

Semantification of Organizational Content as a Transformational Technology for SMEs

Fefie Dotsika

Westminster Business School, UK

Email: dotsikf@westminster.ac.uk

Abstract—This paper identifies the pairing of social media with the semantic enrichment of web content as a factor with high potential impact for organisations. It explores the challenges encountered in combining the power of collaborative technologies with the capabilities of semantics, examines the practical implications and proposes an adoption method that facilitates decision making by means of a requirements-based approach.

Index Terms—knowledge management & decision making, Internet technology and applications, semantic web applications, information management

I. INTRODUCTION

The Semantic Web [1] and the social media of Web 2.0 [2] have changed the way web applications operate in competitive environments by transforming stakeholder communications, information management, system interoperability and knowledge discovery [3], [4], [5], [6]. Their applications have been divergent and the potential promised by their possible union mostly speculated at. In theory, the collective intelligence of social media and the scalable business models of social computing can be resourcefully paired with internal research and knowledge capabilities derived from interoperable information repositories, such as back-end databases hosting information and operations management and legacy systems. Former research shows that semantic technologies coupled with social media and end-user involvement can instigate innovative influence with wide organisational implications [7], [8]. Knowledge sharing capabilities instigate innovation [9] while the social aspect of web-based sales, such as the growth of social shopping, has a far greater impact on organisations than other technologies [10]. Organisational information assets that have been semantified support innovation, increase productivity and can free human resources so that they can be used to better serve business development [11], [12], [13]. The adoption of semantic technologies in large corporations is common place and/or mandatory in industries such as finance, engineering, procurement, oil and gas [14], [15], [16], [17]. Ontology engineering is becoming mainstream through evolution [18], re-engineering and reuse [19]. The adoption however in

small and medium enterprises (SMEs) is uncommon due to shortage of expertise and resources [20], [21]. Lack of semantic alignment between collaborating organisations leads to misinterpretations unless specialised mapping rules are established up-front, incurring elevated costs [22].

This paper presents a framework that enables decision making through a requirements-based approach assisting the identification of the best available method for information modelling and semantic enrichment.

The rest of the paper is structured as follows: Section 2 describes the methodology and approach followed. Section 3 identifies the challenges associated with semantic annotation and proposes a framework for the facilitation of decision making. Section 4 discusses the implications for practice and Section 5 draws our conclusions.

II. METHODOLOGY AND APPROACH

A combination of action research methods was followed. Participant observation [23], process consultation [24] and Soft Systems Methodology [25] were applied. The research drew upon interviews and surveys with practitioners, consultants and knowledge workers, as well as document analysis and participant observations. Sources of input were London SMEs (Small and Medium Enterprises) seeking social web knowledge management solutions and companies providing semantic technologies over a period of five years (2009-2014). Information was collected from knowledge workers, information strategists, developers, end-users, online forums and blogs. The research unearthed a clear message: organisations want to combine social capabilities (collaboration, personalisation) with adaptive (interoperable, integrated) information access.

The approach is impact-led. The semantification of content impact on the organisation was investigated in terms of quality, innovation and sustainability. The overarching rationale is as follows:

1) The organisational impact of the social aspect is based on maximising collective intelligence and has created the need for organisational strategies that reflect the shift in online culture [26]. Social technologies enable innovation through sources of collective content with functionality that gets enriched as more people use them.

2) The organisational impact of adaptive information access on the other hand is based on semantic enrichment leading to system interoperability and personalised information access. Interoperability addresses heterogeneity issues which are present in data and business processes and it ensures information integration across systems, a process of significant cost. Semantified content facilitates interchange, distribution and creative reuse. Adaptive technologies facilitate the tailoring of information access according to given user profiles. Intelligent information integration and agents such as information brokers, filters, personalised search agents and knowledge management services are examples of innovative applications.

III. A DECISION ASSISTING FRAMEWORK

Our findings mirrored the results of the literature review. While social computing has found its way into most organisations, semantification of content is a reality only in large enterprises mostly, where top-down semantic mark-up is often part of the organisational strategy. Small and medium enterprises (SMEs) have neither the budget nor the expertise for it. The research identified the adoption of semantic applications programmers interfaces (APIs) as a potential solution that enables semantic mark-up and facilitates information discovery.

