Effective and Efficient On-line Communication

Dieter Fensel, Birgit Leiter, Stefan Thaler, Andreas Thalhammer and Ioan Toma
STI Innsbruck, University of Innsbruck, Austria
firstname.lastname@sti2.at

Abstract – **As Web, Web2.0 and semantic channels are constantly growing we are facing new challenge in on-line communication. This paper present a new methodology based on distinguishing and explicitly interweaving content and communication as a central means for achieving content reusability, and thereby scalability over various heterogeneous channels. Our approach enables smaller organizations to perform effective and efficient on-line communication and thus to increase their profitability.**

Keywords: Value management, On-line communication, Social Media, Web2.0

I. Introduction

Fax, phone, and later the Internet, have radically changed our communication possibilities. More and more communication has been freed from the geographic barriers that formerly limited their speed and expansion. Now, it is (in principle) possible to instantly communicate with a large portion of the entire human population. Nevertheless, new means also generate new challenges.

Organizations of all sizes, commercial and not-for-profit, regularly face the challenge of communicating with their stakeholders using a multiplicity of channels, e.g. websites, videos, PR activities, events, email, forums, online presentations, social media, mobile applications, and recently structured data. The social media revolution has made this job much more complicated, because:

- the *number of channels* has grown exponentially,
- the communication has changed from a mostly unilateral "push" mode (one speaker, many listeners) to an increasingly fully *bilateral communication*, where individual stakeholders (e.g. customers) expect one-to-one communication with the organization, and the expected speed of reaction is shrunk to almost real-time, and
- the *contents of communication becomes more and more granular* and increasingly dependent on the identity of the receiver and the context of the communication.

Organizations need an integrated solution that provides management and execution of communication goals in a mostly automated fashion, with costs equivalent to mass-media communication, along with the granularity of individual experts, and at the pace of real-time social media. We are aiming to mechanize important aspects of these tasks, allowing scalable, cost-sensitive, and effective communication for small-or-medium sized business units and comparable organizations for which information dissemination is essential but resources are significantly limited. Additionally, it may also help intermediaries such as marketing agencies to extend their business scope by increasing the cost-effective ratio.

We identify yield, brand, and reputation management as three major aspects around which communication may be centered. They can be identified as variations of general value management differing in their short-versus-long-term orientation towards commercial goals, as well as in their overall connection to financial value orientation.

The reminder of this paper is organized as follows: Section II presents our approach. Section III discusses related work and directions for future research activities. Finally, conclusions are provided in Section IV.

II. Approach

We start this section by introducing the underlying idea and major structure of our approach. We then discuss our role model, tool support, and finally, we sketch some applications of our approach.

A. Separating Content and Channel to enable various dimensions of reuse in transactional communication

The core idea of our approach is to introduce a layer on top of the various Internet based communication channels that is domain specific and ***not*** channel specific.[1] So one has:

- *information models*, that define the type of information items in a domain;
- a *channel model* (or communication model), that describes the various channels, the interaction pattern, and their target groups;
- *mappings* of information items to channels through weavers; and finally,
- a library of *implemented wrappers* for actual channel instances.

What is essential is to *distinguish* the communication or channel model from the conceptual descriptions of the information.[2] Our approach requires the creation of a communication model (i.e., an increasingly complete model of channels), and knowledge models for each vertical (such as research projects, research institutes, associations, hotels, restaurants, tourist events, medical doctors, etc.), and finally linking the knowledge model with the communication model

[1] See also as an excellent presentation on this idea: http://www.slideshare.net/reduxd/beyond-the-polar-bear

[2] In analogy to style sheets that separate the contents from its presentation.

Dieter Fensel, Birgit Leiter, Stefan Thaler, Andreas Thalhammer and Ioan Toma: Effective and Efficient On-line Communication. In Proceedings of the 11th International Workshop on Web Semantics and Information Processing (WebS 2012), co-located with DEXA 2012 - 23rd International Conference on Database and Expert Systems Applications, Vienna, Austria, September 3-7, 2012.

through a weaver that weaves concepts with channels. Currently, all commercially available solutions are only channel centric and do not provide any built-in support for what needs to be disseminated or where to disseminate what piece. In our approach, a knowledge-model is built and explicitly linked with the channel model enabling reusability of the same information element for various channels. Reusability is also achieved by developing the information as domain ontology for a certain vertical area such as tourism, accommodations, gastronomy, etc. These elements of reusability deliver the major contribution to the scalability of our approach.

