

A Preliminary Short Survey of State of the Art Enterprise Search Engines for Future Work Place

Selver Softic, Manfred Rosenberger, Markus Zoier, Konstantin Mondelos
and Erik Pillinger

Virtual Vehicle Research and Test Center, Graz University of Technology, Inffeldgasse 21A, Graz, Austria

Keywords: Survey, Evaluation, Enterprise Search, Future Work Place.

Abstract: This short survey represents a first step towards identifying relevant enterprise search engines as possible key enablers for challenges related to Future Workplace like knowledge transfer and individual and organisational information management inside an enterprise collaboration cycle. In the early period of our research we want to deliver a short overview on current state of the art solutions that can contribute to the idea of Future Workplace. We summarised in our literature study most significant parameters from current research on the topic of semantic and enterprise search. Information was collected via product sheets and white paper of the vendors as well using previous studies and accessible information at the web. Four most advanced solutions in this area has been evaluated. We used them to check their appliance for Future Workplace trends and are aiming at expansion of our review for additional solutions in the future.

1 INTRODUCTION

The Future Workplace tries bringing answers as: which tools, services, processes and actors will be the most challenging for a Knowledge Worker in the near future. One of them is the search and consummation of relevant information in a more precise and faceted way. Hereby the choice of adequate retrieval technology has to suffice to the demands like mobility, flexibility, privacy issues as well as precision regarding search results. Search engines cannot suffice the information need and the specific demands of the Knowledge Worker. This rises a requirement for a new generation of search engine frameworks. Especially in the enterprise surrounding inquiry for such solutions is growing with the complexity of data that should be managed for every day working process. We analysed the needs identified by the other studies in this field and defined upon these findings a subset of eminent features as criteria for evaluating of enterprise search systems. Since no other studies treated enterprise search for Future Workplace we decided to choose the most prominent provider in this field and to evaluate them, using the extracted feature set. Enterprise search represents a step forward regarding the strong data cross-linking which is one of the main demands for Future Workplace as current research claims (Denger et al., 2012). In this short

survey we present results of our evaluation for products: Sinequa, Conveawer, Semaphore and IntraFind. This short paper is organised as follow: After introductory note we will describe some related work regarding Knowledge Worker, Future Workplace, enterprise search studies which tried to define the key concepts to evaluate such systems. Further we will move on with precise description of our methodology and describe the results of our evaluation made upon derived feature set on four enterprise search engines. We will finally draw conclusions from evaluation results and describe the ongoing future work.

2 RELATED WORK

Knowledge Worker as such has been natively defined over 50 years ago by Drucker in (Drucker, 1959), as a person who is payed for intellectual work and corresponding knowledge. In today's enterprises the importance of this role grows with the knowledge cumulation through increased information flow and cross linking. Almost everybody involved in decision processes acts as Knowledge Worker inside the collaboration cycle. Future Workplace therefore as such defines new demands on roles as such involved within the whole process with the occurrence of technolo-

gies like search engines, social media, cloud computing etc., as current research on this topic reveals (Denger et al., 2012; Stocker et al., 2012). Further, according to the recent works the need for increased faceted information consumption with the high degree of mobility and flexibility regarding the service which provides support for this demand is one of the key enablers of Future Workplace. Such demands require more than a simple keyword based search as provided by state of the art search engines like Google¹. These demands therefore raised the question of enterprise search engines.(Mukherjee and Mao, 2004) described the main aspects in more generic way. Studies like (Hager, 1997; Ulbrich et al., 2009; Gronau, 2005) have shown that demands on retrieval, data cross-linking and information consumption for an enterprise search system should provide overlap in a certain set of features. Based upon their findings and demands related to the idea of Future Workplace we summarized the most significant aspects and generated evaluation feature set for the semantic enterprise engines in our survey.

3 METHODOLOGY

In our current research we are seeking for solutions that support management of structured, semi-structured and unstructured data. Desired solution should also at least include personalised search with variety of options and visualisation of such results along with navigation facets and search history support. Access rights management on a huge data pool has been also defined as one of the criteria relevant for the choice of evaluation candidates. Sinequa², Con-Weaver³, Semaphore⁴ and IntraFindIntraFind⁵ represent this level of enterprise search solution which has qualified them as candidates for evaluation. Our methodology includes following steps: First, we identified a subset of studies that deal on feature detection (Hager, 1997; Ulbrich et al., 2009; Gronau, 2005) for rating the enterprise search solution. After comparing the findings in this literature work we derived the relevant common results from all studies and rated them as more significant toward other aspects we found. Demands derived out of the interview series done by (Denger et al., 2012) delivered the additional reduction criteria regarding the final choice of the evaluation feature set. Table 1 shows the features set

¹www.google.com

²<http://www.sinequa.com>

³<http://www.conweaver.de/>

⁴<http://www.smartlogic.com/>

⁵<http://www.intrafind.com>

Table 1: Extracted evaluation metrics.