The increasing popularity of web semantics has resulted in a rise of semantic APIs that offer web content classification and discovery [13]. These semantic APIs take unstructured text (including web pages) as input and return the content's contextual framework. Some of the more popular APIs include Dapper (Data Mapper) API [27], OpenCalais [28], TextWise's Semantic Gist API [29], Semantic Engines API [30], Zemanta API [31] and Ontos API Semantic Web Service [32].

But how can an organisation know whether such an adoption is wise or if semantic enrichment of content can make a difference? Moreover, what information aspects have to be considered and how can the various requirements be grouped so that a decision can be reached?

The research carried out aims to fill this gap by proposing a framework based on two basic aspects governing decision making in organisations. These aspects relate to the perennial questions *why* and *how* [33], [34], [35]. The first aspect presents the case for change and addresses the expected impact on the organisation (*why?*), while the second informs and facilitates the choice of method (*how?*). The framework is schematically represented by a tree structure. The "change" here signifies the organisational adoption of semantified web content and corresponds to the root node of the tree, while the two children nodes correspond to the organisational impact and method determination. The various requirements, constraints, assumptions and other relevant information gathered during our research has been integrated into the semantification decision support structure presented in Fig. 1 below.

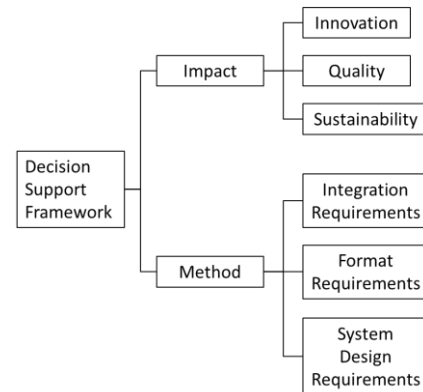


Figure 1. Decision support framework overview

The root node links to the organisational *Impact* and *Method* selection nodes. The *Impact* sub-tree branches out to the *Innovation*, *Quality* and *Sustainability* matrices.

- The Innovation node focuses on technologies and related applications supporting semantic content innovation with organisational implications. Semantified content can model enterprise information and processes with accuracy and consistency, enabling automatic reasoning, concept-based searches, process composition and knowledge discovery. The decision criteria were found to be related to three main criteria: information display (HTML, XHTML, microformats, hCard/Calendar, Topic Maps), syntax and semantics capabilities (XML, RDF, RDFa, N3, NTriples) and rule and inference capabilities (ontologies with RDFS, OWL or folksonomies).
- The Quality node corresponds to information quality issues and has a direct impact on organisational success and profitability, focusing on contextual attributes (relevancy, value-added, timeliness, completeness and volume), representation (interpretation, ease of understanding, concise and consistent representation), accessibility/access security, and intrinsic data qualities (accuracy, objectivity and reliability).
- The Sustainability node addresses organisational change issues and focuses on the impact new technologies bring to organisational processes, functions, values and power and is found to be mostly dependent upon the use of web services and cloud computing. Sustainability is assessed following the underpinning aspects that analyse its conceptual developments (goal orientation and behavioural interaction). There is no evidence that semantic enrichment makes organisations more or less sustainable.

An overview of the decision considerations originating from the Impact branch is presented in Fig. 2.

The Method determination sub-tree branches out to Integration, Format and System Design requirements. This part of the decision making aids web content

semantic enrichment by means of a modular design of requirements-based tools.

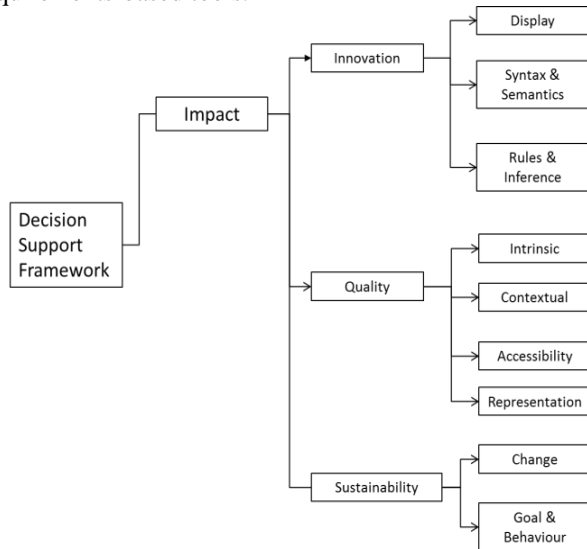


Figure 2. Impact branch expansion

- The Integration requirements are categorised as pertaining to quality, semantic enhancement, mapping completeness and trust/ethics. They focus on the identification of specific issues, highlighting the relevant domain and potential problems. Information modelling quality issues in particular are identified and distinguished from concerns about mapping and semantic clustering.
- The Format requirements aid the decision of semantic format adoption and offer a matrix that covers issues of standardisation, presentation, modelling power (semantics and granularity) and product-related cost constraints in terms of simplicity and implementation.

