USING POPULAR MICROBLOGGING IN SUBJECT-ENVIRONMENT INTERACTION MODEL OF CYBERINFRASTRUCTURE-ASSISTED ENTERPRISES

Bilodray Yuriy V., Cand. Sc. Prof. Mesyura, Volodymyr I.
Vinnytsia National Technical University
Hmelnytske Shose, Vinnytsia 21021, Ukraine

Summary

The study is to evaluate possibility of using highly accessible and scalable tools and techniques of social media to serve as a base of the new generation of enterprise infrastructures, called CyberInfrastructures. The concept of CyberInfrastructure-Based/Assisted Enterprises is a further development of "IT-enabled", "Internet-based" and "the cloud" notions into the field of enterprising. It proposed to exploit societal by nature, highly interconnected Internet systems and apply the Output-Input fusing paradigm, - governing production processes by the both, personal demands of end users/their environments and specific production needs. The Subject-Environment Interaction Model (SEI) additionally suggests embedding of intelligence into the network agents to create a distributed information/decision system together with enterprise databases. The topical question is how to build such systems avoiding high cost and complexity. This paper proposes a solution that uses popular microblogging (Twitter, Jaiku, etc.) as a global communication environment common for agents - humans and third-party systems wherein they setup specific rules and exchange messages in controlled languages, and has all features pertinent to the concept. This lays ground for creating practitioner-ready communication/metadata spaces to layer upon existent enterprise eco-systems as one of realization variants to enable "long-term evolution" to fully enabled CyberInfrastructure-Based/Assisted Enterprises.

Keywords: social media, microblogging, ESSP, cyberinfrastructure-based/assisted infrastructures, O-I fusing, Subject-Interaction Model, SEI, intelligent agents, semantic, sequential filters

1. CyberInfrastructure-Based/Assisted Enterprises and Subject-Environment Interaction Model

The term of CyberInfrastructure-Based/Assisted Enterprise was coined by Prof. Cheng Hsu, embracing results of development in sub-areas of his theory called Digital Connection Scaling under the umbrella of Service Science research that helps explain the dynamics of services in interconnected by digital communications economy and make use of the effect of "population-wide connected value co-creation among persons and organizations" [1].

In work [2] that is the base for all further considerations in this paper, Hsu describes currently happening evolution of "IT-enabled", "Internet-based" and "the cloud" concepts into the emerging notion of CyberInfrastructure-Based/Assisted Enterprise and a "thought model" for the design, administration, and processing of the new cyber-infrastructure. According to Hsu, the main driving force of this evolution is a new mode of micro-economical production function: Output-Input Fusing. Paraphrasing shorter statements of work [2] for conciseness and context, Output-Input Fusing (O-I Fusing) supposes governing production processes by the both, personal demands of end users/their environments and specific production needs together, inducing co-productive recombinant and reconciling processes. It is different from Output-pulling (user/demand-dominate) and Input-pushing (provider/supply-dominate) paradigms considered historical norm, and it has potential to make services person-centered in distinction to driven by classic I-pushing unitarily and massively run processes.

As Hsu [2] indicated, Cyber-infrastructure is societal in nature; exactly this fact allows embedding the effect of O-I Fusing. An enterprise cannot confine itself to proprietary technologies and rather should try to leverage what is available in society and be open to different technologies and proprietary controls. Economy of scale is achieved by concurrent running many co-productive processes that may be considered technically as virtual configurations tuned for specific, even very different, services/productions on shared connected resources of a cyberinfrastructure. Such approach may work well especially for services. The important fact Hsu highlights in [2] is that Internet Service Providers (ISP), Application Service Providers (ASP) and such well-known e-commerce/e-business models as Exchange (forums, blogs, other peer-to-peer sites) and Marketplace (e-shops, particularly Android Market, iTunes, etc.) are examples of already functioning kinds of CyberInfrastructures.

