Beyond Data: Building a Web of Needs

Position Paper

Florian Kleedorfer
Research Studios Austria
Forschungsgesellschaft mbH
Studio Smart Agent Technologies
Thurngasse 8/2/16
1090 Vienna, Austria
http://sat.researchstudio.at
florian.kleedorfer@researchstudio.at

Christina Maria Busch
Research Studios Austria
Forschungsgesellschaft mbH
Studio Smart Agent Technologies
Thurngasse 8/2/16
1090 Vienna, Austria
http://sat.researchstudio.at
christina.busch@researchstudio.at

ABSTRACT

The Web as related to commerce suffers from a fundamental asymmetry. While there is a great number of commercial offers available, consumer needs are rarely represented explicitly. Thus, the most widely applied process of connecting the prospective consumer of a resource with its supplier is Web search. In Web search, the user needs are implicit, driving the interaction, and therefore only the interaction partners can try to deduce them. We present an approach for a) publishing needs on the Web of Data and b) building a protocol that allows decentralized matching of needs and communication between need owners. Albeit inspired by the analysis of marketplaces, the proposed framework allows for a much broader range of social applications, such as collaborative problem solving, help organizing the sharing economy or finding interesting people to meet.

Categories and Subject Descriptors

Information Systems [**World Wide Web**]: Web applications; Information Systems [**World Wide Web**]: Information System Applications—*Collaborative and social computing systems and tools*

General Terms

semantic Web, ontology matching, instance matching, protocols, linked data

1. INTRODUCTION

In[1] any society that is based on division of labour, the same principle is always present in one form or another: *Transfer of resources*. This transfer takes place when one actor has control over a resource she is prepared to give away and another feels a need that can be satisfied by obtaining the resource. Thus, these actors are connected by an asymmetric relationship between a *need* and an *offer*.

Offer and need differ substantially in their ontological status. Both are abstract notions, intentions of taking part in the transfer of the resource in question. However, and this is crucial, an offer can *appear as the thing being offered*, whereas a need always denotes the absence of something, so it can never appear as the thing that is required[2] until it is satisfied and has ceased to exist.

This difference leads to the form of the marketplace, where physical objects are on display on different kiosks, representing offers, i.e., their owners' intentions to transfer them. Buyers must walk from kiosk to kiosk, searching for things that satisfy their needs, simply because the latter cannot be represented by any objects. Thus, buyers' needs are unknown to anyone else until negotiation about a transfer takes place.

Why does the current state of the Web as related to e-commerce follow this form almost exactly, thereby perpetuating the asymmetry of need and offer? Web sites offer goods for sale, amounting to a staggering number of distinct offers. On the other hand, users who want to satisfy a need have to perform Web search and peruse the results in order to find relevant offers. As in the classic marketplace, the users' needs are unknown to anyone but to the users, and a lot of effort is spent in trying to guess the needs through the analysis of browsing or buying patterns and similar approaches. We argue that this form of market is no longer the necessary form. Web technologies allow for needs as well as offers to be represented as documents of equal status, all published on the Web. Automatic matching services can find suitable pairs for a resource transfer. The transfer itself can be mediated by a set of protocols, reducing human interaction to a necessary minimum. Stressing the fundamentally different status of needs in such a system we refer to it as a *web of needs*.

In this paper, we discuss requirements for such an infrastructure, propose a design, also explaining the current state

[1]Note that this work is based on the ISWC 2011 Outrageous Ideas submission *Building a Web of Needs*, see `http://iswc2011.semanticweb.org/fileadmin/iswc/Papers/outrageous/iswc2011outrageousid_submission_1.pdf` [2013/03/07].

[2]It can of course be represented by a description – which is exactly the case in point – but it cannot appear as the *thing actually required.*

LDOW2013 May 14, 2013, Rio de Janeiro, Brazil

of implementation, and address benefits and challenges.