Domain		Feature
A. Indexing	1	<i>Structured Data</i> (e.g databases, xml, MS Excell)
	2	<i>Semistructured and unstructured data</i> (e.g. e-mail, pdf, MS Word, MS Powerpoint)
	3	<i>Persons</i> (e.g. contacts)
	4	<i>Media data</i> (videos, models, pictures etc.)
	5	<i>Meta data</i>
	6	<i>Different sources</i> (file system, databases, social media, web)
B. Rights	1	<i>Early Binding</i> (integrative approach)
	2	<i>Sticky Policies</i> (item specific access)
C. Analytics	1	<i>Semantic</i>
	2	<i>Natural Language Processing</i>
	3	<i>Statistic</i>
	4	<i>Text Mining</i>
D. Results	1	<i>Classification</i> (result type distinction)
	2	<i>Diversity</i> (diff. types at same time)
	3	<i>Locality</i> (location reproducible)
	4	<i>Results personalisation</i>
	5	<i>Performance and scalability</i> (realtime, Big Data)
	6	<i>Alternative result visualisation</i> (diff. visualisation)
	7	<i>360 degree view</i>
	8	<i>Navigation facets - search history</i>
E. Search	1	<i>Keywords</i>
	2	<i>Query Expansion (QE)</i>
	3	<i>Boolean operators</i>
	4	<i>Wildcards</i>
	5	<i>Case sensitivity</i>
	6	<i>Different search modi</i>
F. OS	1	<i>Windows</i>
	2	<i>Linux, Unix</i>
	3	<i>Other</i>
G. Support	1	<i>Forum/Online Portal</i>
	2	<i>E-mail</i>
	3	<i>Chat</i>
	4	<i>Hotline</i>
	5	<i>Other</i>
H. Languages	1	<i>English</i>
	2	<i>German</i>
	3	<i>Other</i>
I. Integration	1	<i>Portals</i>
	2	<i>Applications</i>
	3	<i>CMS/DMS/ERP</i>
	4	<i>Search engines</i>
	5	<i>Other platforms</i>
J. Administr.	1	<i>Installation</i>
	2	<i>Configuration</i>
	3	<i>Indexing control</i>

that has been extracted as the final version of literature study. As evaluation source we used white papers, official web pages, product sheets (IntraFind, 2013a; IntraFind, 2013b; ConWeaver, 2013; Dirsch-Weigand and Schmidt, 2006; Sinequa, 2013; IBM, 2013; Smartlogic, 2013b; Smartlogic, 2013a) other comparisons as well all accessible product facts that can be found via Web. In the reference part of this paper only official sources has been referenced. The evaluation process will use value + for full support, o for partial support, ~ if uncertain information is only available, - for none support and ? if no information could be found. For legend refer Table 2.

Table 2: Evaluation symbols legend.

Symbol	Meaning
+	full support
o	partial support
~	uncertain information
-	none support
?	no information found

4 RESULTS AND EVALUATION

The idea was to evaluate technical, functional, integral and user oriented features interesting for a Knowledge Worker at Future Workplace. Results on evaluation has been summarized in Table 3 and compared in Figure 1. Considering *A.Indexing*, *C.Analytics*, *E:Search* and *D.Results* feature overlapping between the enterprise search engines is quite high. Methodologies of content analysis differ, however the coverage of feature remains the same. For instance Sinequa and ConWeaver reside more strongly semantic knowledge networks while Semaphore and IntraFind rely on Concept and Topic Maps. The 360 degree view remains typical for Sinequa as well as the integrated *Early Binding* access rights support. All of the search engines reveal in evaluation to be highly integrative and modular (refer domain *I.Integration*).

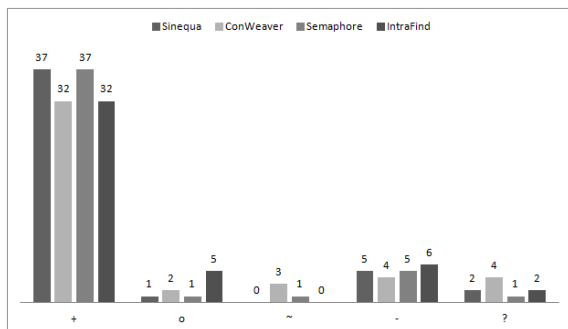


Figure 1: Preliminary feature set comparison.

Some of them more in a generic way, the other ones in customized sense, as in case of IntraFind where tailor made solutions are considered as integration. Multiple language support is claimed to be an extra good integrated feature by all of the providers. Semaphore hand in hand with Sinequa seems to have the most professional *G.Support*, while Sinequa as only product offers an indexing *J.Administration*. Although *Performance and Scalability (D.5)* is characterized as integral part of the product by each of the providers, most impressive information has been delivered from Sinequa⁶. Current installations deal with hundreds of millions of documents. Highest divergence in results occurs in domains *F.OS*, *J.Administration* and *B.Rights*.

