A Survey of Semantic Based Peer-to-Peer Systems

Vijay Srinivas Agneeswaran*,
Distributed Information Systems Lab (LSIR)
Ecole Polytechnique Fédérale de Lausanne.

Abstract

This paper makes a step in identifying the state of the art in semantic P2P systems. On one hand, lot of research in the P2P systems community has focused on fault-tolerance and scalability, resulting in numberous algorithms, systems such as Chord, Pastry and P-Grid. These systems, however, have no notion of semantics and consequently, have difficulty in knowledge sharing. On the other hand, research in the semantic web community have focused on knowledge sharing among different nodes with possibly different schemas. These have tended to use centralized repositories. The obvious benefits of combining P2P and semantic systems would be to have large scale collection of structured data. Several recent efforts have focused on this combination. However, there have been no attempt to have these efforts grouped in one place for easy assimilation and for finding interesting future directions; this paper fills the gap.

1 Introduction

The Semantic Web is a vision and effort by the W3C (http://www.w3.org/) to enable globally dispersed users to share and access structured data, much like a global database. Ideally, the Semantic Web would allow data from diverse sources (possibly stored using different schemas) to be accessed seamlessly, as there would be ways of making the data understandable across

*vijaysrinivas.agneeswaran@epfl.ch

sources. The data representation is mainly in the form of Resource Description Framework (RDF), which has now become a W3C standard (http://www.w3.org/RDF/). RDF serves as a minimalistic representation format for data, similar to the role played by HTTP in the World Wide Web. RDF stores such as Jena [1] and others allow RDF elements to be stored in relational database and allow queries over the RDF data, in the form of SPARQL (http://www.w3.org/TR/rdf-sparql-query/) or RDQL (http://www.w3.org/Submission/RDQL/) queries.

Many existing RDF stores store RDF data in centralized repositories, making it difficult to provide scalability and fault-tolerance. Moreover, the very nature of the semantic web calls for completely decentralized approaches for data management, for reason of freshness and flexibility [2].

Research in Peer-to-Peer systems have focused on scalability and fault-tolerance properties, mainly in file systems. Starting from unstructured systems such as Gnutella [3] and Napster [4], P2P systems have evolved to structured systems such as Chord [5], Pastry [6] and P-Grid [7]. The structured P2P systems provide a scalable mechanism to store and retrieve data based on indexes, even in the presence of node failures. However, the P2P system community have not focused on semantics, with the result that state of the art P2P systems having no notion of semantics. Thus, it makes perfect sense to marry P2P systems with the semantics in order to realize the Semantic Web vision. This has resulted in what is known as semantic based P2P systems. The focus of this paper is to provide a survey of the state of the art in semantic based P2P systems. The rest of the paper is organized as follows:

2 Storing and Querying RDF Data in Semantic P2P Systems

We first focus on research efforts which involve storing and querying semantic data in P2P systems. RDF is the format generally used to capture semantic information. Hence, this category can be looked at as RDF stores and efforts to efficiently query RDF stores in P2P systems.

2.1 RDF and Traditional Query Architectures

This [8] paper provides an architectural perspective of querying RDF data in a distributed environment. It describes the mediator architecture that is at the heart of semantic interoperability solutions. The mediator stores a central schema and is responsible for query reformulation, when it gets a query wherein the target or the destination of the query has a different schema. Different mediator architectures are described in order to improve the scalability and fault-tolerance, including hierarchical mediator architecture (HMA) and a cooperative mediator architecture (CMA). HMA is a first step in reducing the query load on the mediator, as sub-mediators can also perform some query processing. However, the hierarchical nature still limits fault-tolerance. The CMA provides fault-tolerance in addition to scalability, by organizing the mediators as super-peers, with other nodes accessing them for RDF query resolution. The super-peers themselves are connected by a random graph, just like unstructured P2P systems. This makes it difficult to provide guarantees on querying and can also reduce scalability due to flooding. Moreover, the issues relating to layer management [9] including super-peer failures, ratio of peers to super-peers (which impact search performance) etc. are not easy to handle.

2.2 Query Processing in RDF/S-Based P2P Database Systems (SQPeer)

SQPeer [10] is a sophisticated query engine for processing RQL queries, focusing mainly on conjunctive queries involving RQL class and property patterns. Peers advertise their respective bases in the form of RVL views by publishing it in the P2P network. The information advertised by each peer is mainly intensional (schema related). SQPeer handles query formulation in which a given query pattern is augmented with information about peers that can actually answer it. If required, a complex query is broken down into simpler sub-queries. A P2P lookup service is then used to find peer views relevant to the input query. A query planner generates a query execution plan. The plan is executed with the help of appropriate peers.