2. OUTLINE OF THE INFRASTRUCTURE

Our overall goal is to create a decentralized infrastructure that allows people to publish documents on the Web which make it possible to contact each other, and this should only happen if all parties involved have an interest in doing so. The said document may contain a description of a product or service required or offered, a description of a problem to be solved with the help of others, an invitation to social activities, or anything else users may think of. On this abstract level of description, the document can be said to represent an interest in or a need for some kind of interaction with others. Therefore, we refer to this document as a *need*. It is the central entity of the system we propose. Each need has a globally unique identifier and an *owner*, i.e., a person or other entity that creates and controls it. When need owners want to communicate with each other, a *connection* object is created for each need involved. The connection is the second important entity in our design.

For the interchangeable formulation of needs, a common modeling language is required along with a publishing mechanism. The resource description framework (RDF)[3] as a basic technology and the principles of linked data publication[4] are well suited for this purpose. Moreover, with the GoodRelations ontology[7], groundbreaking work has already been done with respect to the description commercial offers and needs. For describing the resources being sought or offered themselves, however, no single language can be expected to fit all needs, so the graph describing a need contains an extension point where an arbitrary RDF graph can be inserted. This should make it possible to re-use existing RDF content. Furthermore, the description language must allow constraining and combining needs, as well as constraining need combinations, which is required for formulating complex needs.[5]

When needs are published on the Web, independent matchmaking services can crawl them (or be informed of them in other ways) and look for suitable matches. A protocol has to be in place to define how these services inform the need owners of possible matches. Matching services are required to honor the description language specific to needs; in addition to that, they must also be able to discover matches between resources described in different vocabularies. Simple solutions can use off-the-shelf information retrieval technology like Solr[6] (or, more specifically, SIREn[3]). For more detailed matching, ontology and instance matching tools[4] can be adapted to this problem domain and serve as starting points for building robust matchmaking services. The approach developed by Abramovich and Sheu[1] could be used to compute solutions for complex problems automatically.

Finally, a protocol is required to allow for interested parties to communicate. The most simple and probably most important variant is a chat protocol. However, for interactions where natural language is inappropriate, in retail, for example, the protocol should allow for predefined workflows encapsulated in distributed transactions.

[3]See http://www.w3.org/TR/rdf-primer/ [2013/03/09]

[4]See `http://www.w3.org/standards/techs/linkeddata#w3c_all` [2013/03/07]

[5]For example, one may need a flight to location A, a hotel in A, car rental from A to B, a flight back from B.

[6]See `http://lucene.apache.org/solr` [2013/02/25]

3. ARCHITECTURE

3.1 Nodes

Conceptually, the proposed infrastructure is a network consisting of at least three different types of nodes: *owner applications*, *web of needs (WON) nodes*, and *matching services*.

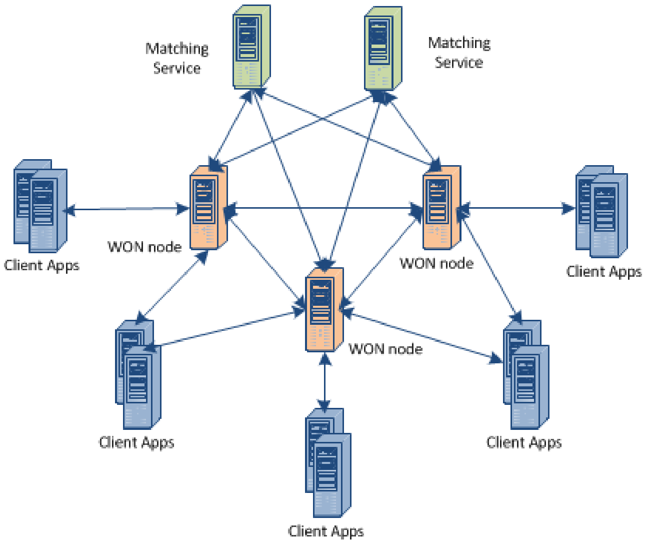

Figure 1: Deployment diagram indicating types of nodes and communication paths. Needs are published as linked data on WON nodes, matching services crawl the data and report good matches; applications make the functionality available to users.