1) Information Model

An information model is an ontology [5] that describes the information items that are used in typical communication acts in a certain domain. We do not aim to model a domain as such, through a very deep model that allows arbitrary (transactional) applications. We can rather focus on the major and typical information items that are used in the on-line dissemination and communication processes. Therefore, the size of these ontologies in our case studies, were moderate (around 100 concepts and properties), and many of these concepts and properties could be reused between different use cases. After defining an informal model, we formalized this ontology (see [2] for more details) in a simple sublanguage of OWL-2, since we foresee little need for reasoning about it. We model structured information items as concepts and non-structured ones as properties, i.e., we assume simple non-structured values for properties. Actually, we separated a generic ontology common to all three use cases and specific refinements.

In general, our models are rather simplistic, but this reflects the fact that we do ***not*** model a domain with all its complexity but rather the information chunks that are disseminated about it.

In an intermediate phase of our journey, we also tried to directly use some LOD vocabularies to model these ontologies, including Dublin Core, FOAF, schema.org, and GoodRelations. However, we took some important lessons from this exercise: (1) our domain models were suddenly unintelligible to the domain experts, (2) it was extremely hard to decide which term to take from which vocabulary. The terminologies were either redundant or terms had different but overlapping coverage and (3) suddenly, we had to deal with a large number of properties for which we had never asked.

We draw an important conclusion from this**:** *LOD vocabularies are not means to describe our content models, i.e. they were not really useful for deriving domain models.* We decided it would be better to *interpret them as channels*. That is, we model our information items in a Domain Ontology that is understandable by the domain experts. We then provide mappings, though our weaver, that export and/or import information to or from terms of various LOD vocabularies. *For us, LOD vocabularies are means to disseminate and share information and not means to model information*. By weaving our models with LOD vocabularies we see a gain in broadening our range of communication through them.

2) Channel Model

In our approach we see a channel as means of exchanging information in the on-line space. There is a close relationship between URIs and channels as each URI can be used as a channel to spread or access information. However, not each channel directly refers to an URI. For example, Facebook provides around forty different methods of spreading information not distinguished by a URI. Additionally, individual information items spread through Facebook are not distinguished by URIs. In general, a channel can be interpreted as a “place” where one can find or leave information, whether it is unanimously referred by a URI or addressed through a service. However, even this is not broad enough. As described previously, a channel can also be the URI of a vocabulary (or the formalisms such as RDFa or microformats) that are used to publish the information. Through use of this URI, only humans or software agents that “speak” this dialect are able to access this information. Here, the communication channel cannot be interpreted as a place, but rather as a way to express or refer to the information. In the following, we want to distinguish channels by the communication mode they support.

Communication is based on the broadcasting of information. Therefore, we define the first category of our channel classification system as channels used for ***broadcasting***. Here we make a distinction between the publication of mostly ***static information*** and ***dynamic contents*** that express the timeliness of an information item.

Websites are example of channels used for broadcasting of static information. Information that reflects the structure of the contents is provided through websites and they offer a smooth way for users to access this content. Broadcasting of dynamic content is done via a multitude of channels including blogs, RSS feeds, Twitter, email, email lists and chatting.

Sharing is the first form of cooperation. Explicit ***collaboration*** through a shared information space is the next cooperation category we have identified. Collaboration between individuals leads to groups of people actively organizing their communication and cooperation. Social networking sites that support ***groups*** *of people* in their information needs are instances of this next category we have identified. Obviously, the boundaries between these categories are fluid and many channel providers try intentionally to establish services covering several of them. Still, it is often possible to identify a major category for them, often based on the major usage patterns of their users. An important approach to broaden the scope of a dissemination activity is to add ***machine-processable semantics*** to the information, this being our final channel category. With this approach, search and aggregation engines can provide a much better service in finding and retrieving this information.

