UNIVERSITY
OF TRENTO

DEPARTMENT OF INFORMATION AND COMMUNICATION TECHNOLOGY

38050 Povo – Trento (Italy), Via Sommarive 14
http://www.dit.unitn.it

PEER-TO-PEER KNOWLEDGE MANAGEMENT

Matteo Bonifacio, Fausto Giunchiglia and Ilya Zaihrayeu

Arpil 2005

Technical Report # DIT-05-042

Also: Proc. of the 5th International Conference on Knowledge Management (I-KNOW), 2005 (to appear)

Peer-to-Peer Knowledge Management

Matteo Bonifacio
(Center for Scientific and Technological Research (ITC-irst), Italy
bonifacio@itc.it)

Fausto Giunchiglia
(University of Trento, Italy
fausto@dit.unitn.it)

Ilya Zaihrayeu
(Center for Scientific and Technological Research (ITC-irst), Italy
zaihrayeu@itc.it)

Abstract: Peer-to-Peer (P2P) is a decentralized networking paradigm where autonomous parties have equivalent capabilities in providing other parties with data and/or services. On the other hand, Knowledge Management (KM) is viewed as a core capacity in order to compete in the modern social and economic environment. In the view of the emerging semantic web technologies, P2P is looking for knowledge-driven domains to better exploit its technological potential. At the same time, driven by economical and social trends, KM is questioning its centralized nature assumption and is looking for a technological paradigm in order to benefit from exploiting its distributed dimension. In this paper we discuss the state of the art and trends in both the P2P and KM fields, discuss what possible synergies can benefit integrated P2P KM solutions, and present an implemented P2P KM system.

Keywords: Peer-to-Peer Computing, Knowledge Management, Semantic Web
Categories: H.1.1, I.2.4, I.2.11

1 Introduction

Peer-to-Peer (P2P) computing has received significant attention from the side of research labs and academia, largely due to the popularity of commercialized P2P file sharing applications such as Napster, Morpheus and KaZaa. In the P2P model, peers exchange data and/or services in a completely decentralized distributed manner. Peers are autonomous, and are free to choose what other peers to interact with. In this point-to-point interaction, peers possess equal functional capabilities.

On the other hand, Knowledge Management (KM) is increasingly viewed as a core capacity in order to compete in the modern social and economic environment ([Devenport 02], [Senge 90], [Nonaka 95]). Researchers and practitioners agree that those intellectual assets [see Stewart 02] that are embedded in working practices, social relationships, and technological artefacts constitute the only source of value that can sustain long term differentiation, quality of services, innovation, and adaptability. Nonetheless, even due to a debatable success of current KM implementations, it is still unclear how such matter should be managed in highly complex, distributed, and heterogeneous settings.

In the last couple of years, P2P and KM have followed different but converging paths. In fact, P2P technologies have left their initial "computational", "anarchoyd", and spontaneous fashion to embrace more service level domains and business settings. On the other hand, KM is questioning its centralized assumption based on the implicit belief that knowledge is managed successfully when it can be standardized and controlled. In this sense, it seems that while P2P is looking for value added domains to better exploit its technological potential, KM is looking for a technological paradigm more able to fit an emerging distributed organization of knowledge.

This paper gives a short overview of the P2P model and applications, and explains the increasing requirement for P2P applications to address knowledge-driven domains (Section 2). The paper then gives a short introduction to the KM field, and describes the advantages of distributed KM solutions (Section 3). In Section 4 we discuss what synergies can benefit integrated P2P KM solutions, and present a concrete example of a P2P KM application. Section 5 provides the conclusions for the paper.

2 Towards Knowledge-Driven Peer-to-Peer computing

In the Peer-to-Peer (P2P) model computational peers interoperate in a completely decentralized distributed environment, providing to and requesting from each other data and/or services. Peers are largely autonomous in what and how they store in their local (knowledge) bases, in what data and services they provide to other peers, in what other peers they "talk to", etc. Peers come and go, make spontaneous acquaintances with other peers, and, eventually, drop them. Peers collaboratively process user requests, and the overall performance of the network emerges from local point-to-point interactions of (all) peers on the network.

P2P applications cover a number of domains such as file sharing (e.g., Kazaa[1]), distributed computing (e.g., SETI@Home[2]), collaborative networking (e.g., Groove[3]), and instant messaging (e.g., ICQ[4]). However, most of the P2P systems are *hybrid* – peers, (mainly) for discovery purposes, rely on a centralized resources indexing server. An example of a totally decentralized P2P application is Gnutella[5], which is a file sharing application. In file sharing applications, all peers rely on the same schema that describes the content of the files they share. What here needs to be underlined is that, although technologically distributed, P2P (data sharing) systems are at least conceptually centralized as far as they have to assume some shared semantics in order to exchange data meaningfully (e.g. names and meaning of categories). For a comprehensive overview of the P2P technology and applications see, for instance, [Milojicic 02].