- *WON nodes*, or *nodes* for short, store needs, receive *hints* from matching services, and serve as communication relays between need owners. All needs stored on a WON node are published as linked data.
- *Matching services* constantly crawl the Web of needs, i.e., the part of the linked open data graph located on WON nodes. Whenever they find needs that satisfy each other's matching criteria, the respective need owners are given the appropriate information.[7]
- *Owner applications* are applications that make use of the technology, either by direct interaction with users or by connecting to enterprise resource planning systems or other corporate software. They connect to one or more WON nodes to manage (create, update, delete) needs that are stored there and to communicate with other need owners.

3.2 Communication Protocols

As shown in figure 1, the WON node is the central building block of the network. It communicates with matching services as well as with owner applications. WON nodes serve as proxies for the owner applications, thereby hiding

[7]Evidently, other communcation models are conceivable here, for example WON nodes could query a matching service each time a new need enters the system. This and other communication models are not covered in this paper.

them from the rest of the network. Only the node can link a published need document to the owner application that controls it. All communication directed at a need is routed by the WON node to reach its owner. Furthermore, figure 1 shows that matching services are independent of nodes. They may communicate with any number of them, but do not communicate directly with owner applications. Similarly, owner applications may communicate with a number of nodes, but not with matching services directly. The communicative acts between the different participants in this system can be categorized by their respective purpose:

- Owners performing CRUD (create, read, update, delete) operations with needs
- Matching services crawling needs
- Matching services informing owners of suitable matches
- Owners engaging in communication with other owners so as to satisfy their needs

As there are three types of communication (owner-node, node-node, matcher-node), we propose three communication protocols:

The *owner protocol* governs the communication between owner application and WON node. It comprises CRUD operations and specifies how messages directed at the need are relayed to the owner. This protocol conceptually is a two-way client-server protocol.

The *need protocol* governs the communication between needs (and therefore, also between WON nodes). It defines how a connection can be established and managed between two needs, and how messages are exchanged over this connection. This protocol, too, is a two-way client-server protocol.

The *matcher protocol* defines how matching services inform needs about possibly interesting matches. In the description at hand, this is a one-way client-server protocol where the node is the server and the matching service is the client.

In addition to these, one may name the principles of linked data publishing (and, more specifically, the linked data platform[8]) as the protocol that allows for publishing and crawling needs in a standardized way. Figures 2, 3, and 4 illustrate the message exchanges taking place for creating a need, sending hints, and connecting needs. As shown in figure 4, each time a need owner initiates a connection between one of her needs and another one, a connection object, serving as a communication relay, is created at each end. These connection objects are also published as linked data, referring to each other and to their respective needs.[9] Thus, information about ongoing and past transactions is available as linked data and can be used to discover WON nodes starting from a given node, to collect information for the improvement of matching systems, and to gain up-to-date market insight.

[8]See `http://www.w3.org/TR/2012/WD-ldp-20121025/` [2013/03/07]

[9]Of course, no message content is published.

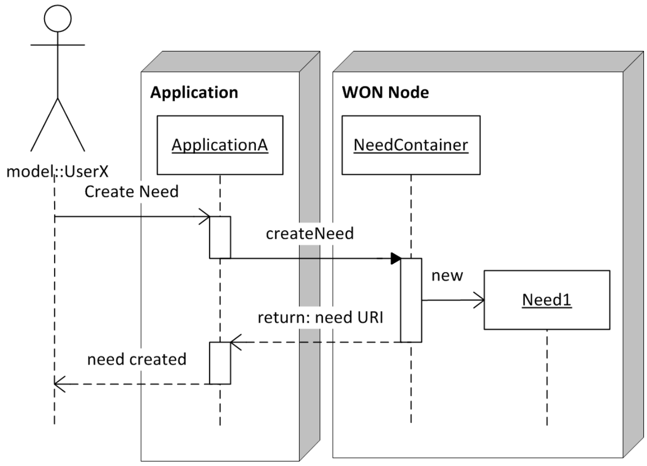

Figure 2: Sequence diagram illustrating the communication when a need is created by a user. The need object and the need container are published as linked data.

