "Kleenk Article" tag which contained the relevant bibliographic items and the kleenks between them.

The visualisation of the graph induced by the kleenks is done, as mentioned in the description of the general model, using consacrated layout methods. Specifically, in our case, we use an attraction-force model.

kleenk.com is a linked data application, conforming to Berner-Lee's vision of the future of the web. Contents, kleenks and tags all have persistent URIs that can be dereferenced in order to obtain linked data. One other interesting side-effect of this is that interesting scientific applications can emerge on top on the data contributed by the users to kleenk. For example, new scientometric indicators based on kleenks could be computed by a 3^{rd} party application.

7.3 Use case example

7.3.1 Obtaining feedback for a recently published article

Alice is a fresh PhD student in Semantic Web, who is overwhelmed by the vast amount of publications on this topic. Being a first year student, she has to complete a document describing the state of the art by the end of the year. Being a Facebook user, it's easy for her to create an account using one click on kleenk.com, since it features integration with Facebook's login service. Once logged in, she adds her colleagues who already have a Kleenk account as research fellows and now can easily see their shared tags. She studies the visualisations and grows to see a few important articles which are in the center of most tags, and starts reading them. Since she pays close attention to her news feed, she can easily see in real time what connections her colleagues are creating, and they all obtain quick feedback from their advisor, via comments and ratings.

Since she will be writing a survey article as well, she started creating a tag specifically for the bibliography of the article. First, the tag is private, since it is a work in progress and she doesn't want to share it with anyone. As the text of the article and the bibliography mature, she changes the visibility of the tag from private to shared, so that her research fellows can express their opinion on the connections she is making. After receiving the final approval for publication, she makes the tag public and includes the visualisation of the bibliography in a presentation for her department.

8 Conclusions and Future Works

This article discusses the current context of the Web of Data, analyzes a few of its current limitations and focuses on the need to engage regular users in the creation of semantic links. We propose a new approach inspired by the success of Web 2.0 techniques such as wikis, blogs and social networks.

The main contribution of this paper is the concept of *kleenk* which is a collaborative link that contains unstructured data in addition to the classical RDF predicate. We discuss the importance of allowing users to add unstructured data to the Web of Data and how this approach could lead to the creation of links which would otherwise be impossible to automatically parse from existing datasets.

We also identify the main challenges of a platform allowing users to create kleenks: access to entities, collaborative ontology creation, visualization of kleenks and privacy. These challenges have to be properly addressed for a system to succeed in applying kleenks. We finish by introducing a free online platform, www.kleenk.com, which applies successfully the concept of kleenk in the scientific research domain and discuss how the identified challenges have been addressed.

Future works include:

- testing kleenks in other domains in order to see what would be the specific problems in adopting them for those domains
- building a common kleenk schema in order to describe kleenks
- defining scientometric metrics which are kleenk-related instead of the old citation-related approachess
- populating the kleenk.com database automatically with kleenks for citations in order to bootstrap the community use

REFERENCES

[1] Christian Bizer and Richard Cyganiak. Quality-driven information filtering using the WIQA policy framework. *Web Semant.*, 7:1–10, January 2009.

[2] Christian Bizer, Jens Lehmann, Georgi Kobilarov, Sören Auer, Christian Becker, Richard Cyganiak, and Sebastian Hellmann. DBpedia - A crystallization point for the Web of Data. *Web Semantics: Science, Services and Agents on the World Wide Web*, 7(3):154–165, September 2009.

[3] K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor. Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, pages 1247–1250. ACM, 2008.

[4] Renata Dividino, Sergej Sizov, Steffen Staab, and Bernhard Schueler. Querying for provenance, trust, uncertainty and other meta knowledge in RDF. *Web Semant.*, 7:204–219, September 2009.

[5] Claudio Gutierrez, Carlos A. Hurtado, and Alejandro Vaisman. Introducing Time into RDF. *IEEE Trans. on Knowl. and Data Eng.*, 19:207–218, February 2007.

[6] http://www.w3.org/TR/rdf primer/. RDF Primer, February 2004.

[7] Audun Jøsang, Roslan Ismail, and Colin Boyd. A survey of Trust and Reputation Systems for Online Service Provision. *Decis. Support Syst.*, 43:618–644, March 2007.

[8] Mauro Mazzieri and Aldo Franco Dragoni. Uncertainty reasoning for the semantic web i. chapter A Fuzzy Semantics for the Resource Description Framework, pages 244–261. Springer-Verlag, Berlin, Heidelberg, 2008.

[9] Simon Schenk. On the Semantics of Trust and Caching in the Semantic Web. In *Proceedings of the 7th International Conference on The Semantic Web*, ISWC '08, pages 533–549, Berlin, Heidelberg, 2008. Springer-Verlag.

[10] Antoine Zimmermann, Nuno Lopes, Axel Polleres, and Umberto Straccia. A General Framework for Representing, Reasoning and Querying with Annotated Semantic Web Data. *Elements*, pages 1437–1442, 2011.