3) Weaver

The central element of our approach is the separation of content and communication channels. This allows reuse of the

Dieter Fensel, Birgit Leiter, Stefan Thaler, Andreas Thalhammer and Ioan Toma: Effective and Efficient On-line Communication. In Proceedings of the 11th International Workshop on Web Semantics and Information Processing (WebS 2012), co-located with DEXA 2012 - 23rd International Conference on Database and Expert Systems Applications, Vienna, Austria, September 3-7, 2012.

same content for various dissemination means. Through this reuse, we want to achieve scalability of multi-channel communication. The explicit modeling of content independent from specific channels also adds a second element of reuse: Similar agents (i.e., organizations active in the same domain) can reuse significant parts of such an information model.

Separating content from channels also requires the explicit alignment of both. This is achieved through a weaver. Formally, a weaver is a set of tuples of nine elements:

1. An *information item*: As discussed in Section 2, it defines an information category that should be disseminated through various channels.
2. An *editor*: The editor defines the agent that is responsible for providing the content of an information item.
3. An *editor interaction protocol*: This defines the interaction protocol governing how an editor collects the content.

Elements 1 to 3 are about the content. They define the actual categories, the agent responsible for them, and the process of interacting with this agent. Elements 4 to 9 are about the dissemination of these items.

4. An *information type*: We make a distinction between three types of content: an instance of a concept, a set of instances of a concept (i.e., an extensional definition of the concept), and a concept description (i.e., an intensional definition of a concept).
5. A *processing rule*: These rules govern how the content is processed to fit a channel. Often only a subset of the overall information item fits a certain channel.
6. A *channel*: The media that is used to disseminate the information.
7. *Scheduling information*: Information on how often and in which intervals the dissemination will be performed which includes temporal constrains over multi-channel disseminations.
8. An *executor*: It determines which agent or process is performing the update of a channel. Such an agent can be a human or a software solution.
9. An *executor interaction protocol*: It governs the interaction protocol defining how an executer receives its content.

A weaver is basically a large collection of tables specifying what is disseminated by whom to where. Interaction protocols, rules, and constraints further guide this process. Such a manual is of extreme importance to manage the on-line communication process. Obviously, it determines the need to implement and mechanize essential aspects of it, improving its scalability. However, a major step is to structure the process towards a mechanizable routine.

4) *A Spiral as Process Model*

We have adopted the *transactional model* of communication and its underlying premise that individuals are simultaneously engaging in sending and receiving messages (cf. [1]).[3]

Consequently, our approach not only disseminates information, but also deals with the aggregation of feedback and impact by simply going through the dissemination chain in the opposite direction, collecting responses in the various channels and integrating them under the appropriate knowledge item. We not only talk, we also listen to responses. And we do not get these responses scattered over multiple channels. Instead, they are aggregated and presented at the level of the domain specific concept to which they refer. Finally, our approach can also aggregate information from channels without prior publication activity.

Therefore, a holistic methodology for supporting communication must support the following subtasks that basically form a circle or spiral: (1) *design* of an information item, (2) *dissemination* of the information item using suitable channels and places, (3) *observation* of communication acts and (4) *measure*, *analysis*, and *aggregation* of the information published

These activities form a circle that we call the *life cycle model of communication*. Reactive communication starts with the observation task; active communication starts with the design phase. Obviously, these tasks can and should be parallelized once initialized.

B. *Role Model*

Editors are assigned to information items, responsible for producing or collecting their information instances. In general, an editor can also be the executor, directly publishing the information. However, expertise in a certain information domain may not necessarily correspond to technical expertise and even if it does, it may not be a very efficient way of distributing labor. Only if a fully automated and easy to use software solution is provided can this model make sense. Otherwise, a person with more technical skills often helps in disseminating the information. Again, an interaction protocol has to be defined for interacting with this person. Recursively, some of his tasks may be manual (importing contents into a content management system such as Drupal) and some can be fully mechanized (like producing a feed and a tweet automatically for certain information items introduced into Drupal). We identified five different roles involved in this process: (1) the *communication manager* that actively reads and writes information in the multi-channel space and manages the overall communication process, (2) the *quality manager* that routinely checks the outcome of the process and the impact that is achieved through it, (3) the *editors* that provide information that should be disseminated or that infer actives from information provided by others, (4) the *web manager (executor)* is an expert in web technology who is able to publish information with current web technology including CMSs such as Drupal, email lists and Web 2.0