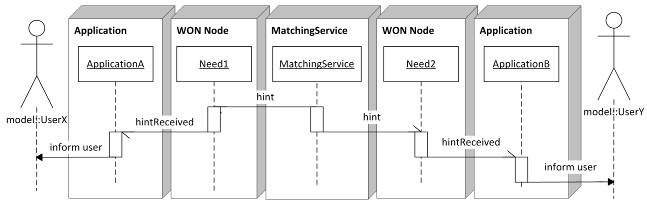

Figure 3: Sequence diagram illustrating the communication taking place when a matching service sends a hint message to two need owners.

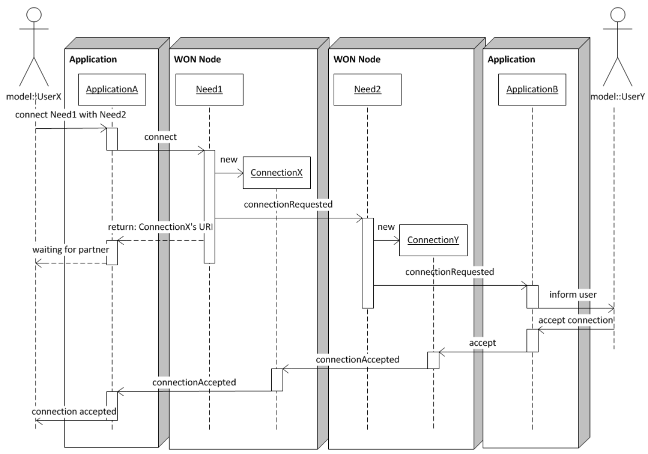

Figure 4: Sequence diagram illustrating the communication taking place when the owner of a need initiates a connection with the owner of another need and the counterpart accepts the connection. Both need objects and both connection objects are published as linked data.

4. NON-FUNCTIONAL REQUIREMENTS

In the following, we state important requirements that the solution must meet.

Access. The only technical prerequisite for using the system should be access to the Internet.

Usability. The system should allow users to post needs, connect with others and negotiate in a simple way.

Fairness. All published needs should be treated as equal.

Simplicity. The technical adoption of the system should be as simple as possible.

Scalability. The problem of matching all needs with each other naively has a complexity $O(n^2)$.[10] Matching quality and speed should not be affected by a growing number of needs in the system.

Timeliness. Fast matching results are important for a good user experience. However, in some situations, longer waiting times may be acceptable, for example, when a user defines a need for something outrageous that doesn't exist yet.

Robustness. The system we propose will not be very useful if users experience frequent service outage or inconsistencies, and therefore refrain from using the system for important aspects of their lives.

Security. A system that people rely on for performing important tasks must be immune against eavesdropping, manipulation, impersonation, and more threats of this sort.

Privacy/Anonymity. The central aspect of our proposed infrastructure is the publication of personal needs. It is evident that this is a privacy disaster if those needs can be traced back to their respective authors, An ideal solution therefore allows for fully anonymous use, which may be very hard to achieve in some cases.

5. OPEN ISSUES

In this section, we explain aspects of the infrastructure that are implied by the design as stated so far but are too unclear to include them in the design.

Need routing. For achieving timely results, merely crawling linked data before matching cannot be sufficient, though it provides a robust foundation. Moreover, in an open system, many independent matching services could emerge, and they could specialize with respect to domain, location, or other aspects. This would entail that some matching services produce better results for a given need than others. It would therefore be desirable to have a routing infrastructure in place that passes information about new needs to the the best suited matching services. Possible solutions could build on message oriented middleware with publish-subscribe queues for all relevant need categories or locations, and there may be entirely different solutions.