According to some studies carried out at Gartner, P2P has passed its "peak of inflated expectations", and now true technology's applicability, risks and benefits

---

[1] Kazaa; Link: http://www.kazaa.com
[2] SETI@Home; Link: http://setiathome.ssl.berkeley.edu
[3] Groove; Link: http://www.groove.net
[4] ICQ; Link: http://www.icq.com
[5] Gnutella; Link: http://www.gnutella.com

need to be understood[6]. At this stage, exploration of application domains where P2P can better exploit its technological potential becomes vital. According to the Gartner report, it will take from 5 to 10 years before the real-world benefits of the P2P technology are demonstrated and accepted.

On the other hand, the emerging Semantic Web (SW) technologies open new horizons for P2P. The vision of the SW is to enable machines to retrieve and process meta-data (i.e., information about data), and exploit this knowledge for intelligent, meaningful and context-driven interoperation with users and other applications[7]. In the SW P2P scenario, users encode their knowledge in a formal structure, such as ontology, and then share it with other users and applications; they create communities of knowing which gradually evolve as new knowledge is brought in. Moreover, the SW allows it to overcome current limitations of P2P data sharing applications. Namely, it allows for richer and mutually heterogeneous peers' schemas, while still ensuring interoperability. Note that when peers' schemas are heterogeneous, the role of centralized resource indexing servers diminishes, as there is no common schema to index resources.

Another knowledge-intensive domain is (organizational) Knowledge Management (KM). There is very little in the literature about what P2P can do for KM. However, Ovum reports that "*Of all the application domains we have studied, knowledge management is the one where the benefits of peer-to-peer and a clear and straightforward business model for suppliers are most evident*" [Axton 02]. Ovum identifies at least two major areas where P2P can have impact on KM: collaboration and knowledge discovery. The former area includes the support for virtual teams and organizations, and for communities of practice. The latter one is about finding content, social patterns and specialists' profiles in enterprise and personal P2P networks. At the technical level, these tasks require expressive formalisms to represent and share knowledge. From this point of view, SW technologies become very useful.

Currently, none of the commercialized P2P applications can be seen as a comprehensive KM solution. However lots of research is being done on this topic in academia. For instance, the Semantic Web community discusses the role of ontologies in KM systems (e.g., [Ehrig 03], [Fensel 02]). The database community proposes several solutions for a completely decentralized P2P database system with the support of heterogeneous schemas [Bernstein 02], [Halevy 03]. There are many other areas that contribute to building viable P2P KM solutions – personal knowledge management [Tsui 02], semantic matching of ontological structures to facilitate peers interoperability [see Giunchiglia 03], and so on.

3 Towards Distributed Knowledge Management

The traditional architecture of KM systems have embodied the assumption that, to share and exploit knowledge, it is necessary to implement a process of knowledge-

---

[6] Gartner Hype Cycle 2002 – Information Technology Trends

[7] For more information on the Semantic Web, see the W3C Semantic Web activity page: http://www.w3.org/2001/sw/

extraction-and-refinement, whose aim is to eliminate all subjective and contextual aspects of knowledge, and create an objective and general representation that can then be reused by other people in a variety of situations [Bonifacio 00], [Bonifacio 02b], [Bonifacio 02a]. Very often, this process is finalized to build a central knowledge base, where knowledge can be accessed via a knowledge portal. This centralized approach – and its underlying objectivist epistemology – is one of the reasons why so many KM systems are deserted by users, who perceive such systems either as irrelevant or oppressive (see [Alvesson 01], [Tsoukas 01]). As clearly pointed by the Report on Knowledge Management to the European Commission, 2004 "KM is a crucial competence in the new competitive arena… but the degree of predictability which has been inherent in KM thinking, reflecting the general belief in linearity, is now seriously questioned."

During the last year, the evidence that knowledge is a distributed, contextual and subjective matter have led to an alternative vision, the so called Distributed Knowledge Management (DKM). As described in [Bonifacio 02b], DKM is an approach to KM based on the principle that the multiplicity (and heterogeneity) of perspectives within complex organizations should not be viewed as an obstacle to knowledge exploitation, but rather as an opportunity that can foster innovation and creativity.