2.3 Cayley DHTs - A group Theoretic Framework for Analyzing DHTs Based on Cayley Graphs

This paper [11] analyzes the different DHTs based on group theory algebra and concludes that a number of DHTs are based on Cayley graphs, a general group theoretic model for analyzing interconnection networks. For instance, Chord is a special ciculant graph which are Cayley graphs obtained from the cyclic group and an arbitrary (inversion-closed) generator set, while CAN is a d-dimensional torus, which is also a Cayley graph. The paper further shows that even non-Cayley DHTs such as P-Grid and Pastry/Tapestry also have several interesting properties of Cayley graphs, resulting in desirable features such as symmetry and load balancing, hierarchy, fault-tolerance, proximity, hamiltonicity (the ability to embed a ring structure for dynamic routing). The overall goal is to investigate if the algebraic design approach enables one to build DHTs in which data and nodes need not necessarily be hashed. This would be very useful for the semantic web, as non-hashed structured P2P systems would allow one to apply semantic web and database technologies for data organization and query processing, combined with the desirable features of DHTs - scalability and fault-tolerance.

2.4 Semantic Query Routing in Unstructured Networks Using Social Metaphors

This paper [12] describes an algorithm knows as REMINDIN for searching/querying RDF data in an unstructured P2P system using social metaphors. The key idea is to draw inspiration from human social networking; a human being may observe the communication between others and knows who may be able to answer his/her queries. Each peer can select at most two peers from a set of known peers for a given query to avoid flooding the network. If a peer fails to select any peers for a given query, then it uses query relaxation to broaden the query until eventually, all peers can be reached. By using results from a real implementation and a simulation, the authors show that REMINDIN provides better recall than other algorithms. They also show that network load reduces over time, in terms of messages per query.

2.5 Expertise-Based Peer Selection

This paper [13] provides an semantic overlay comprising peer expertise advertisements. Peers advertise their expertise in the network, with the expertise being described using a shared ontology. Queries submitted to a peer are sent to appropriate nodes and not flooded to the network. Selecting an appropriate node to evaluate a query is based on a semantic matching of the query with the peer expertise. The authors describe a bibliographic scenario, namely the bibster system, in which each peer stores an RDF knowledge base comprising set of bibliographic meta-data items. The shared ontology in this case is the ACM topic hierarchy. An advertisement of a peer is a set of topics for which this peer is an expert. Using simulation and real experiments, the authors show that expertise-based peer selection improves the performance of peer selection significantly. This paper [14] also describes the Bibster bibliographical system, focusing on the recommender functionality.

2.6 RDFPeers

RDFPeers [15] was one of the first efforts at building a P2P RDF store using structured overlays. The key idea was to index the RDF triple three times - one each for subject, object and predicate by using globally known hash functions and hence, to store the triple in three nodes which have closest matching ids as the corresponding indexes. This ensures that triple based queries will be successful if the triple exists and it will be found within O(log(N)). Another important advantage of RDF peers is that it does not require schema definitions apriori (before triples can be inserted into the network). RDFPeers is built over Multi-attribute Addressable Network (MAAN) [16], which itself was an extension of Chord [5] to handle multi-attribute and range queries. RDFPeers is built as a layer over MAAN that provides RDF specific functionalities including RDF storage, retrieval and load balancing. RDF triples are also stored in neighbouring peers (based on the replication factor) in order to provide fault-tolerance. RDFPeers can support triple based queries - atomic triple patterns, conjunctions with the same variable and constant subjects and different constant predicates. RDFPeers does not support arbitrary joins, selections and projections.

2.7 Evaluating Conjunctive Queries Over Structured Overlays

This paper addresses the open problem of evaluating full conjunctive queries for RDF triples over a P2P network [17]. A Query Chain (QC) algorithm is proposed for the same. The query initiator in QC takes a full conjunction of RDF triples and forms a query chain by sending each triple pattern to a (possibly) different node, based on the hash value of constant part of the triple. Each of these nodes are sent a message comprising the query itself, index of the triple pattern to be computed by it, the query originator's IP address and an intermediate relation to hold intermediate values. The authors mention that the order of evaluation of the triple patterns is important for optimizations, but currently assume a left to right order. The paper also presents an improved algorithm known as Spread By Value (SBV) that spreads the query to more nodes (multiple query chains) than QC by using the intermediate query results. They also show that the performance of queries can be improved by using query caching - a process where IP address of overlay neighbour is stored along with query, so that the overlay routing can be skipped next time, saving network traffic and query time.