The Subject-Environment Interaction Model (SEI) is a supplementary part of the concept stated in [2]. It suggests embedding of intelligence into the network agents (operation systems, devices and sensors) to create a distributed adaptive information/decision system together with enterprise's service/production control model and database. The system feeds real-time sensor data into the enterprise database on the one hand, and provide for

adaptive control of the agents based on the latest information in the database, on another hand. This behavior has common traits with proposed by Trompenaars & Hampden, cited in [3]:

The relationship between centralization and decentralization should not be a matter of "either/or" or even "and/and", but instead should be viewed as "through-through" relationship. The centralized body accumulates knowledge through the input of local experts and disseminates that knowledge through the organization. In turn, the local bodies improve their activity through leveraging the global body of knowledge.

2. Social media as communication-centric environment for enterprises

Recent years, significant effort is put to understand prospective usage scenarios of social media for conducting, managing and supporting directed collective activities, particularly for supporting enterprise activities. Classic enterprise ecosystems, based on sets of siloed or not well connected systems, often are locking valuable data inside - that is a well established fact. Another basic aspect of such ecosystems is that the manipulation with formal data stores is usually separated from spaces of "live" collaboration (e-mail, instant messaging and social media tools). We call such approach "system-centric".

Social media tools (Facebook, Twitter, etc.) tend to expose their APIs to third party social, technical and management systems. That allows information exchange between people and third party systems/sensors (for ex., simplest event notifications), with role of the social media as information routers. Such communication approach can be called "communication-centric". Figure 1 demonstrates the two approaches along.
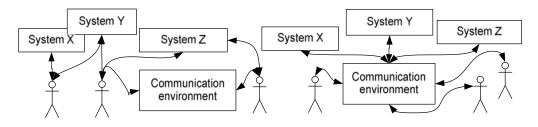

Fig.1. Classic system-centric vs. new communication-centric approach with social media

One can imagine social media as an enterprise environment where people and systems exchange messages in peer-to-peer mode getting information and requesting each other to perform some actions. With specifically chosen communication protocols enabling mutually interpretable and actionable messages people and systems exchange, this can turn into a meta-information environment, the "digital nervous system" mentioned in [2]. Usage of existent popular social media tools means getting already available highly accessible and scalable infrastructure penetrating many industries and activity aspects, possessing variety of free and proprietary applications (web, mobile social media clients, sms, etc.) as well as reaching many linked or having potential to link to third-party systems-satellites of social media.

3. Realizing SEI model and building metadata systems on microblogging

According to McAfee [4], social media display apparent qualities of complex adaptive systems (CAS) that combine relevant information into high-level patterns naturally due to effect of the emergence; they are Emergent Software Social Platforms (ESSP). It is possible to filter meaningful information, visualize its patterns, give users control over the visualization and allow information to trigger actions. Managing directed group activities, such as managing projects, requires capability to manipulate data at aggregated and at once at precise fact-atomic levels in nearly-real-time mode. Progress of such systems is impeded by complexity of NLP/IE processing, semantic representation and querying semantic data.

Microblogging (Twitter, Jaiku, etc.) is one of the most used tools and techniques of social media being in essence a platform that allows exchange short sms-like messages suitable for transferring brief actionable, historical information and directives, with optional references to other multimedia documents. Such method proved to be very effective for rapid distributing information world-wide. It possesses techniques of manual pre-annotation that easies semantic analyses, for example, #hash-tags, "nanoformats", like format of Twitter Data [5], and functions applying the practice of "folksonomies" or controlled vocabularies. Use of such pre-annotations is a noticeable trend among user communities.