[3] Or in Web2.0 terms, users are *prosumers*, i.e., consumer and producer of information.

services and (4) the *repository manager (executor)*, an expert in semantic web technology in terms of syntax, implementations via repositories, and various vocabularies used to publish this information.

C. Technologies

Obviously there is an important need for methods and integrated tools that cover the multi-channel bi-directional aspects of value management and provide highly scalable and effective solutions. We are developing a communication platform and methodology for providing this based on Semantic Technologies, Human Computation, and Social Media Analysis.

From the Semantic Technologies field we use content management tools such as Drupal 7 to include RDFa, microdata, and microformats in the web documents. The data will also be exported from Drupal into OWLIM[4] to support direct RDF and SPARQL (see Figure 1). Either the editors, or alternatively, the dissemination manger enrich the content for on-line presentation by adding links and tags to the presented information. For this purpose, tools such *KIM*[5], *OpenCalais Web Service*[6] or *Zemanta*[7] can be used.

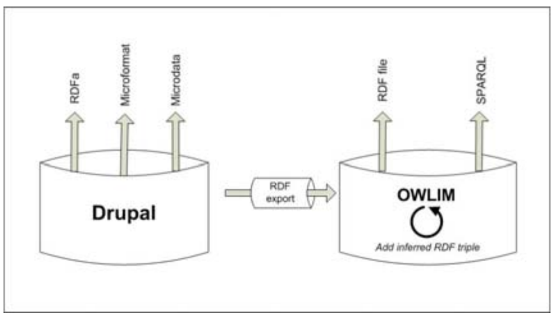

Figure 1. Technical means to publish Semantic Data

From Human Computation we adopt the principles of integrating humans for tasks that cannot be fully automized. This includes crowd-sourcing[8] where we are using methods to define incentives for large user communities to provide content needed for communication strategy through a distributed community effort.[9]

From Social Media Analysis field we adopt techniques and tools that enable us to study properties of large, complex social networks. For reputation management, we require support for impact analysis, multi-triggered bi-directional communication, and standing queries over streams tools. Tools such as *Boardreader*[10], *Radian6*[11], *Social Media Dashboard*[12] or *TweetDeck*[13] are just some of possible.

D. Use Cases

We developed and applied our approach in three major case studies: (1) the *European Semantic Web Conference Series (ESWC)*[14] which bring together researchers and practitioners dealing with different aspects of semantics on the web, (2) the *PlanetData* project[15] aiming to create a durable community made up of academic and industrial partners working on large-scale data management and (3) the *Semantic Technology Institute (STI) International* research association[16], a global network engaging in research, education, innovation and commercialization activities on semantic technologies.

Around 80% of the information items of ESWC, PlanetData, and STI International are interchangeable due to some simple renaming (e.g., core and associate partner versus partner and member). This is excellent news and a hint for scalability especially given the fact that we talk about a research *project* and a research *association*.

Based on our approach ESWC, PlanetData, and STI International are now managing their on-line appearance. In total, we have identified around *five hundred* different semantic and non-semantic channels in these case studies that are used to disseminate elements of the information model. Obviously, such a bandwidth requires a structured and mechanized approach. Based on our approach, around 300 concepts and properties, 500 channels, i.e., more than 100,000 potential content-to-channel mappings are run efficiently by a very small dissemination team.

III. Related and Future Work

A. Related Work

We see two specifically related areas: *Ontology-based content management systems (CMSs) for websites* and *Semantic matchmaking of senders and receivers of content*.

