Modeling the Web of Things from an IR approach

Héctor Cristyan Manta Caro
Faculty of Engineering, District University of Bogotá Francisco José de Caldas, 110311, Colombia
hcmantac@udistrital.edu.co

Juan M. Fernández-Luna
Department of Computer Science and AI, CITIC-UGR, University of Granada, 18071, Spain
jmfluna@decsai.ugr.es

ABSTRACT

Internet and Web technologies have evolved remarkably from their conceptualization. Nowadays, the origin of two novel paradigms have been triggered by the possibility of interconnecting not only traditional devices, smart phones, and wearable computing but also any object in the real world, and publishing Web-based services with dynamic content and data in real time. They are called the *Internet and the Web of Things*, respectively. The emergence of such paradigms implies a redefinition of the systems which they interact with, such as Information Retrieval systems. Thereby, it is essential to develop abstract models of Web representation, and simulation in order to establish new approaches in Information Retrieval for the Web of Things. A proposal for modeling the Web of Things based on a structured XML representation is described in this paper. This model has been designed with flexibility and modularity to allow the representation of multiple scenarios, being the conceptual source for future IR Systems development.

Categories and Subject Descriptors

H.1 [**Models and Principles**]: Miscellaneous;
H.3.5 [**Information Storage and Retrieval**]: Online Information Services—*Data Sharing, Web-based services*

Keywords

e-Things, Information Retrieval, Internet of Things, Search Engines, Sensor Web, Web of Things, Wisdom Web of Things (W2T)

1. INTRODUCTION

Nowadays, only 1% of real-world objects are connected to the Internet, but the expected number of connected devices is 50 to 100 billion by 2020 [6]. This new paradigm is referred to as the *Internet of Things (IoT)*, which describes technologies and research disciplines that enable the Net to adopt some intelligence and to venture into the real world of physical objects that are interconnected [4]. In addition, if we enable advanced Web access through virtual elements, which are abstract representation of things in the real world, we can create intelligent spaces which appears as the new paradigm called the *Web of Things (WoT)* [1]. At present, these paradigms bring new perspectives and challenges to systems interacting with them. In the context of Information Retrieval (IR), the new paradigms introduce dynamic factors to consider in detail: the WoT will abstract a huge amount of objects in the real world continuously producing a vast amount of information. Then its Web representation will incorporate status information or critical variables of interest that must be updated in real time and with frequent state changes, leading to highly dynamic and very large information sources.

This strong dynamism has not been well explored or evaluated in conventional retrieval systems. One of the most important open research topics, is the design, development, implementation and adaptation of real-time search engines that allow finding things, and information on variables of these things, as well as the features and services provided by them. This paper presents an approach for modeling the WoT as a basis for subsequent research in the development of new IR systems dealing with the dynamism and amount of "documents". However, for this purpose it is necessary to firstly develop an abstract model and a structured Web representation. In the next section, we present a state of the art in searching on the Internet of Things, and the Web of Things. The proposals are compared from different perspectives, conceptualizations and ordered according to sophistication. From simple search of embedded information on micro-devices attached to objects to elaborated search of things on semantic Web. We describe our model of the Web of Things from an IR perspective in Section 3, and our proposal of a structured WoT representation. Finally, Section 4 sums up and concludes.

2. SEARCHING ON THE WEB OF THINGS

The search for real-world entities (people, places, tangible and intangible things) will become an important and crucial service similar to current Web content search for media and documents [9]. There is also an increase in the relevance and worth of the information captured by sensors, the state, the properties, capabilities, functions and services that things may provide in the real world through their Web abstraction. The use of IR systems on the Web of Things is an issue of considerable complexity that imposes large demands on the design as the information is highly dynamic, inherently

ECIR'14 Information Access in Smart Cities Workshop (i-ASC 2014).
April 13, 2014, Amsterdam, the Netherlands..

distributed and with potentially colossal number of expected interconnected things[3]. In the rest of this section, several existing approaches are discussed, based on their area scale: i) only information found on the device connected to the thing, ii) things on a personal, iii) local, iv) metropolitan, as smart cities, or v) global area. Another distinction can be made by the range or possible results of search: sensor data, only sensors, sensor groups, things or physical objects based on constraints, or smart spaces.

Regarding IoT, Gander [2] presents a conceptual model of search engine for personalized networked spaces based on sampling of expressiveness and protocols responsive of space and time, focusing on the design of queries and data models resolved *in-situ*. The main contribution is its focus on the here and now, putting into consideration the high dynamism of information classifying the data as ephemeral, but with a low scalability factor. A similar hybrid approach is proposed in [7], where its main contribution is the novel and robust architecture that takes into account the dynamic collection and content. In this proposal, there is no Web representation of entities or a formal abstract model for the WoT.