With help of microblogging, it is possible to gather specific service/production information or information of general character: process performance, logistics, problems and solutions, information for further calculation, for example, EVM analysis of Project Management, sensor data, task assignments to humans and systems, and also use commands to achieve desirable system's feedback. As the input of the system there should be messages qualified for the interpretation, better in some form of controlled language, and the system must have base data (for ex., ontology as proposed in [2]) that describe specific substances for the interpretation. Examples of possible input messages:

Work package WP2 has EV=1.1 and AC = 0.9;

Project XYZ has workpackage WP1 and workpackage WP2. WP2 is half-complete;

An accident on crossing of Street X and Street Y. Height of tidal wave is 1.5m near the east coast.

Line #5 needs 18 blank kits "Type A" by 2 PM. Generate Resource Utilization for Project N.

Figure 2 shows a schema of connecting a metadata system to a microblogging platform like Twitter. The actors are: the users (humans), the users (third-party systems), the user-Information Router system (IR), and the metadata system itself ("system" hereafter). All of them are subjects of peer-to-peer two-directional communication, and multi-directional in the general context of the information interchange. All the elements are registered in the standard way as the users of the microblogging platform and system. IR is a part of the system, but is represented as a separate element for demonstrating its role as a central point of connection of the microblogging environment and the system. IM also can publish own messages, which can be read by other users. The aggregate of the connected in such a way elements form communication space of the metadata system. IR plays role of "global node" in terms of [2].
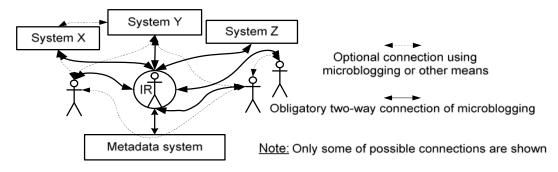

Fig.2. Connecting a microblogging platform to the metadata system

It is important to note, that the schema of Figure 2 was found similar to the schema, Figure 6 "Concurrent Systems Integrator Architecture" of paper "The Metadatabase for Manufacturing Systems Integration" [6] by Hsu et al. With current state of technology and due to the fact that microblogging clients (network agents) are pieces of software set up on computers or mobile devices with operation systems, it is possible to realize SEI model [2], place such network agents into any destination with Internet or phone/SMS wireless coverage and embed into machinery. Needless to say, such clients have capabilities to act as intelligent network agents. The system of such kind is CAS with adaptive, self-optimizing sub-optimal behavior and high robustness under condition of proper configuration with more than one global node.

Utilizing currently available Natural Language Processing (NLP) and Information Extraction (IE) toolsets like GATE; semantic stores like Sesame, Ontotext OWLIM, Jena, etc., it is possible to represent information from microblogging in RDF format with good visibility of interrelations and possibility to infer hidden relations. In such a case the NLP/IE realization can be simpler and quality better than for not pre-annotated large text corpora usually considered sources of annotation. Gathering and parsing short text messages is not new application. A bright example is GyroQ application by Gyronix with MindReader add-in by ActionOwner for managing projects in GTD style using mindmapping application MindManager by Mindjet. It is also somewhat similar to the model of Global Query System (CQS) with its syntax-free user interface described in [6] by Hsu et al.
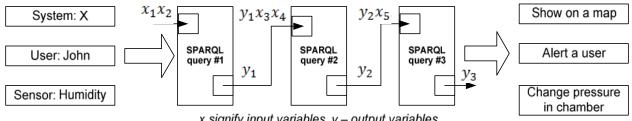
In regard to Twitter Data microblogging pre-annotation format compatible with RDF, further semantic transformation can be based on CLIE (Controlled Language Information Extraction) [7] with an extension that allows a dynamic constituent of the controlled language phrases, basing on predicates of Twitter Data/RDF

statements, that makes the controlled language vocabulary restricted only by vocabularies specific for particular applied usage. Another work in progress is dedicated to this topic.

Virtual process configurations in metadata systems based on microblogging

For constructing controlled visualizations of information patterns mentioned in the previous section, the sequential approach of progressive information filtering is a proven choice. It is similar to the principle used in software for control and filtering video- and audio-streams and, what is a close analogy, similar to GTD project management software by Gyronix, named Result Manager, where users create sequential filters chained as mindmapping topics, using logical expressions.