- The System design requirements focus on the design architecture, end-user involvement, automation, cost, evaluation process and issues of information loss, customisation requirements and semantic enrichment power and/or granularity.

An overview of the decision considerations originating from the Method branch is presented in Fig. 3.

The last part of the framework assists with the potential choice of a semantic API. A comparative matrix for the most popular semantic APIs was constructed based on the attributes deemed most important by the London SMEs: cost, user support, input classification schemes (incl. custom or standard) and output formats. One attribute is conspicuously missing: performance. Sample usage, demos and pilot studies were unable to reach informed conclusions, so the estimates were not included. Table I shows the comparative matrix.

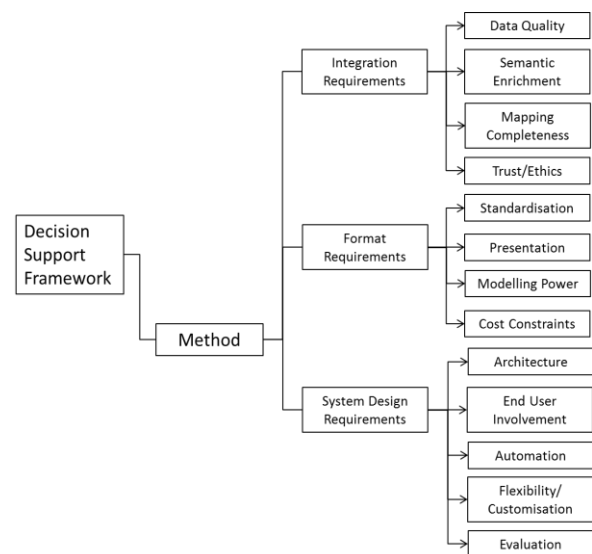


Figure 3. Method branch expansion

TABLE I. POPULAR SEMANTIC APIS

Feature API	Cost	User Support	Input: Classification scheme			Output: Semantic tagging method						
			Custom taxonomy	Standard taxonomy	Classification scheme used	Micro formats	RDF	N3	RDFa	OWL	Topic Maps	Output formats
Dapper	free	yes	yes	yes	User thesaurus or standard taxonomy	n/a	n/a	n/a	yes	n/a	n/a	XML JSON CSV RSS
OpenCalais	free - ££ (Calais Professional)	yes	n/a	yes	RDFS schema [a number of set entities, events and facts]	yes	yes	n/a	n/a	n/a	n/a	XML RDF Microformats JSON CSV
Semantic Gist	free to ££ (depending on agreement)	yes	yes	yes	Open Directory Project (ODP)	n/a	yes	n/a	n/a	n/a	n/a	XML RDF JSON tag cloud formats
Semantic Engines	££	yes	yes	yes	Open Directory Project (ODP)	n/a	yes	n/a	n/a	n/a	n/a	XML RDF JSON tag cloud formats
Zemanta	free to £££	yes	yes	yes	Pre-indexed database of content	n/a	yes	n/a	n/a	n/a	n/a	XML RDF JSON
Ontos	free demo versions available	yes	yes	yes	Fixed ontology that can be enhanced with additional concepts/instances	n/a	yes	yes	n/a	n/a	n/a	XML, RDF N3 JSON

IV. FINDINGS AND IMPLICATIONS FOR PRACTICE

Combining collaborative development with standardised semantics transforms information management and leads to organisational innovation through improving system interoperability, enhancing information find-ability, promoting organisational intelligence, assisting decision making and encouraging employee engagement and participation. The in-depth investigation showed that assisting decision-making in SMEs needs to take into account the possible lack of expertise and resources while focusing on a requirements-based process. The outcome of the research is a decision-support framework which assists practitioners with the semantification of organisational content.

