Differences + Triple Spaces = Active Triple Spaces

Vlad Tanasescu

Knowledge Media Institute
Open University
Walton Hall, Milton Keynes, MK7 6AA, UK
v.tanasescu@open.ac.uk, vladtn@gmail.com

Abstract

As the Semantic Web (SW) is being automatically populated with large number of RDF triples scalability issues related to wide scale reasoning occur. We believe these difficulties are due to the use of global reasoning engines, which carry all the load of collecting and handling all relevant triples, and they can be alleviated by distributing the load of the reasoning amongst the meaningful entities represented by the triples themselves. Therefore, as an application of a knowledge representation model based on Differences and on top of the triple space computing model, we introduce Active Triple Spaces, triple spaces managing triples acting as differences, i.e. processes subscribing to active queries and presenting their result as a new triple. Active triples allow the caching of the result of each new query, as well as its rapid update by subscription mechanisms.

1. Introduction

As the number of RDF triples available on the Semantic Web increases, reasoning is becoming more and more problematic: indeed complete query matching at web scale is hardly realistic, and therefore localized heuristic methods are advocated [1]. Moreover, knowledge structures principally based on taxonomies and subsumption are not always appropriate in some contexts such as geospatial systems and alternative knowledge models are needed and advocated [2].

We first describe differences as an alternative knowledge representation model, then discuss triple spaces [3] as a support platform for triple querying and reasoning before discussing the advantages of merging the two paradigms.

2. Differences

Differences, or difference spaces are a knowledge representation (KR) model inspired by philosopher Gilles Deleuze's work on the concepts of Difference and Identity, as well as by the emergence of collaborative tagging frameworks. According to Deleuze, differences pre-exist to identities as what allow us to identify things in the first place. Collaborative tagging gives absolute freedom in the elicitation of a label, or *tag*, independently of any pre-established categorisation.

Differences try to alleviate the KR difficulties originating in the use of category based ontologies for some domains. Indeed, while ontologies seem well suited to the description of scientific domains such as medicine and biology which are already semi-formal and organized by categories and part-of relationships, some communities such as geospatial scientists only accept with scepticism the exclusive usage of ontologies to describe their domains [4]. Arguments in favour of using alternative KR models include, amongst other, the inadequacy of category based reasoning to represent some aspects of reality [5], the absence of grounding of symbolic systems [6], the need of different representations of the same entity according to the context [7], as well as the difficulty to represent psychological concepts such as *affordances* in a hierarchical way [8]. Indeed, we are still waiting for ontologies to be flexible enough to match the representational complexity of the human mind. Therefore each of these KR models is used when needed, in an *ad hoc* manner, either to the exclusion of others, which creates new representational difficulties, or merely as a useful addendum to host theories with which they only superficially integrate. Differences attempt to include all of them in a new KR model.

A *difference* is a snippet of meaning. It is anything that emerges from a background and can be isolated by an *actual* process. For example particular colours, shapes, a distinct word, a sound, an action, an event etc., but only if there exist a process able to detect them. Indeed differences do not have meaning independently of the process of detecting them. For example I may not be able to distinguish as many shades of a colour as an artist, i.e. they may exist but have no meaning for me. Therefore difference-based KR is distinct from symbolic KR in that differences are not static symbols, but always the result of processes. These processes themselves operate on other differences, i.e. algorithms operating on data in an information science context, or chemical processes operating on modifications of the physical environment in a biological context, or activity processes operating on a changing environment in an embodied psychology context. Independently of the nature of the process, the result is always a difference, the production of a snippet of meaning. Therefore, independently of the domain and of the nature of the underlying process, differences can always be used, or *folded*, by other processes to form another difference, more meaning. Moreover, a difference can always be *unfolded*, i.e. bypassed to access other differences, until no process can be devised.

For example the difference space detecting some *emergency* may unfold into difference spaces detecting the presence of *danger* or particular weather hazards such as a *snow-storm*, which itself depends of the fact that a snow storm has been announced or a direct control of the snow level, depending of the process (cf. Figure 1).

Figure 1. A difference space. Each difference is not a symbol but a specific process.

As every difference is a process *reasoning* becomes similar to *execution*. Indeed, in our example the sensors immediately detect a particular combination of quantity of humidity and temperature, which is interpreted by a numerical computational process as snow level, itself transformed in the qualitative statement stating that the *snow-level is high*, which can be read by a human agent. Alternatively or in addition to sensor information, depending on the particular cognitive process, the human agent can trust the fact that a snow-storm has been announced by trusted channels. Etc. We see that reasoning occurs, but it is broken in elementary parts by the differentiation process. Of course, particular differences can be unfolded, i.e. bypassed, and the human agent can directly read the low-level sensors, or even go and check for himself, bypassing the sensors, or rather than bypassing them, replacing them with perceptual processes, i.e. his or her own senses.