The patterns can be considered "information dashboards". In the new system, the dashboards comprise sets of sequential filters that are filtering information in semantic stores. The filters in essence are sets of SPARQL queries to RDF graphs. Simplicity of use is primary, enterprise personnel and users should be able to construct the filters even without semantic and SPARQL knowledge in a simple visualized manner (e.g., drag-&-drop). The idea lies in employing automated generation of SPARQL queries to RDF graphs using sequential approach as analogy of the method of Mokin & Yasholt [8] for SQL queries to RDBMS: ontological classes are used instead of the tables, and input and output variables are not fields of the tables, but properties of the ontological classes. Another work in progress is dedicated to this topic. Figure 3 demonstrates a sequence's example.

x signify input variables, y – output variables

Figure 3. Handling information using SPARQL passive and active dashboard filters

The dashboards may have more than one branch. Various dashboards can serve a particular purpose showing some information aspects (*e.g. state of a project, performance of a mechanism, status of city traffic, etc.*). The dashboards are run periodically on demand or triggered by event – process of "collation" of information patterns against RDF data is performed, that is, in other words, running sets of SPARQL queries. An important point is that the dashboards can not only passively show information however perform a calculation utilizing an internal capability or an external service, or trigger predefined events when some value of a process described by metadata circulating in the microblogging is changing or crossing a threshold. Alerts or microblogging directives can be emitted into the channels of microblogging or other communication channels to warn interested parties or change behaviors of the system or agents. Microblogging messages can be either directed to a particular user or to a group of users, or be broadcasted, realizing ability to scale geographical and organizational reach. The general schema of information processing by Metadata system is shown in Figure 4. Manual pre-annotation is not needed for messages generated by systems. They are to be automatically formed in controlled language.
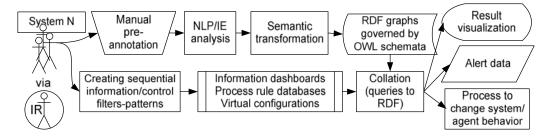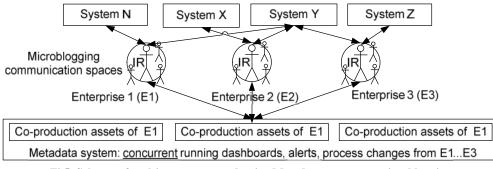

Figure 4. General schema of information processing by Metadata system on microblogging

At such an angle, the system can be considered an enterprise processes' rule database, a monitoring-and-control mechanism or a self-adapted mechanism that is congruent with the notion of virtual configurations in [2] and Metadatabase model with its rule-oriented shells described in [6].

4. Multi-tenant metadata system on microblogging for concurrent co-productions

To reach economy of scale using the concept proposed by Hsu [2], it is necessary to implement mechanisms of concurrent co-productions. As mentioned above, the aggregate of the users connected with IR form communication space of the metadata system/cyberinfrastructure. As microblogging can maintain significant number of such communication spaces, one can characterize such construction as a concurrent co-production multi-tenant environment of communication spaces, and as a service to enterprises.

Fi.5. Schema of multi-tenant co-productive Metadata system on microblogging

Metadata system can be owned by the provider of microblogging service or by another provider and is considered as a separate from microblogging service. Concurrent co-production is achieved by simultaneous running of all dashboard generation, dashboard visualization, alerting and process change jobs in as much paralleled as possible way. The process can be optimized by grouping those by some criteria, cashing, partial reification of RDF data, etc. Schema of such interaction is depicted in Figure 5.

5. Authentication and security aspects

Authentication and security are aspects of enterprise metadata systems/cyberinfrastructures on popular microblogging, needing thorough attention.