3 Semantics, DistrIbuted Schemas in P2P Systems

This category of papers focus on semantic interoperability in a possibly peer data setting. This means these efforts attempt to solve the problem of exchaning and querying semantic data, with each peer having its own independent and possibly unknown schemas.

3.1 Semantic Coordination of Heterogeneous Classifications Schemas

This paper [18] addresses the semantic interoperability problem by focusing on the (pairwise) mapping between dynamic schemas. Even if the schemas are not known apriori (as may be the case in a P2P system where peers are free to come and go), a mapping may be established. This is possible by two new ideas: One, the mappings must be semantic relations (which have well defined model-theoretic interpretations); Two, the meaning implicit in each schema element must be made explicit, so that the mapping becomes a problem of finding a semantic relation between two schemas, with each schema containing meaningful labels for the elements. Semantic elicitation proceeds by using three different levels of knowledge: namely, lexical knowledge (knowledge about words in the labels), world knowledge (knowledge about the relations between the concepts expressed in words) and structural knowledge (structure of schema graph). Once the schemas are encoded using the three knowledge types, the problem translates into one of logical deduction via implication.

3.2 Semantic Mapping by Approximation

This paper [19] addresses the issue of coordinating or finding an agreement between heterogeneous schema models. A method to discover approximate mappings between two given concepts is proposed. An approximation value that indicates "how strongly" a concept is a subconcept of the second is calculated for each pair of concepts. The numerical value indicates the error in the subsumption relation, with close to 0 meaning most semantic concept of first is there in second. The paper applies the ideas in the context of music ontologies, by extracting music meta-data from music sites on the Internet (common music sites such as those found on Amozon and Yahoo). They use a formalism based on propositional logic for arriving at the approximate mapping.

3.3 Satisficing Ontology Mapping

This paper [20] explores semantic mapping from the perspective of satisfying a decision maker by achieving sufficient quality. Consequently, the algorithm proposed named as Quick Ontology Mapping (QOM) addresses semantic translation between peers, where large ontological structures have to be mapped to each other in a few seconds. QOM has lesser run-time complexity compared to other approaches, illustrating the trade-off in sacrificing accuracy in favour of efficiency. The word "satisficing" is inspired from human decision making coined by Herbert Simon (Nobel Laurrete in Economics) which targets not optimality but with limited resources, time, attention and efforts at solutions that are good enough.

4 Peer Data Management Systems

This category of papers attempt to solve the problem of building a scalable infrastructure for semantic systems, with the focus being on the scalability part rather than on the semantics, compared to the previous category, where the focus was on semantic interoperability but not scalability.

4.1 Active XML

A simplistic view of a traditional distributed database management system is that it uses a centralized server to provide a global schema and ACID properties through transactions. Several approaches have extended these techniques to work in a decentralized manner, to apply to Internet or P2P systems. Active XML [21] provides dynamic XML documents over web services for distributed data integration. It is a model for replicating (whole file) and distributing (parts of a file) XML documents by introducing location aware queries in X-Path and X-Query. It also provides a framework by which peers perform decentralized query processing in the presence of distribution and replication. It allows peers to optimize localized query evaluation costs, by a series of replication steps.

4.2 Edutella

Edutella [22] attempts to design and implement a schema based P2P infrastructure for the semantic web. It uses W3C standards RDF and RDF Schema as the schema language to annotate resources on the web: achieving a mark up for educational resources. Edutella provides meta-data services such as querying, and replication as well as semantic services such as mapping, mediation and clustering. Edutella services are built over JXTA [23], a widely used framework for building P2P applications. Edutella query service provides the syntax and semantics for querying both individual RDF repostitories and for distributed querying across repositories. Edutella uses *mediators* to provide coherent views across data sources through semantic reconciliation. Edutella was visualized to provide a platform for educational institutions to participate in a global information network, retaining autonomy of learning resources.