Exposed as differences, meaning is transported from chemical processes detecting molecular variations to mathematical, logical, or cognitive processes. Making reasoning immediate or quasi immediate depending on the speed of the corresponding differentiation processes. To summarize the difference based KR model:

1. *There are only differences*: categories, classes, or types are not first class elements, only a type of difference.
2. *Differences process differences*: a difference can only be deduced from other differences through a process.
3. *A difference can always be unfolded*: resulting in other differences.

3. Triple Spaces

Triple Space Computing (TSC), is a new paradigm for Web service communication aiming to comply with the following basic principles of the Web:

- Stateless communication of resources
- Persistent publication of resources
- Unique identification of resources
- Non-destructive read access to resources

TSC is based on the evolution and integration of several well-known technologies such as Tuple Space computing, Shared Object Space, and Semantic Web technologies (in particular the Resource Description Framework) [9].

In practice, triple spaces are data represented as RDF triples published to a space where a client can retrieved it using a url. Triples are expressed in RDF to provide machine interoperability and link to the Semantic Web. The containing spaces provide 5 fundamental operations to clients: *write*, *read*, *take*, *waitToRead* and *waitToTake*, which manipulate triples or sets of triples (graphs). Therefore clients can *write* triples or sets of triples to a space, that other clients can then *read* in a non destructive way, or consume (*take*), whether they wait for the right data to become available or not (taking the risk of failing). These operations can be combined with query languages, to retrieve not only a triple by its address, but the set of triples returned by the query. Moreover, triple spaces support the *publish-subscribe* paradigm. This becomes particularly relevant with the addition of a query language as a client can *subscribe* to receive updates when the set of triples that satisfies a query changes.

Following our example, a triple space describing an emergency is depicted in Figure 2. We can see three spaces, *TS1*, *TS2* and *TS3*, which contain triples. *TS1* contains sensor related information. *TS2* contains some circumstantial information, while *TS3* contains triples indicating states of the world.

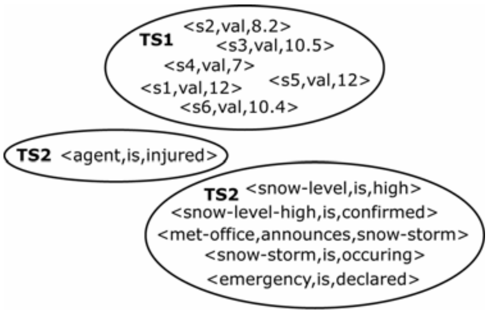

Figure 2. Triple Spaces

4. Active Triple Spaces

Triple spaces are by nature passive. Indeed, due to the constraint of persistency, they only extend the capabilities of triple-stores by allowing the aforementioned operations to active clients. Any type of reasoning implies the retrieval of all relevant triples in the accessed space or in all relevant ones, a daunting task in itself, as well as the treatment of a large number of triples if the query requires it. These triples are accessed by the client or by the space itself if it supports query languages. This approach is relevant for truly global queries such as counting the number of users that have a given characteristic. However, for most queries, where the meaning of some situation is involved, we propose to apply the differences KR paradigm to the organisation of data and the retrieval of information. Indeed, differences as processes, can make use of the publish-subscribe mechanisms of triple spaces, to allow processes to subscribe to differences which are relevant to them. By doing so, each differentiation process produces a triple which represents its state; therefore each triple, being the result of a unique differentiation process, can be said active.

As an example, following our emergency scenario, the triple spaces described before start becoming active if each sensor writes the triple that corresponds to the difference it detects, triples which are subscribed to and read by the process deciding if the snow level is high, etc. (cf. Figure 3) By following the chain of subscriptions the triple produced by each process is updated when the triples to which the process subscribes to change.

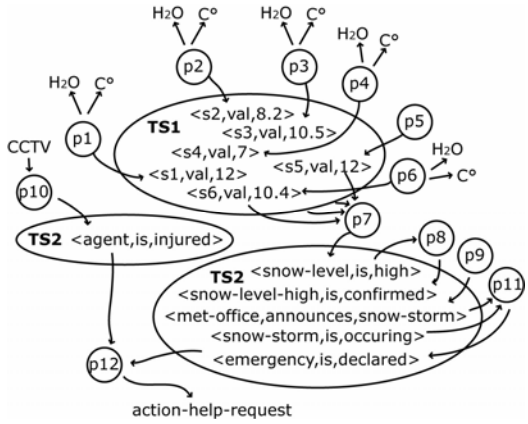

Figure 3. Active Triple Spaces.