The field of *semantics-based or enhanced CMSs* has already been quite thoroughly explored. As an early pioneer for ontology-based website management, OntoWebber system [7] introduces an integration layer that adapts to different data sources. This is related to our weaver concept but, in contrast, the weaver adapts to different channels rather than to different information sources. [6] introduces “The Rhizomer Semantic Content Management System” which integrates services with metadata browsing, editing, and uploading, continuing their earlier work on the Knowledge Web portal. [4] proposes a Linked Data extension for Drupal that enables content annotation with RDFa and provides a SPARQL endpoint. As

[4] http://www.ontotext.com/owlim
[5] http://www.ontotext.com/kim
[6] http://www.opencalais.com/
[7] http://www.zemanta.com/
[8] http://en.wikipedia.org/wiki/Crowd_sourcing
[9] See the results of the Insemtives project at http://www.insemtives.eu/
[10] http://www.boardreader.com/
[11] http://www.radian6.com/
[12] http://hootsuite.com/
[13] http://www.tweetdeck.com/
[14] http://eswc-conferences.org/
[15] http://www.planet-data.eu/
[16] http://www.sti2.org/

Dieter Fensel, Birgit Leiter, Stefan Thaler, Andreas Thalhammer and Ioan Toma: Effective and Efficient On-line Communication. In Proceedings of the 11th International Workshop on Web Semantics and Information Processing (WebS 2012), co-located with DEXA 2012 - 23rd International Conference on Database and Expert Systems Applications, Vienna, Austria, September 3-7, 2012.

reported in [3], BBC's World Cup 2010 site[17] is based on semantic repositories that enable the publishing of metadata about content rather than publishing the content itself. While the data input is fixed, different schemas for the output are defined. Finally, the European project Interactive Knowledge Stack (IKS)[18] focuses on porting semantic technologies to CMS software solutions.

Compare to these approaches, our approach supports the overall management of content dissemination in a multi-channel and bi-directional communication setting. Further, we augment the technical approach with a methodology and the approach of using vertical domain models, which are shared and reused in a vertical area instead of being used for a single application only.

Semi-automatic matchmaking is a well-studied field. Obviously we can only select a small sample of approaches in this area, which focus on matchmaking in regard to content. [9] presents a selective information dissemination system that is based on semantic relations. In their paper, the terms in user profiles and terms in documents are matched through semantic relations that are defined using a thesaurus. The system introduced in [11] uses RDF, OWL, and RSS to introduce an efficient publish/subscribe mechanism that includes an event matching algorithm based on graph matching. Our approach, in contrast, matches information items to channels rather than events to users. Also, instead of graph matching, we use predefined weavers for channel selection. While [12] uses fuzzy linguistic modeling and NLP techniques for semiautomatic thesaurus generation and performs a matching based on statistical analysis, we use semantics to manually define the connections between information items and the channels. Since we aim for high precision and professionalism in on-line communication, we see little use for statistical based semantic methods. We want to allow the user to abstract from the channel level to the content level, but we see the need for human involvement in defining the content-channel mapping and at the content level. Fortunately, a large number of such web analytical toolkits already exist, [8] lists a large number of them that cover parts of these tasks. However, there is an important need for methods and integrated tools that cover the multi-channel bi-directional aspects of value management and provide highly scalable and effective solutions.

B. *Future Work*

Introducing a semantic layer on top of communication channels will enable reusabilty. However, this combination of research fields opens a broad variety of new challenges yet to be solved. We list below some possible research directions for our approach: (1) *modeling and interweaving feedback* with content items previously published, (2) *modeling target groups* and *engaging support with target groups* in different platforms, (3) *adapting content* of information items by converting them into different formats and/or automatic transforming their content to fit the target group, (3) *using crowd sourcing* to smartly enhance sentiment analysis, natural language processing, and translating algorithms with human intelligence where needed, (4) implement *mechanisms of trust*, and *quality management*, (5) *quantification of social values* as well as *quantification of brand and reputation* and (6) *enrichment of yield management* which is currently based on statistical analysis on different parameters, such as pricing, capacity, and demand (cf. [10]) to take also into account social relationships.