The fact that different individuals and communities may have very different perspectives, and that these perspectives affect their representation of the world (and therefore of their work) is widely discussed – and generally accepted – in theoretical research on the nature of knowledge. Knowledge representation in artificial intelligence and cognitive science has produced many theoretical and experimental evidences of the fact that what people know is not a mere collection of facts; indeed, knowledge always presupposes some (typically implicit) interpretation schema, which provides an essential component in sense-making (see, for example, the notions of context [McCarthy 93], [Bouquet 98], [Ghidini 01], mental space [Fauconnier 85], partitioned representation [Dinsmore 91]). Moreover, studies on the social nature of knowledge stress the social nature of interpretation schemas, viewed as the outcome of a special kind of "agreement" within a community of knowing (see, for example, the notions of scientific paradigm [Kuhn 79], frame [Weick 95]), thought world [Dougherty 92], perspective [Boland 95]). In this sense, rather than linear, knowledge dynamics are better represented by evolving constellations of autonomous and heterogeneous "knowledges" that consolidate their knowledge locally and seek for some form of coordination with other "knowledges" by means of semantic negotiation.

Despite this large convergence, the need to preserve an idea of control which is inherent to the very notion of business resource, led organizations and managers to neglect what increasingly happens behind the scene of any corporate intranet: people and groups still continue to develop an interrelated web of local systems that better fit their needs. On the other hand, the risk that these constellations become unmanageable, and the maturity that is being reached by P2P technologies are increasingly attracting managers towards alternative perspectives. Among these, at least the opportunity to weakly control knowledge constellations making them visible, instead of hiding them behind the official but ephemeral claims of control provided by the knowledge portal. In this less ideal but more realistic landscape, P2P systems seem particularly suitable to implement this view in which the existence of knowledge "lobbies" is recognized and inter-community cooperation is supported through the provision of adequate coordination facilities. In the next section, we propose a system that tries exactly to go in this direction.

4 Peer-to-Peer Knowledge Management – Prospects and an Application

The convergence of P2P and KM technologies creates new challenges for researchers to address: new methodologies to model, design, and deploy distributed KM solutions; theories and algorithms to represent the social and semantic dimensions of a knowledge network; mechanisms to cope with the dynamic autonomous nature of P2P and to provide means to support emergent network self-organization. New technologies should be provided in order to support full operational functioning of P2P KM systems, ensuring high extensibility of the solutions along several dimensions, such as scalability in the number of peers, size and kind of supported knowledge bases, level of heterogeneity in knowledge representation, etc.

In return, P2P KM applications give the prospects of robustness, large pool of shared resources and semantics-driven tools to effectively operate these resources; local autonomy and comprehensive support to each peer provided by a collaborative effort of other peers. P2P KM solutions are not dependent on the presence of certain peers or content on the network; instead, peers bring new knowledge which flows along semantic links between peers, being enriched and completed on the fly. The "knowledge base" of a P2P KM network is formed dynamically; peers forge and break knowledge groups based on a common interest, etc. From this point of view, it turns out that operation on a P2P KM network naturally complements usual economic and social patterns.

As an example of an existing P2P KM system, we propose in this paper a P2P DKM architecture, named KEEx, which is the result of the research project EDAMOK[8] (Enabling Distributed and Autonomous Management of Knowledge). In KEEx, each community of knowing (or Knowledge Nodes (KN), as they are called in [Bonifacio 02a]) is represented by a peer, and the two principles above are implemented in a quite straightforward way: (i) each peer provides all the services needed by a knowledge node to create and organize its own local knowledge (autonomy), and (ii) by defining social structures and protocols of meaning negotiation in order to achieve semantic coordination (e.g., when searching documents from other peers). Built on top of this architecture, distributed local "knowledges" can emerge and aggregate through a bottom-up process from the individual level, to the organizational one, passing through the establishment of communities (group of peers that share a similar interest) and zones (networks of peers that relate to a neighbourhood). Moreover, peers' knowledge bases can be run on top of industrial solutions such as existing content management applications or databases. As a consequence, knowledge bases become to be virtual, flexible and

[8] EDAMOK; Link: http://edamok.itc.it

temporary aggregation of both individual and more institutionalized knowledge sources.