Therefore active triples alleviate the reasoning scalability problem in two ways:

1. *Query caching as difference*: each difference, or active triple, can be thought as the result of a query. For example, *p7* determines that the snow level is high by subscribing to sensor values (*s1* to *s6*), which corresponds to the result of a given queries using the value of the sensors. Identically, further queries on sensor values, such as for example the selection of all sensors which value is bigger than a given threshold, may first a) access the sensor values, b) subscribe to the triples accessed in order to be notified of future changes, c) expose the result as a triple which can be accessed by further queries. This triple can contain, for example, links to the triples constituting the result set. This process of caching the result as a difference can be made transparent by the system: to achieve this, the system only needs to fork, after each query, a process subscribing to the related triples and reasoning on them. In this way, the first time any query is executed, the normal computation time is needed, but as it is cached as a difference, updates to the result of the process, i.e. the query, will occur by subscription and further similar queries will get the result instantly, by reading the resulting triple.
2. *Query decomposition through differences*: complex queries can make good use of queries previously cached as differences. If the system determines that a query uses a query which has previously been cached, it will access this particular triple, and not execute this part of the query again. For example requesting the regions where the snow value is higher than 4 millimetres, can involve the previously cached query of all sensors with values higher than 4mm. If another query is necessary to determine the spatial regions defined by each sensor, this query itself can be cached, and the complex query will subscribe to these two differences (the set of sensors with values higher than 4 mm, and the set of regions define by the

sensors) to determine its result, which can itself be cached in order to be reused.

3. *Lazy update of active triples*: since query caching as differences implies for each query, the creation of a process that subscribes to triples relevant to the query, the resulting activity may become computationally intensive, especially in a quickly changing environment where differences will update themselves quickly, continually firing subscription notices and executing processes. To avoid this issue, the processes acting as differences may become lazy, querying the differences they subscribed to only if the requesting difference necessitates it. For example some queries are only concerned by values at a given time, or at a given level of precision, and therefore do not need to be updated each time a subscription fires. Moreover, the system may decide to notify an active triple related to sensors and originating in the caching of a previous query that several of the sensors used have changed, only if a query involving them occur. These mechanisms implementing lazy active triples should ensure that the load on the system remains stable.

5. Conclusion and Acknolegements

Applying a knowledge representation model based on difference to triple spaces computing may allow for cognitively sounder reasoning and scalability. Indeed, very naturally, the system acquires some kind of working memory, through the caching of queries as differences, and also resolves queries in a cognitively sounder way, focusing on what is necessary to a given query instead of updating all information every time. As a test bed of these ideas, we are working on an implementation using Javaspaces.

Thanks to Carlos Pedrinaci for fruitful discussions and feedback as well as to all COSIT 2007 participants for useful feedback about differences.

6. References

[1] Fensel, D., Harmelen, F.: Unifying Reasoning and Search to Web Scale, IEEE Internet Computing, 2007.

[2] Tanasescu, V.: Spatial Semantics in Difference Spaces, COSIT, 2007.

[3] Fensel, D., "Triple-Space Computing: Semantic Web Services Based on Persistent Publication of Information," Intelligence in Communication Systems, 2004.

[4] W. Kuhn: Why Information Science needs Cognitive Semantics - and what it has to offer in return, 2003.

[5] A. Wierzbicka: “Apples” Are Not a “Kind of Fruit”: The Semantics of Human Categorization American Ethnologist, JSTOR, 11, 313—328, 1984.

[6] P. Gärdenfors: How to Make the Semantic Web More Semantic Formal Ontology in Information Systems, Proceedings of the Third International Conference (FOIS), 17-34, 2004.

[7] C. Vangenot, C. Parent and S. Spaccapietra: Modeling and manipulating multiple representations of spatial data Proc. of the Symposium on Geospatial Theory, Processing and Applications, 2002.

[8] J. Gibson: The Ecological Approach to Visual Perception, Lawrence Erlbaum Associates, 1979, p.42-43

[9] R. Krummenacher et al., Triple Space Computing D1.2 v1.1, http://tsc.deri.at/deliverables/D12v11.html, 2006.