Building metadata systems/cyberinfrastructures on microblogging requires capability to connect infinite number of users (people and third-party systems) to microblogging platforms and allow users to login into third-party systems easily from within links in microblogging messages therefore the question of smooth and secure authentication across third-party systems comprising a particular cyberinfrastructure is topical. Insufficiently strong authentication, rights administration and security mechanisms of existent platforms of popular microblogging does not allow deploying cyberinfrastructures of organizations sensitive to information disclosure straightforwardly; however the development has future potential. Particularly promising is QAuth protocol that enables websites or applications (Consumers) to access protected resources from web-service (Service Provider) via an API, without requiring Users to disclose their Service Provider credentials to the Consumers [9]. This protocol is created by microblogging activists and much attention is paid to its simple integration with microblogging. A good example of a shell for managing social media connections and enterprise teams is HootSuite Pro, Social Media Dashboard. In particular, it has an administrative module for managing multiple contributors on various microblogging platforms and sharing data and access without sharing passwords and feature of Secure Profile for additional protection for Enterprise teams.

The user profiles of popular microblogging usually can be public and private and thus there are possible gradations of the metadata/cyberinfrastructure "openness" to external environment, useful and threatening, not analyzed in the scope of this work. Only a fact to mention, if all authorized users of popular microblogging have private profiles, the entire system will be closed for access from outside, however information databases of the private profiles may be accessed from side of the owner of the microblogging platform - usually it has less strict security requirements than cloud enterprise services do. Another peculiarity, - it is possible that owner of microblogging platform and owner of Metadata system with its own RDF data storage are two different organizations so the user information will double in their storages in the two forms. Generally, providers of popular microblogging services should improve security of data they store to be capable of supporting enterprise cyberinfrastructures. For example, BlackBerry-like message securing or text stream ciphering like BoxCriptor can be

used. It is important to highlight that unauthorized access to the metadata system does not automatically mean easy unauthorized access to its connected systems; the latter can be embraced by another security perimeter. For organizations not sensitive to information disclosure or seeking openness, the security factor is not or is less important.

6. Correspondence with Hsu's models of cyberinfrastructure-based/assisted enterprises

The following principles are stated in language of paper [2] with explanations of how the approach is proposed to realize in this paper. In Hsu's model, the three basic categories promoting the O-I fusing paradigm are distinguished:

- *Basic elements of the cyber-infrastructure (the totem pole)* for the three classes of deployment are integrated together through use of popular microblogging as a common standard and an open connectivity solution for proprietary third party systems:
- *the enterprise*: satisfied requirement of integrating third-party information systems into microblogging space by connecting them to microblogging in peer mode; concurrent co-production is achieved by grouping processes from virtual configurations and enterprises by some criteria, cashing, doing partial reification of RDF data in stages, etc.
- *the person (and other moving objects concerned)*: is integrated into microblogging communication space in the same way as enterprises, giving clients/customers levers to monitor and control production processes proactively providing communication/person-centered environment;
- *the environment*: is integrated in the same way as *the person* through the use of sensors with intelligent microblogging clients capable or adaptive behavior.
- *Design and administration of the cyber-infrastructure* (the three-schema model):
- *internal schema* (defining the access methods of the physical data storage): uses proprietary third-party system APIs, QAuth-like authorization, integrating sensors via microblogging clients/SMS, etc. through devices like wireless mobile devices;
- *conceptual schema* (defining the logical data objects and their semantics for the enterprise as a whole): the schema of all area/enterprise-specific substances is maintained in a semantic repositories like Sesame, Ontotext OWLIM, Jena as hierarchically joined OWL schemata from general ones down to more specific with ability to supplement / modify by enterprise personnel and users in a simplified well visualized way;
- *external schemas* (defining the virtual data objects and semantics that individual classes of users and/or applications use): achieved by extensions of the conceptual schema by users and enterprise personnel building their RDF-OWL virtual objects on the base of the *conceptual schema* in the say simplified well visualized way.
- Application of the cyber-infrastructure (engineering of on-demand enterprise processes): achieved through giving users simple visualized means to build information dashboards comprising sequential filters-SPARQL queries to the RDF graphs hiding their complexity.