The same authors have also attempted to use super-peer based organization of the Edutella peers to make searching more efficient. The paper [24] describes an organization of the super-peers based on HyperCup, a structured P2P system based on the Hypercube topology [25]. The super-peers maintain meta-data for a set of peers, instead of each peer maintaining its own meta-data. The super-peers themselves are connected using the Hypercup overlay. This makes searching for meta-data quite efficient, as searches are executed only in the super-peer overlay. They also use super-peer indices based on schema information to facilitate faster search.

4.3 Piazza

Piazza [26] is a peer data management system that facilitates decentralized sharing of heterogeneous data. Each peer contributes schemas, mappings, data and/or computation. Piazza provides query answering capabilities over a distributed collection of local schemas and pairwise mappings between them. It essentially provides a decentralized schema mediation mechanism for data integration over a P2P system. Peers in the system contribute to stored relations, similar to data sources in data integration systems. Query reformulation occurs through stored relations, stored either locally or at other peers. Piazza also addresses the key issue of security, which would enable users to share their data in a controlled manner. Another paper [27] describes the way a single data item is published in protected form using crptographic techniques. The owner of the data item encrypts the data and can specify access control rights declaratively, restricting users to parts of the data.

4.4 PIER

P2P Information Exchange and Retrieval (PIER) [28] is a P2P query engine for query processing in Internet scale distributed systems. PIER provides a mechanism for scalable sharing and querying of finger print information, used in network monitoring applications such as intrusion detection. PIER uses four guiding principles in its design. First, it provides relaxed consistency semantics - best effort results, as achieving ACID properties may be difficult in Internet scale systems [29]. Second, it assumes organic scaling, meaning that there are no data centres/warehouses and machines can be added in typical P2P fashion. Third, the query engine assumes data is available in native file systems and need not necessarily be loaded into local databases. The fourth principle is that instead of waiting for breakthroughs on semantic technologies for data integration, PIER tries to combine local and reporting mechanisms into a global monitoring facility. PIER is realized over CAN, the hypercube based P2P system [30].

4.5 PeerDB

PeerDB [31] is an object management system that provides sophisticated searching capabilities. PeerDB is realized over BestPeer [32], which provides P2P enabling technologies. PeerDB can be viewed as a network of local databases on peers. It allows data sharing without a global schema by using meta-data for each relation and attributes. The query proceeds in two phases: in the first phase, relations that match the user's search are returned by searching on neighbours. After the user selects the desired relations, the second phase begins, where queries are directed to nodes containing the selected relations. Mobile agents are dispatched to perform the queries in both phases.

4.6 GridVine: Building Internet-Scale Semantic Overlay Networks

It can be observed from a reading of the previous sections that several research efforts focus on query answering and semantic mapping reconcilation in federated databases. Peers could have different schemas and data storage mechanisms and are logically interconnected through schema mappings, forming a semantic overlay network. The other research focus has been on structured overlay networks and how scalable, fault-tolerant applications can be built over DHTs. However, these efforts did not focus on semantics. Hence, GridVine [33] is the first attempt at using a structured overlay network (namely P-Grid) to realize semantic overlays. GridVine realizes semantic overlays by separating a logical layer from a physical layer, applying the well known database principle of data independence.

The logical layer has operations to support semantic interoperability including attribute based search, schema inheritance, schema management and schema mapping. Schema inheritance allows users to derive new schemas from existing well known base schemas, fostering interoperability by reusing sets of conceptualizations belonging to the base schemas. It also supports a new schema reconciliation technique known as semantic gossiping for semantic interoperability in decentralized settings [34].

The physical layer provides efficient realizations of operations that exploit the structured overlay network. This includes mapping data and meta-data to routable keys, introduction of a namespace for resources, so that resource requests can be resolved and implementations of traversals of the semantic network, taking advantage of intermediate schema mappings.

5 Conclusions

The paper has presented a survey of semantic P2P systems. The first category of papers were ones which addressed storing and querying of RDF stores. The second set of efforts were focused on addressing issues of semantic interoperability in a peer setting. The third category included attempts to build scalable infrastructures for semantic systems, putting scalability as the central focus. GridVine is one of the systems which addressed both the semantic interoperability and scalability issues.

This survey has opened up several research possibilities. One such is in the area of benchmarking semantic P2P systems. Simple benchmarks for P2P systems have been proposed, including the *find owner* and *locate replicas* [35]. Recent efforts have also proposed benchmarks for P2P information retrieval [36]. However, none of these address semantic P2P systems. There was an effort to benchmark RDF queries using the Barton data set [37]. The data set, however, is centralized. Thus, there have been no real efforts at benchmarking semantic P2P systems.