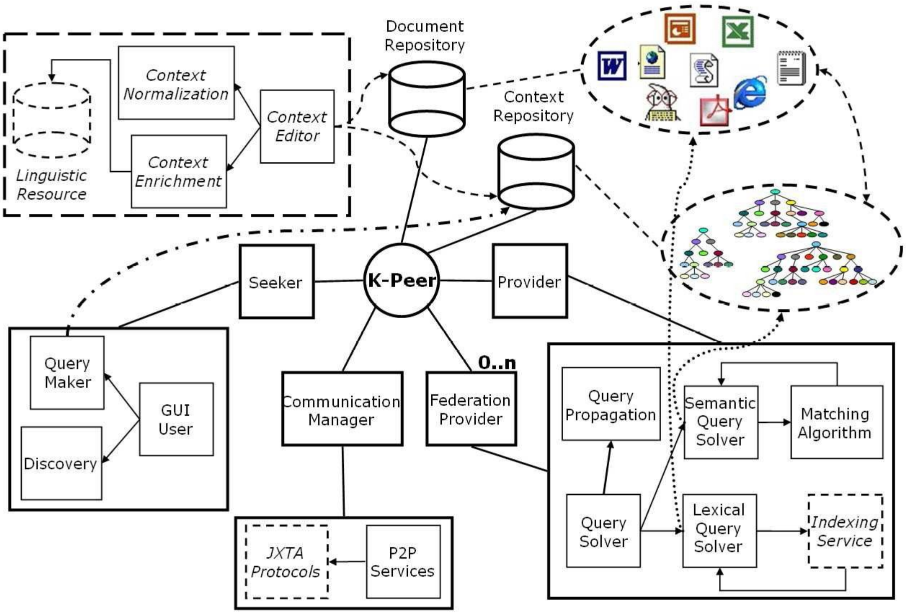

Fig 1. The KEEx's main components

The main components of KEEx are shown on Figure 1. Each Knowledge Node (also known as *K-peers*) can play two main roles: *provider* and *seeker*. In the former case, a K-peer "publishes" in the system a body of knowledge, together with an explicit representation on it. In the latter case, a K-peer searches for information by explicitly specifying a query as a part of its own perspective. K-peers store their local knowledge in *document repositories*. K-peers formally describe the real world (from their own perspective) in the approximate and partial form of a *context*. Contexts are also used by seekers for query representation. A K-peer may have more than one context, and it stores its context(s) in a *context repository*. Contexts are created, manipulated and used by K-peers by means of the *context management module*, which includes a *context editor* and a *context browser*. Apart from this, KEEx allows for semantic matching of contexts, for forming federations of K-peers based on knowledge that the peers have in common, and for peers discovery. KEEx is implemented on top of the P2P platform JXTA[9]. For a throughout discussion of KEEx's components and functionality, see [Bonifacio 02c].

5 Conclusions

In this paper, we argued that technological architectures, when dealing with processes in which human communication is strongly involved, must be consistent with the social architecture of the process itself. In particular, in the domain of KM,

[9] JXTA; Link: http://www.jxta.org

technology must embody a principle of distribution that is intrinsic to the nature of organizational cognition. Here, we suggest that P2P infrastructures are especially suitable for KM applications, as they naturally implement meaning distribution and autonomy. It is worth noting at this point that other research areas are moving toward P2P architectures. In particular, we can mention the work on P2P approaches to the semantic web [Arumugam 02], to databases [Giunchiglia 02], to web services [Papazoglou 03]. We believe this is a general trend, and that in the near future P2P infrastructures will become more and more attractive to all areas where we can't assume a centralized control.

References

[Alvesson 01] Alvesson M., Kärreman D.: "Odd Couple: Making Sense of the Curious Concept of Knowledge Management"; Journal of Management Studies, 38(7), 2001.

[Arumugam 02] Arumugam M., Sheth A., Budak Arpinar I.: "The peer-to-peer semantic web: A distributed environment for sharing semantic knowledge on the web"; in WWW Workshop on Real World RDF and Semantic Web Applications. Honolulu, Hawaii (USA), 2002.

[Axton 02] Axton C., Gear R., Macehiter N., Woods E.: "Peer-to-Peer Computing: Applications and Infrastructure"; Ovum Report, February, 2002.

[Bernstein 02] Bernstein P., Giunchiglia F., Kementsietsidis A., Mylopoulos J., Serafini L., Zaihrayeu I.: "Data Management for Peer-to-Peer Computing: A Vision"; in proceedings of the Fifth International Workshop on the Web and Databases (WebDB 2002), Madison, Wisconsin, June 6-7, 2002.

[Boland 95] Boland J.R., Tenkasi R.V.: "Perspective making and perspective taking in communities of knowing"; Organization Science, 6(4):350–372, 1995.

[Bonifacio 00] Bonifacio M., Bouquet P., Manzardo A.: "A distributed intelligence paradigm for knowledge management"; In Working Notes of the AAAI Spring Symposium Series 2000 on Bringing Knowledge to Business Processes. AAAI, March 18-20, 2000.