The first part of the framework aids the evaluation of the anticipated impact of the adoption on the organisation. The impact is assessed through a series of considerations that place emphasis on the resulting quality of organisational information and knowledge, the effect of adopting innovation-enabling technologies and issues regarding change and sustainability. In relation to organisational knowledge assets in particular, the method can be used to assess the influence of semantic enrichment on content generation, distribution, retrieval and re-use.

The second part of the framework assists practitioners in deciding what method is to be followed. The method is determined based on a comprehensive list of requirements grouped around system design, integration considerations and format issues such as tailoring and personalisation. System design incorporates matters of design architecture as well as automation, end-user involvement, cost and issues of information loss, customisation requirements and the desired power of semantic enrichment.

The last part of the framework corresponds to a comparative matrix that aids decision making in choosing an automated method of semantification (semantic API). The APIs are compared in terms of a set of criteria based on practitioner requirements, consultants input and participant observation results. Product information and requirement-based decision planning criteria have been factored into the comparison matrix.

V. CONCLUSIONS

The research reinforced the indication that while information is the most important strategic factor in corporate operations, the information and knowledge management challenges in SMEs are idiosyncratic and pose a significant and often decisive obstacle to the flow of knowledge inside the organisation. Semantification of organisational content contributes positively to all aspects of organisational content generation, distribution, retrieval and reuse. It enhances the sharing of a common understanding of a domain among the members of the community, facilitates the analysis and re-use of domain knowledge and makes explicit any assumptions on this

domain. The resulting framework informs and influences organisational policy by providing a decision support mechanism that facilitates semantification and enables information quality enhancement, champions innovation and aids content generation, distribution and retrieval.

REFERENCES

- [1] T. B. Lee, J. Hendler, and O. Lassila, "The semantic web," *Scientific American*, 2001.
- [2] T. O'Reilly. (2005). What is Web 2.0? [Online]. Available: <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>
- [3] A. Fred, J. L. G. Dietz, K. Liu, and J. Filipe, "Knowledge discovery, knowledge engineering and knowledge management," *Communications in Computer and Information Science*, vol. 128, 2009.
- [4] M. Buckland, "What kind of science can information science be?" *Journal of the American Society for Information Science and Technology*, vol. 63, no. 1, pp. 1-7, 2011.
- [5] R. Perrigot, et al. "Antecedents of early adoption and use of social media networks for stakeholder communications: Evidence from franchising," *Journal of Small Business Management*, vol. 50, no. 4, pp. 539-565, 2012.
- [6] S. Firmenich, G. Rossi, M. Winckler, and P. Palanque, "An approach for supporting distributed user interface orchestration over the web," *International Journal of Human-Computer Studies*, vol. 72, no. 1, pp. 53-76, 2014.
- [7] T. Gruber, "Collective knowledge systems: Where the social web meets the semantic web," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 6, no. 1, pp. 4-13, 2008.
- [8] J. Joo, "Adoption of semantic web from the perspective of technology innovation: A grounded theory approach," *International Journal of Human-Computer Studies*, vol. 69, no. 3, pp. 139-154, 2011.
- [9] J. Maes and L. Sels, "SMEs' radical product innovation: The role of internally and externally oriented knowledge capabilities," *Journal of Small Business Management*, vol. 52, no. 1, pp. 141-163, 2014.
- [10] I. Lee, S. J. Yoo, M. J. Choi, and D. H. Shon, "Determinants of social shopping performance in small and medium-sized social merchants: Theories and empirical evidence," *Journal of Small Business Management*, vol. 53, no. 3, pp. 735-747, 2015.
- [11] T. Bass, "Information management challenges on the path to net-centric operations," in *Proc. of Military Communications Conference*, 2005, pp. 2742-2749.
- [12] U. Bojars, J. G. Breslin, V. Peristeras, G. Tummarello, and S. Decker, "Interlinking the social web with semantics," *Intelligent Systems*, vol. 23, no. 3, pp. 29-40, 2008.
- [13] F. Dotsika, "Semantic APIs: Scaling up towards the semantic web," *International Journal of Information Management*, vol. 30, no. 4, pp. 335-34, August 2010.
- [14] M. J. Murphy, M. Dick, and T. Fischer, "Towards the semantic grid: A state of the art survey of semantic web services and their applicability to collaborative design, engineering, and procurement," *Communications of the IIMA*, vol. 8, no. 3, pp. 2. 2014.
- [15] W. E. Mayer, M. Stumptner, G. Grossmann, and A. Jordan, "Semantic interoperability in the oil and gas industry: A challenging testbed for semantic technologies," Ph.D. dissertation, Association for the Advancement of Artificial Intelligence, 2013.
- [16] T. Butler and E. Abi-Lahoud, "A mechanism-based explanation of the institutionalization of semantic technologies in the financial industry," in *Proc. Creating Value for All Through IT*, 2014, pp. 277-294.
- [17] M. Miroslav, M. Miloš, Š. Velimir, D. Božo, and L. Đorđe, "Semantic technologies on the mission: Preventing corruption in public procurement," *Computers in Industry*, vol. 65, no. 5, pp. 878-890, 2014.
- [18] F. Zablit, G. Antoniou, M. d'Aquin, G. Flouris, H. Kondylakis, E. Motta, and M. Sabou, "Ontology evolution: A process-centric survey," *The Knowledge Engineering Review*, pp. 1-31. 2013.