[17] http://www.bbc.co.uk/worldcup
[18] http://iks-project.eu/

IV. CONCLUSIONS

The following core features characterize our approach:

- We use ontologies to model content in order to have a representation layer independent from the communication channel. The alignment of content and channel is achieved through a weaver that aligns ontological items with channels.
- These ontologies are not case-specific, but model a certain vertical domain.
- Our approach is bi-directional.
- We support in an integrated fashion, the dissemination via traditional web channels, Web 2.0, and semantic based channels, using various formats and vocabularies.

Based on our approach, ESWC, PlanetData, and STI International are now managing their on-line appearance. Currently, we are performing additional case studies looking into using our approach in the dissemination of other research projects and associations as well as applying it to more commercial areas such as eTourism.

REFERENCES

[1] D. C. Barlund: A transactional mode of communication. In C. D. Mortensen (eds.), *Communication Theory* (2nd ed.), New Jersey Transaction, 2008.

[2] T. Bauereiß, B. Leiter, and D. Fensel: Effective and Efficient On-lin Communication, STI TECHNICAL REPORT 2011-12-01, 2011. http://www.sti-innsbruck.at/TR/OnlineCommunication

[3] B. Bishop, A. Kiryakov, D. Ognyanoff, I. Peikov, Z. Tashev, and R. Velkov: OWLIM: A family of scalable semantic repositories, Technical report, 2010.

[4] S. Corlosquet, R. Delbru, T. Clark, A. Polleres, and S. Decker: Produce and consume linked data with Drupal, *Springer Constraints Journal*, 1380:763-778, 2009.

[5] D. Fensel: Ontologies: Dynamic Networks of Formally Represented Meaning. 2001. http://sw-portal.deri.at/papers/publications/network.pdf

[6] R. Garcia, J. M. Gimeno, F. Perdrix, R. Gil, and M. Oliva: The Rhizomer Semantic Content Management System. In *Proceedings of the 1st World Summit on The Knowledge Society*: Emerging Technologies and Information Systems for the Knowledge Society, *WSKS'08*, pages 385-394, Springer, 2008.

[7] Y. Jin, S. Decker, and G. Wiederhold. Ontowebber: Modeldriven ontology-based website management. *In Proceedings of the Semantic Web Working Symposium (SWWS)*, Stanford University, 2001.

[8] H. Kasper, M. Dausinger, H. Kett, and T. Renner: Marktstudie Social Media Monitoring Tools, Frauenhofer Verlag, 2010.

[9] I.-E. Katzagiannaki and D. Plexousakis: Information dissemination based on semantic relations. In *CAiSE Short Paper Proceedings'03*, 2003.

[10] S. E: Kimes: A strategic approach to yield management. In A. Ingold, U. McMahon-Battie, and I. Yeoman (eds.), *Yield Management: Strategies for the Service Industry*, International Thomson Business Press, 2000,

Dieter Fensel, Birgit Leiter, Stefan Thaler, Andreas Thalhammer and Ioan Toma: Effective and Efficient On-line Communication. In Proceedings of the 11th International Workshop on Web Semantics and Information Processing (WebS 2012), co-located with DEXA 2012 - 23rd International Conference on Database and Expert Systems Applications, Vienna, Austria, September 3-7, 2012.

[11] J. Ma, G. Xu, J. L. Wang, and T. Huang: A semantic publish/subscribe system for selective dissemination of the RSS documents. In *Proceedings of the Fifth International Conference on Grid and Cooperative Computing (GCC'06)*, pp. 432-439, 2006.

[12] J. M. Morales-del-Castillo, R. Pedraza-Jimenez, A. A. Ruiz, E. Peis, and E. Herrera-Viedma: A semantic Model of Selective Dissemination of Information for Digital Libraries, *Information Technology and Libraries*, 28(1):21-31, 2009.

Dieter Fensel, Birgit Leiter, Stefan Thaler, Andreas Thalhammer and Ioan Toma: Effective and Efficient On-line Communication. In Proceedings of the 11th International Workshop on Web Semantics and Information Processing (WebS 2012), co-located with DEXA 2012 - 23rd International Conference on Database and Expert Systems Applications, Vienna, Austria, September 3-7, 2012.