Distributed matching and need refinement. As stated above, a need can be combined with other needs to form arbitrarily large and complex structures. It is conceivable that these structures are created by users over the course of multiple interactions with the system, during which matching services propose to replace a single need by a more complex structure. This could be beneficial in situations where a need describes a very complex problem for which no singular solution exists, but when broken down into smaller elements, solutions can be found easily for each of them. Such functionality would require additions to the matcher protocol and novel matching service functionality.

Trust. For many kinds of transactions, it is not irrelevant whether we can trust our counterparts. When formulating a need, it would be desirable to be able to make constraints about the level of trust required by other needs in order to be elegible as matches. This idea may conflict with the requirement of anonymity, however.

Workflow protocol and distributed transactions. For communication between end-users, we believe that chat is the most important communication pattern, but it is not for B2B or B2C communication. For example, between a retail store and a consumer, chat is quite unpractical. For such cases, the concept of a more rigid protocol, allowing for distributed transactions[11], may be much better suited. Ideally, the selection of the most suitable communication protocol is part of an initial connection negotiation between need owners.

Integration of payment solutions. For the proposed infrastructure to become generally useful, it is required that payment solutions be integrated such that commercial transactions can be performed. The availability of a protocol allowing for distributed transactions would enable payment solution providers to offer their services within the system at hand.

Stepwise disclosure of private information. Users may not feel at ease publishing their home address and phone number on the Web along with their needs. In accordance with the requirement of privacy, such information should be protected by the system. However, during negotiation with the owner of another need, it may become necessary to disclose such information, for example so as to arrange for delivery of goods. In chat-style communication this is of course at the discretion of the user; however, if a more rigid protocol is in place, processes could be automated to a much further degree if the disclosure of such data could be tied to certain protocol states.

User interfaces. Users should be able to define and publish their needs with ease. Research on linked data authoring has shown the difficulties of a domain-independent RDF editing tool [2]. It is unclear if domain-specific editors need to be created for the most important domains, or if it is possible to devise a domain-independent one. Moreover, a solution to this problem may be a system using domain-specific GUI definitions, similar to the Fresnel system[8], that can be loaded and used dynamically. Such definitions may be created by independent parties and published on the Web. To complicate things further, natural language could prove to be the best form of input if the wealth of past needs and transactions proves sufficient to bootstrap machine learning systems for the task. Other possible approaches to the problem of eliciting user needs encompass dialog systems similar to the popular Web game akinator[12] and systems allowing drill-down through relevance feedback.

Technological basis. Because of the availability of protocol specifications and implementations for security[13], distributed transactions[14], and many others that seem relevant

[10] One may argue the complexity be lower as offer/offer or need/need pairs need not be matched.

[11] For example, WS-BusinessActivity, `http://docs.oasis-open.org/ws-tx/wsba/2006/06` [2013/03/07]

[12] See `http://en.akinator.com` [2013/03/07]

[13] See `https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=wss` [2013/03/07]

[14] See `https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=ws-tx` [2013/03/07]

in our context, we decided to base our implementation on SOAP Web services. However, as much of the future developments is still unclear, REST[5] may be a better technology for implementing the protocol, as REST is simpler and well suited for use with the linked data platform. Another approach we are currently evaluating is using XMPP as a basis.
Cardinality of communication. So far, we only covered 1:1 communication. Although most intuitive and important (at least in the beginning), this is an inappropriate generalization. In general, multiple people may share the same need, and they may want to work out a solution collectively. It is worth noting that the proposed infrastructure would make it much easier than it is with current Web infrastructure to detect such situations; this in turn calls for an appropriate handling thereof. A solution supporting this N:M communication, possible with workflows for decision making could be created and integrated as an addition to the protocols.