[Bonifacio 02a] Bonifacio M., Bouquet P., Cuel R.: "Knowledge Nodes: the Building Blocks of a Distributed Approach to Knowledge Management"; Journal of Universal Computer Science, 8(6):652–661, 2002.

[Bonifacio 02b] Bonifacio M., Bouquet P., Traverso P.: "Enabling distributed knowledge management: managerial and technological implications"; Novatica and Informatik/Informatique, III (1), 2002.

[Bonifacio 02c] Bonifacio M., Bouquet P., Mameli G., Nori M.: "KEx: a Peer-to-Peer Solution for Distributed Knowledge Management"; in Proceedings of the Fourth International Conference on Practical Aspects of Knowledge Management (PAKM02), Vienna, Austria, 2-3 December 2002. Lecture Notes in AI, Vo. 2569, pp.490-500, Springer Verlag, Heidelberg 2002

[Bouquet 98] Bouquet P.: "Contesti e ragionamento contestuale. Il ruolo del contesto in una teoria della rappresentazione della conoscenza"; Pantograph, Genova (Italy), 1998.

[Devenport 02] Devenport T.H., Prusak L.: "Working Knowledge"; Harvard Business School Press., 2000.

[Dinsmore 91] Dinsmore J.: "Partitioned Representations"; Kluwer Academic Publishers, 1991.

[Dougherty 92] Dougherty D.: “Interpretative barriers to successful product innovation in large firms”; Organization Science, 3(2), 1992.

[Ehrig 03] Ehrig M., Tempich C., Broekstra J., van Harmelen F., Sabou M., Siebes R., Staab S., Stuckenschmidt H.: “SWAP - ontology-based knowledge management with peer-to-peer technology”; In Proceedings of the 1st National Workshop Ontologie-basiertes Wissensmanagement (WOW2003), 2003.

[Fauconnier 85] Fauconnier G.: “Mental Spaces: aspects of meaning construction in natural language”; MIT Press, 1985.

[Fensel 02] Fensel D., Staab S., Studer R., van Harmelen F.: “Peer-2-Peer enabled Semantic Web for Knowledge Management”; In J. Davis et al. (eds.), Towards the Semantic Web: Ontology-Driven Knowledge Mangement, Wiley, 2002.

[Ghidini 01] Ghidini C., Giunchiglia F.: “Local Models Semantics, or Contextual Reasoning = Locality + Compatibility”; Artificial Intelligence, 127(2):221–259, 2001.

[Giunchiglia 02] Giunchiglia F., Zaihrayeu I.: “Making peer databases interact – a vision for an architecture supporting data coordination”; In 6th International Workshop on Cooperative Information Agents (CIA), 2002.

[Giunchiglia 03] Giunchiglia F., Shvaiko P.: “Semantic Matching”; In the Knowledge Engineering Review Journal, vol. 18(3): 265-280, 2003.

[Halevy 03] Halevy A. Y., Zachary G. I., Mork P., Tatarinov I.: “Peer Data Management Systems: Infrastructure for the Semantic Web”; WWW Conference, 2003.

[Kuhn 79] Kuhn T.: “The structure of Scientific Revolutions”; University of Chicago Press, 1979.

[McCarthy 93] McCarthy J.: “Notes on Formalizing Context”; In Proc. of the 13th International Joint Conference on Artificial Intelligence, pages 555–560, Chambery, France, 1993.

[Milojicic 02] Milojicic D. S., Kalogeraki V., Lukose R., Nagaraja K., Pruyne J., Richard B., Rollins S., Xu Z.: “Peer-to-Peer Computing”; HP Laboratories Palo Alto, report HPL-2002-57

[Nonaka 95] Nonaka I., Takeuchi H.: “The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation”; Oxford University Press, New York, 1995.

[Papazoglou 03] Papazoglou M.P., Jang J., Kraemer B.J.: “Leveraging web-services and peer-to-peer networks”; CAiSE: 485-501, 2003.

[Senge 90] Senge P.: “The Fifth Discipline: The Art and Practice of the Learning Organization”; Doubleday, 1990.

[Stewart 02] Stewart T.A.: “The wealth of knowledge”; Doubleday, New York, 2002.

[Tsoukas 01] Tsoukas H., Vladimirou E.: “What is organizational knowledge?”; Journal of Management Studies, 38(7), 2001

[Tsui 02] Tsui E.: “Technologies for Personal and Peer-to-Peer (P2P) Knowledge Management”; CSC Leading Edge Forum (LEF) Technology Grant report, 2002

[Weick 95] Weick E. K.: “Sensemaking in Organizations”; Sage Publications Inc, 1995