The creation of a Web infrastructure to facilitate the integration, search, and interaction with smart things is presented in [10]. The proposal treats the location of a smart thing as the main property, structured hierarchically according to location identifiers. Searching space is larger than a personal smart space, considering the entire Web infrastructure. Another important contribution is the spatial hierarchy of things and their changing spatial association. The search engines return results at level of things, services or interfaces. However, the sensor level is not considered, nor are the dynamic changes on Web content. A line of approaches, we classified as State-based Searching of Entities on the WoT, propose the retrieval of things, which are in a particular state at the query time. For example, [9] introduces a method of ranking sensors formally modeled as random variables. A key contribution is the fact that each entity in the system is represented by a virtual counterpart with its own URL. The query language is not only based on keywords but also on the dynamic properties of entities captured by the associated sensors. The problem of searching for entities with dynamic content in real-time is addressed by different design dimensions.

On the other hand, other contributions introduce Web ontologies or semantic enrichment, mostly aimed at searching only at the level of a thing. For example, [8] presents a discovery system of RFID objects in smart spaces using a domain composed of two ontologies: one of general knowledge and another of specific domain user knowledge. In relation, [1] introduces a process and tools that allow users or applications to find connected objects that match a set of requirements and expectations. This work is based on the creation and use of semantic profiles of the connected objects, the establishment of similarities between the profiles to gather objects in groups, and a way of calculating a ranking for associating the context to incoming queries, also allowing for the selection of the most appropriate search algorithms. The limitation of this approach is that the possible outcomes are aimed at things, regardless of the sensor level, or spatial extent. In [5] Guinard introduces a vision and architecture of the Semantic WoT based on vocabularies, an abstraction of things and their high-level state, and semi-automatic sensor description generation. Searching of sensors and things is based on the high-level state of them. Both sensor and entities have well-defined semantic representations and considerations for their retrieval. The contribution of this work is its integration proposal based on Linked Sensor Data (RDF dataset of US weather stations sensor data), as well as the inclusion of semantics to the architecture of the WoT. However, the Semantic Web representation described adds complexity to the model, and the spatial-temporal context is not clearly defined in terms of the mobility of things and belonging to a space.

3. ABSTRACT MODEL OF THE WOT

We propose a WoT model, which begins with the abstraction of the real world, which mainly consists of two elements: things (tangible or intangible type), and spaces in which these things are contained or have a certain relationship. The model of the IoT involves the physical infrastructure that interconnects these two elements of the real world with the Net. Thus, there is a sensor layer in order to obtain real-time information on the properties of things in the real world, also on spaces. This information from multiple sensors is added to a nano or micro electronic device, called a data node. This by itself can get connectivity to other data nodes, or through gateway nodes, which would perform protocol conversion functions or provide connection to the Internet (see Fig 3). Our model of the WoT is comprised of five levels of abstraction involving the entire universe of elements that we consider relevant. Compared to other proposals, our model achieves completeness and balance by considering the spatial context on three levels, together with formal models of a virtual sensor and a virtual thing.

3.1 Description of the WoT Model

Given that the main and final motivation to model the WoT is publishing and establishing the information of smart things to Web services as a basis for developing future IR systems, our work proposes the representation of the WoT based on three main components of abstraction: Virtual Sensor, Virtual Thing, and Smart Space, and more important their corresponding hierarchical relationship. In contrast to [2], where there is neither a formal abstract model nor a hierarchy of elements in the proposals of searching on personalized networked spaces. However, to provide flexibility to the representation model, we add two additional spatial components: the possibility that a space is formed by sub-spaces, and further by sub-sub-spaces, and the ability to federate smart spaces in so-called Intelligent Zones. The proposed representation simplifies the lower layers, allowing to focus on a greater extend of the application and composition layers of the WoT. Therefore, the sensor node is associated in the model of WoT with an abstraction called Virtual Sensor, whose function is to allow its Web representation, composition features and high-level information to be viewed from the data collected by sensor nodes, that performs aggregation functions or information fusion. We propose to balance the model taking into account the spatial context and also creating a web abstraction of the sensor level. Some unbalanced models exist like [10] in the spatial search for smart things, where the model does not include a sensor level although does so with a spatial-oriented model. Each sensor as an abstraction model element has a URI that identifies its dynamic XML document, containing its description, properties and data. Things in the real world