6. CONSEQUENCES AND BENEFITS

Assuming, for the sake of the argument, the implementation of the proposed infrastructure at Web scale with a large user base, we see far-reaching potential consequences. It could allow for marketplaces currently fragmented in multiple dimensions, such as location, type of goods, customer segments, type of transfer (such as buying, rental, or barter), or simply by Web site (Amazon, Ebay, . . .) to amalgamate into one marketplace, raising competition, and possibly lowering price dispersion[6]. Moreover, consumers could have a single interface to that market – their preferred need management service provider – and could get rid of the burden of Web search or search on different platforms. As needs remain available online after a transaction is finished, users could be enabled to formulate new needs based on past ones, making recurring tasks easier to perform. A publicly available history of needs and how they were satisfied would represent a valuable resource for making informed political decisions or performing market research. Services could emerge that focus on needs instead of offers, e.g. allowing the re-use and collaborative improvement of need combinations that have already been used successfully by others.

In the long run, we may observe a unifying effect on the use of RDF vocabularies or ontologies as more and more market participants (producers, consumers and retailers) perform next-generation SEO to improve their need descriptions and thus obtain better matching results.

Finally, the growth of the sharing economy, currently fueling the rise of two-sided platforms such as airbnb[15], suggests that there is a need for a generic infrastructure that, like the one we propose, allows consumers to connect directly with each other.

7. FUTURE WORK

Current development focuses on delivering a proof-of-concept that is available as open source[16] and will be refined iteratively. We will provide an implementation of each of the required node types and run an online demo of each of them. Further work will provide solutions for authentication, security, and privacy, a more usable user interface, and a more scalable matching service. Moreover, we plan to research how users structure their needs cognitively so as to guide the design process further down the road.

[15]See `http://www.airbnb.com/` [2013/03/09]
[16]See `https://github.com/researchstudio-sat/webofneeds` [2013/03/08]

8. CONCLUSION

We have motivated the creation of a decentralized Web based infrastructure for the management and satisfaction of human needs. We have given an overview of its design, leveraging the semantic Web and linked data technology stacks and named important requirements that the solution must meet. Our proposal defines the overall shape of the system, and by doing so, it opens a range of new possibilities and questions which we have enumerated.

9. ACKNOWLEDGMENTS

This work was supported by the Austrian Federal Ministry of Science and Research and the Austrian Research Promotion Agency (FFG) in the BRIDGE project *WIN - Web of Needs Infrastructure*.

10. REFERENCES

[1] A. M. Abramovich and P. C.-Y. Sheu. Towards linked needs. In *Proceedings of the 2010 IEEE Fourth International Conference on Semantic Computing*, ICSC '10, pages 456–461, Washington, DC, USA, 2010. IEEE Computer Society.

[2] S. Davies, J. Hatfield, C. Donaher, and J. Zeitz. User interface design considerations for linked data authoring environments. In *LDOW*, 2010.

[3] R. Delbru. Siren: entity retrieval system for the web of data. In *Proceedings of the Third BCS-IRSG conference on Future Directions in Information Access*, FDIA'09, pages 29–35, Swinton, UK, UK, 2009. British Computer Society.

[4] J. Euzenat, C. Meilicke, H. Stuckenschmidt, P. Shvaiko, and C. Trojahn. Ontology alignment evaluation initiative: Six years of experience. In S. Spaccapietra, editor, *Journal on Data Semantics XV*, volume 6720 of *Lecture Notes in Computer Science*, pages 158–192. Springer Berlin Heidelberg, 2011.

[5] R. T. Fielding. *Architectural styles and the design of network-based software architectures.* PhD thesis, 2000. AAI9980887.

[6] T. Gupta and A. Qasem. Reduction of price dispersion through semantic e-commerce: A position paper. In *Proceedings of the Semantic Web Workshop*, pages 1–2. Citeseer, 2002.

[7] M. Hepp. Goodrelations: An ontology for describing products and services offers on the web. In *Proc. of the 16th Int. Conf. on Knowledge Engineering and Knowledge Management (EKAW2008)*, volume 5268 of *Springer LNCS*, pages 332–347, Acitrezza, Italy, September 2008.

[8] E. Pietriga, C. Bizer, D. Karger, and R. Lee. Fresnel: A browser-independent presentation vocabulary for rdf. *The semantic web-ISWC 2006*, pages 158–171, 2006.