- [19] M. C. Suárez-Figueroa, A. Gómez-Pérez, E. Motta, and A. Gangemi, "Ontology engineering in a networked world," 2012.
- [20] P. D. Gupta, S. Guha, and K. S. Subramanian, "SME growth and influence of internal and external environmental factors," *Encyclopedia of Creativity, Invention, Innovation and Entrepreneurship*, pp. 1654-1665, 2013.
- [21] M. Marchesnay, "Strategic scanning of SME," *Encyclopedia of Creativity, Invention, Innovation and Entrepreneurship*, pp. 1749-1756, 2013.
- [22] C. A. Marín, et al. "Devolved ontology in practice for a seamless semantic alignment within dynamic collaboration networks of SMEs," *Computing and Informatics*, vol. 30, no. 1, pp. 31-55, 2012.
- [23] L. Jepsen, L. Mathiassen, and P. Nielsen, "Back to the thinking mode: Diaries for the management of information systems development projects," *Behaviour and Information Technology*, vol. 8, no. 3, pp. 207-217, 1989.
- [24] E. Schein, *Process Consultation: Its Role in Organizational Development*, Reading, MA: Addison-Wesley, 1969.
- [25] P. Checkland and S. Holwell, "Action research: Its nature and validity," *Systemic Practice and Action Research*, vol. 11, no. 1, pp. 9-21, 1998.
- [26] C. Li and J. Bernoff, *Groundswell: Winning in a World Transformed by Social Technologies*, Boston: Harvard Business Press, 2008.
- [27] Dapper. (2005). [Online]. Available: <http://open.dapper.net/>
- [28] Calais. (2008). [Online]. Available: <http://www.opencalais.com/>
- [29] TextWise. (2015). [Online]. Available: <http://www.textwise.com/api> (accessed 2nd June 2015).
- [30] Semantic Engines. (2009). [Online]. Available: <http://www.semanticengines.com> (accessed 2nd June 2015).
- [31] Zemanta. (2009). [Online]. Available: <http://www.zemanta.com/api/>
- [32] Ontos. (2009). [Online]. Available: <http://www.ontos.com/>
- [33] A. M. Pettigrew, Longitudinal field research on change: Theory and practice," *Organization Science*, vol. 1 no. 3, pp. 267-292, 1990.
- [34] P. Quattrone, and T. Hopper, "What does organizational change mean? Speculations on a taken for granted category," *Management Accounting Research*, vol. 12, no. 4, pp. 403-435, December 2001.
- [35] R. Todnem, "Organisational change management: A critical review," *Journal of Change Management*, vol. 5, no. 4, 2005.



Dr Fefie Dotsika is a Senior Lecturer in Business Information Management at Westminster Business School. Fefie's research interests lie in the area of web information and knowledge management and include web information modelling and classification schemes, emerging web technologies and system interoperability. She is particularly interested in knowledge transfer between universities and organisations and leads research projects that transfer the required expertise and skills to companies in order to improve their competitiveness and profitability through drawing on the knowledge, technology and skills found in Higher Education. Her teaching is in the area of web information and knowledge management as well as the management of web resources in organisations.