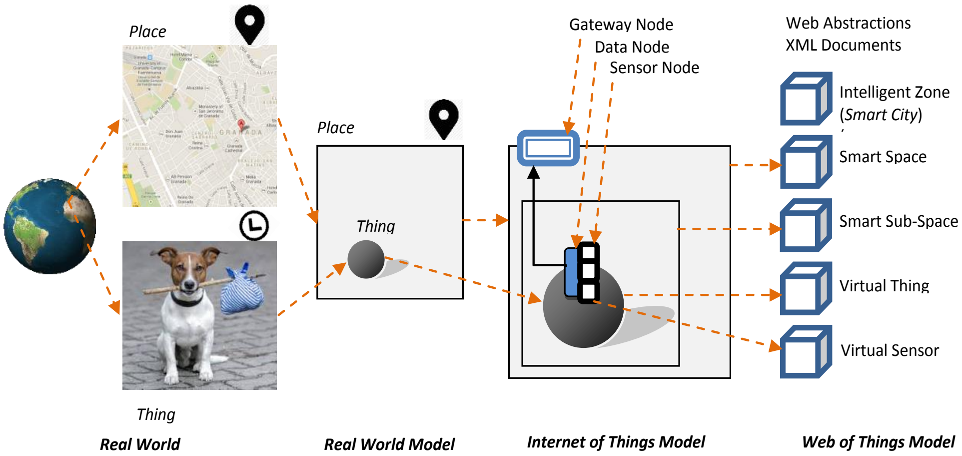

Figure 1: Proposed models of real world, IoT, WoT and their relationships

of tangible and intangible nature are modeled by the Web abstract component called Virtual Thing. Virtual things not only consolidate the information available at the virtual sensors linked to them, but also contain features, functionality or services that things through their Web abstractions can provide. Similarly, it has been decided that each element of the model, in this case the Virtual Thing, is uniquely identified by a URI, related to a dynamic XML document containing the real time information.

Virtual things, like their real counterparts, are confined in smart spaces that correspond to abstractions of places and sites of the real world that have been endowed with intelligence. Virtual things through their virtual location sensors have the potential to change not only their state, but also the smart space where they are. Thus, there will be a change in links between documents, and belonging to a place. Environments, sites and places in the real world are modeled using an abstract component called Smart Space, which condenses the characteristics of the environment in which they are located, bordering one or more virtual things. In comparison with other models, [9] proposes for state-based search of entities to have a stochastic sensor model and space considerations where the possibility of search results are limited to the level of entities, where information of sensors is given no relevance, and the constrained search for spaces are not contemplated, in our models the possible results can be in all the levels: data, sensors, things, and spaces. We propose to extend the modeling to allow a wider range, to enable searching also smart spaces that meet certain restrictions and/or contain certain things or types of sensors or data/states.

For example, a pet can be modeled as a virtual thing, which is contained somewhere within a smart city, in our model an Intelligent Zone. As illustrated in Fig. 1: the Dog is connected to Internet by means of an IoT infrastructure (Data and Gateway Nodes). Additionally, real-time information as Dog's location can be collected with the sensor node and published to the Web abstraction for further access. The model includes the possibility that a smart space can be composed of one or more sub-smart spaces, so the place where the Dog is, establish a sub-space as well as any other particular location for example a free car parking closest to the Dog. A Real-Time Search Engine for the WoT can be resolved the query to found the closest free car parking to the current Dog's location. In the search for entities in the semantic web, some proposals does not considered the sensor level nor the spatial context [8], [1], however [5] has a balanced model that includes both levels so similar to ours, even though the spatial context is not clearly defined. These works explore the semantic enrichment by means of ontologies. Our proposed Web representation model is presented in the next section. Our main motivation is feeding an IR system with a collection of dynamic documents originating from the abstract model, with the aim of retrieving the relevant ones, given a query. The representation of elements of the model can be as simple as using the associated metadata of the physical objects, or as complex as using ontologies and semantic profiles, however following the philosophy of the WoT, the reuse of web technologies, our proposal is to employ XML as a simple, structured, semantically enriched vehicle, containing the information of the items that can be retrieved in the model and, secondly, to consider advances based on XML sensor networks.

3.2 Structured WoT Representation

Each of the components proposed in each of the model layers corresponds to a XML schema, that together define a dynamic collection of XML documents for the WoT, as illustrated in Fig. 2.