5 CONCLUSIONS

Despite the awareness that a literature study can unveil only partly aspects of certain technical systems, results of this survey can be still interpreted in different ways. Choosing the best solution in given context is no valid inference. This work is more overview then rating. The main benefit of this contribution relies on extracted feature set which can be re-used and refined for further evaluations regarding the search requirements for Future Workplace concepts where search as integral part plays a very decisive role. In the future work we are aiming at refinement of feature set and expanding the survey on other enterprise search providers. As stated in the introductory part all evaluations rely on the information provided by the vendors via web pages, product sheets and white papers as well from the one very massively elaborated literature comparison done by (Gronau, 2005). The authors made no technical evaluation of described solutions which means that this evaluation is not intended to be exhaustive and claims not to be complete.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the financial support of the "COMET K2 - Competence Centres for Excellent Technologies Programme" of the Austrian Federal Ministry for Transport, Innovation and Technology (BMVIT), the Austrian Federal Ministry of Economy, Family and Youth (BMWFJ), the Austrian Research Promotion Agency (FFG), the Province of

⁶<http://www.sinequa.com/en/page/solutions/big-data.aspx>

Table 3: Survey results.

	Sinequa	ConWeaver	Semaphore	IntraFind
A.1	+	+	+	+
A.2	+	+	+	+
A.3	+	+	+	+
A.4	+	?	+	?
A.5	+	+	+	+
A.6	+	+	+	+
B.1	+	-	-	-
B.2	-	-	?	-
C.1	+	+	+	+
C.2	+	+	+	+
C.3	+	+	o	+
C.4	+	+	+	+
D.1	+	+	+	+
D.2	+	+	+	+
D.3	+	+	+	+
D.4	+	+	+	+
D.5	+	~	+	+
D.6	+	+	+	+
D.7	+	~	+	?
D.8	+	+	+	o
E.1	+	+	+	+
E.2	+	+	+	+
E.3	+	+	+	+
E.4	+	+	+	+
E.5	+	+	+	+
E.6	+	+	+	+
F.1	+	+	+	+
F.2	-	o	-	+
F.3	-	o	-	-
G.1	-	-	+	-
G.2	+	+	+	+
G.3	-	-	-	-
G.4	+	+	+	+
G.5	+	+	+	+
H.1	+	+	+	+
H.2	+	+	+	+
H.3	+	+	+	+
I.1	+	+	+	o
I.2	+	+	+	o
I.3	+	+	+	o
I.4	o	o	+	+
I.5	+	~	+	o
J.1	+	?	+	+
J.2	?	?	+	+
J.3	?	?	?	+

Styria and the Styrian Business Promotion Agency (SFG) .

REFERENCES

ConWeaver (2013). Conweaver white paper. Website. <http://www.conweaver.de/PDF/ConWeaver.Whitepaper.pdf>.

Denger, A., Schmeja, M., and Stocker, A. (2012). *Future Workplace: eine Untersuchung sozio-technischer*

Einflüsse auf den Arbeitsplatz der Zukunft. Berichte aus der Wirtschaftsinformatik. Shaker.

Dirsch-Weigand, A. and Schmidt, I. (2006). Conweaver.: automatisierte wissensnetze fr die semantische suche. *Information - Wissenschaft & Praxis*, 57(6/7):367–371.

Drucker, P. (1959). *The Landmarks of Tomorrow*. Heine-mann.

Gronau, N., editor (2005). *Anwendungen und Systeme für das Wissensmanagement*. Reihe Wirtschaftsinformatik: technische und organisatorische Gestaltungsoptionen. Gito-Verl., Berlin, 2. aufl. edition.

Hager, S. (1997). *Evaluierung von Suchmaschinen fr den Einsatz im BMWIntranet*. Technische Universität München.

IBM (2013). Ibm partnerworld, global solutions directory, sinequa description. Website. <http://www-304.ibm.com/partnerworld/gsd/solutiondetails.do?solution=32061&expand=true&lc=en>.

IntraFind (2013a). ifinder enterprise search unternehmensweit informationen finden. <http://www.intrafind.de/de/produkte/ifinder>.

IntraFind (2013b). ifinder enterprise search, product data sheet. Website. http://www.intrafind.de/tl_files/documents/Produkte/INTRAFIND_Datenblatt_iFinder4_7_DE_022013.pdf.

Mukherjee, R. and Mao, J. (2004). Enterprise search: Tough stuff. *Queue*, 2(2):36–46.

Sinequa (2013). Product overview. Website. <http://www.sinequa.com/en/page/product/product.aspx>.

Smartlogic (2013a). Semantic enhancement server white paper. Website. http://www.smartlogic.com/request-download?f=35/SL_SES_WPNT0012011_EN.pdf.

Smartlogic (2013b). Semaphore overview a smartlogic white paper. Website. http://www.smartlogic.com/requestdownload?f=121/SL_SemaphoreOverview_WPNT0052011_EN.PDF.

Stocker, A., Denger, A., Hübler, A., Ruckriegel, H., Maletz, M., and Klimisch, M. (2012). Arbeitsplatz der zukunft mit fallstudien von bmw und avl. *HMD - Praxis Wirtschaftsinform.*, 287.

Ulbrich, A., Kraker, P., and Luidolt, C. (2009). Ergebnisse einer untersuchung von anbietern von suchmaschinen-technologien. In Hinkelmann, K. and Wache, H., editors, *Wissensmanagement*, volume 145 of *LNI*, pages 118–127. GI.