7. Conclusions

This paper proposes a solution that uses popular microblogging (Twitter, Jaiku, etc.) as a global communication environment common for agents - humans and third-party systems wherein they setup specific rules and exchange messages in controlled languages, and has all features pertinent to the models of cyberinfrastructure-based/assisted enterprises by Cheng Hsu and Emergent Social Software Platforms by Andrew McAfee. The main benefits are following: no need of special communication infrastructure as it is already existent and being continuously optimized by other parties, world-wide reach, simplified semantic representation and simplified interaction with enterprise personnel and end users. Issues of such systems are: handling rapid processes with a lot of metadata and security aspects. Currently a practical investigation is going to commence in Vinnytsia National Technical University, Ukraine. Design Science method will be applied to build a prototype of an integrated system of managing directed group activities on microblogging. This lays ground for creating practitioner-ready communication/metadata spaces on popular microblogging to layer upon existent enterprise eco-systems as one of realization variants to enable "long-term evolution" to fully enabled CyberInfrastructure-Based/Assisted Enterprises.

References:

1. Hsu, Ch. Cheng Hsu (Ph. D), a biography. New York, US-NY [Web resource]: Department of Decision Sciences and Engineering Systems, Rensselaer Polytechnic Institute, 2011. (http://viu.eng.rpi.edu/team/hsuc.html)
2. Hsu, Ch. Models of cyberinfrastructure-based enterprises and their engineering. An evolutional journey. In books: Service enterprise integration. An Enterprise Engineering Perspective. Under edition: Hsu, Ch. New York,

US-NY: Department of Decision Sciences and Engineering Systems, Rensselaer Polytechnic Institute, 2007, pp 209-246. (http://viu.eng.rpi.edu/publications/SEIBookHsuCh08.pdf)

3. Manas, J. Managing the Gray Areas: Essential Insights for Leading People, Projects & Organizations. Minnetonka, US-MN: RMC Publications, Inc, 2007, Premier Edition.

4. McAfee, A. Enterprise 2.0: New Collaborative Tools for Your Organization's Toughest Challenges. Boston, US-MA: Harvard Business School Press, 2009, 1st edition.

5. Fast, T., Copsa J. Twitter Data. A simple, open proposal for embedding data in Twitter messages. [Web resource]: Creative Commons License, 2009. (http://twitterdata.org/)

6. Hsu, Ch., Babin G., Bouziane, M., Rattner, L., Yee, L. The Metadatabase for Manufacturing System Integration. New York, US-NY: Department of Decision Sciences and Engineering Systems, Rensselaer Polytechnic Institute, 1990. (http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.118.9023)

7. Talban, V., Polanjnar, T., Cunningham, H., Botcheva, K. //User-friendly ontology authoring using a controlled language// Department of Computer Science, University of Sheffield, 2006. (http://gate.ac.uk/sale/lrec2006/clie/clie.pdf)

8. Mokin V. B., Yasholt A. R. //New method of automated formation of information systems reports based on sequential approach// Visnyk of Vinnytsia Technical Institute – scientific journal of VNTU, 2008, 3, pp 50-56. (http://visnyk.vstu.vinnica.ua/en/2008/3/)

9. QAuth Core Group. QAuth Core 1.0 Revision A. [Web resource]: QAuth Core Group, 2009. (http://oauth.net/core/1.0a/)

SEI-

,

Билодрей Ю.В., Месюра В.И.

(   )

,

.   CyberInfrastructure-Based/Assisted

     "IT-   ", "   " "   "
                                                             .

          ,   -   ,
                              (Input-Output fusing)

          ,   . Subject-Environment Interaction (SEI)

               /   .
                    ,
          .   (Twitter,

Jaiku,   .)   -

     ,
     ,.
                         /
                    -                        ,