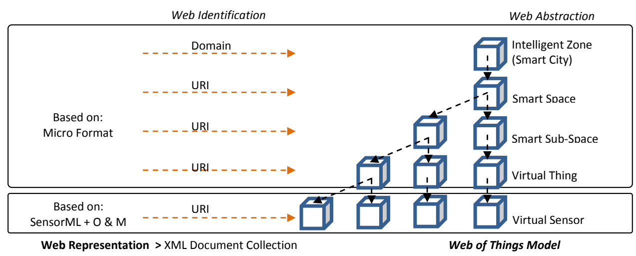

Figure 2: Dynamic collection of XML documents on the Web of Thing

The proposed virtual sensor XSD schema is composed of a group of general information tags, which contains elements of keywords, identification and classification. A group of references with contact elements, with the possibility of specifying and elements of role-based documentation specified in the XML implementation of model OM Observations and Measurements, Modeling Language SensorML. The next section of the XML schema contains the group of properties allows the characterization of the virtual sensor, the description of the sensor capabilities, the high level state, along with a membership element to associate the virtual sensor to the virtual thing sensed. Also, there is a history and events field, and Observing elements of the virtual sensor XML schema, are based scheme O&M of OpenGIS, with an element of sampling time, the time of the result, the feature of interest and the result. The Virtual Thing component should capture the information of the observed physical phenomenon, so it is proposed to use elements of the SWE scheme. The proposed scheme follows the structure of the XML representation of the virtual sensor using groups and elements: general info, references, properties and history. The property group has been enriched with an element of availability of the object, and the field of membership is associated with higher hierarchical level, indicating the smart space where the virtual object. An element location and a list of attached sensors are added. The XML schema of the Smart Space component is built based on microdata Place scheme in schema.org. This contains the same elements of the above components with a list of: virtual things, and subspaces. The Intelligent Zone follows a similar scheme with Web domain identification, and the list of smart spaces that compose it.

4. CONCLUSIONS

The WoT imposes a different dynamic to consider in the design and development of applications and systems for IR, given the change in the location of things in the physical world, the change in the collection of documents on behalf of inserting new sensors or recording new stuff, or depletion of life-time connectivity and thus removal of these documents, and in the same update real-time data collected by sensors. Most models of the WoT located in the first layer to the sensors, and from there a series of overlapping layers, according to the vision and purpose of each investigation. As core layer of most models is a abstraction layer of entities or things in the real world, with multiple alternatives for description and representation of both: non-Web, or using Web technologies like metadata, microformats, microdata, or ontologies. Given the flexibility, simplicity and use of XML standards has been selected for the construction of representation schemes for the WoT. The abstract model of the WoT, and the proposed dynamic representation take elements, and oriented considering the efforts of organizations such as the OpenGIS, W3C, ISO to standardize technologies that point to the interconnection of the real world. Our proposed model of the WoT, consider a real-world view that gives importance to the spatial context, adding relations between things and spaces. In addition the temporal context is added via elements of historical events. Future work is planned to study different conventional and semi-structured indexing information methods, focusing on XML, and real-time to assess their suitability for the WoT. The model is being used to build a Web of Things discrete event simulation with XSD schemas as inputs and to marshal an extend collection of XML documents.

Acknowledgments. This work has been funded by the Spanish Ministry of Ciencia e Innovación, and the Regional Government of Andalusia through the TIN2011-28538-C02-02 and TIC-04526 projects, respectively. Special thanks to the Iberoamerican University Association for Postgraduate AUIP, for their support to the implementation, progress and development of this work.

5. REFERENCES

[1] C. Benoit, V. Verdot, and V. Toubiana. Searching the web of things. In *Fifth IEEE International Conference on Semantic Computing Proceedings*, pages 1–8. IEEE, September 2011.

[2] Z. Ding, X. G. J. Dai, and Q. Yang. A hybrid search engine framework for the internet of things. In *Ninth Web Information Systems and Applications Conference Proceedings*, pages 57–60. IEEE, November 2012.

[3] B. M. Elahi, K. Romer, B. Ostermaier, M. Fahrmair, and W. Kellerer. Sensor ranking: A primitive for efficient content-based sensor search. In *International Conference on Information Processing in Sensor Networks Proceedings*, pages 217–228. IEEE, April 2009.

[4] M. A. Feki, F. Kawsara, M. Boussard, and L. Trappeniers. The internet of things: The next tehnological revolution. *Computer*, 46(2):24–25, February 2013.

[5] D. Guinard. *A Web of Things Application Architecture - Integrating the Real-World into the Web.* TH EidgenÃűssische Technische Hochschule ZÃijrich, Zurich, Switzerland, 2011.

[6] S. Hodges, S. Taylor, N. Villar, J. Scott, D. Bial, and P. Fischer. Prototyping connected devices for the internet of things. *Computer*, 46(2):26–84, February 2013.

[7] S. Mayer, D. Guinard, and V. Trifa. Searching in a web-based infrastructure for smart things. In *3rd International Conference on the Internet of Things Proceedings*, pages 119–126. IEEE, October 2012.

[8] D. Pfisterer, K. Romer, D. Bimschas, and et. al. Spitfire: Toward a semantic web of things. *IEEE Communications Magazine*, 49(11):40–48, November 2011.

[9] K. Romer, B. Ostermaier, F. Mattern, M. Fahrmair, and W. Kellerer. Real-time search for real-world entities: A survey. *Proceedings of the IEEE*, 98(11):1887–1902, November 2010.

[10] M. Ruta, T. D. Noia, E. D. Sciascio, F. Scioscia, and E. Tinelli. A ubiquitous knowledge-based system to enable rfid object discovery in smart environments. In *2nd International Workshop on RFID Technology - Concepts, Applications, Challenges Proceedings*, pages 87–100. IWRT, June 2008.