Consequently, the key issue of performance of the semantic P2P system has not been focused on by the research community. Once benchmarks are in place and existing systems are found inadequate (in terms of performance), as the case may be, building efficient, scalable semantic P2P systems will become an important area of research.

References

[1] Brian McBride. Jena: A Semantic Web Toolkit. *IEEE Internet Computing*, 6(6):55–59, 2002.

[2] Staab, Steffen and Stuckenschmidt, Heiner, editors. *Semantic Web and Peer-to-Peer*. Springer-Verlag, Berlin Heidelberg, 2006.

[3] Gnutella. The Gnutella protocol specification v0.4. http://www9.limewire.com/developer/gnutella protocol 0.4.pdf, 2000.

[4] NAPSTER. Napster media sharing system. http://www.napster.com.

[5] I. Stoica, R. Morris, D. Karger, M. F. Kaashoek, and H. Balakrishnan. Chord: A Scalable Peer-to-Peer Lookup Service for Internet Applica-

tions. *IEEE/ACM Transactions on Networking*, 11(1):17–32, February 2003.

[6] A. Rowstron and P. Druschel. Pastry: Scalable, Distributed Object Location and Routing for Large-Scale Peer-to-Peer Systems. In *Proceedings of the 18th IFIP/ACM International Conference on Distributed Systems Platforms (Midleware 2001)*, pages 329–350, Heidelberg, Germany, November 2001.

[7] Karl Aberer, Philippe Cudre-Mauroux, Anwitaman Datta, Zoran Despotovic, Manfred Hauswirth, Magdalena Punceva, and Roman Schmidt. P-Grid: a Self-Organizing Structured P2P System. *ACM SIGMOD Record*, 32(3):29–33, 2003.

[8] Richard Vdovjak, Geert-Jan Houben, Heiner Stuckenschmidt, and Ad Aerts. RDF and Traditional Query Architectures. In Steffen Staab and Heiner Stuckenschmidt, editors, *Semantic Web and Peer-to-Peer*. Springer-Verlag, Berlin Heidelberg, 2006.

[9] Li Xiao Zhenyun Zhuang and Member-Yunhao Liu. Dynamic Layer Management in Superpeer Architectures. *IEEE Transactions Parallel and Distributed Systems*, 16(11):1078–1091, 2005.

[10] George Kokkinidis, Lefteris Sidirourgos, and Vassilis Christophides. Query Processing in RDF/S-Based P2P Database Systems. In Steffen Staab and Heiner Stuckenschmidt, editors, *Semantic Web and Peer-to-Peer*. Springer-Verlag, Berlin Heidelberg, 2006.

[11] Changtao Qu, Wolfgant Nejdl, and Matthias Kriesell. Cayley DHTs - A Group-Theoretic Framework for Analyzing DHTs Based on Cayley Graphs. In Steffen Staab and Heiner Stuckenschmidt, editors, *Semantic Web and Peer-to-Peer*. Springer-Verlag, Berlin Heidelberg, 2006.

[12] Christoph Tempich and Steffen Staab. Semantic Query Routing in Unstructured Networks Using Social Metaphors. In Steffen Staab and Heiner Stuckenschmidt, editors, *Semantic Web and Peer-to-Peer*. Springer-Verlag, Berlin Heidelberg, 2006.

[13] Ronny Siebes, Peter Haase, and Frank van Harmelen. Expertise Based Peer Selection. In Steffen Staab and Heiner Stuckenschmidt, editors, *Semantic Web and Peer-to-Peer*. Springer-Verlag, Berlin Heidelberg, 2006.

[14] Peter Haase, Marc Ehrig, Andreas Hotho, and Bjorn Schnizler. Personalized Information Access in a Bibliographical Peer-to-Peer System. In Steffen Staab and Heiner Stuckenschmidt, editors, *Semantic Web and Peer-to-Peer*. Springer-Verlag, Berlin Heidelberg, 2006.

[15] Min Cai and Martin Frank. RDFPeers: a scalable distributed RDF repository based on a structured peer-to-peer network. In *WWW '04: Proceedings of the 13th international conference on World Wide Web*, pages 650–657, New York, NY, USA, 2004. ACM Press.

[16] Min Cai, Martin Frank, Jinbo Chen, and Pedro Szekely. MAAN: A Multi-Attribute Addressable Network for Grid Information Services. In *GRID '03: Proceedings of the Fourth International Workshop on Grid Computing*, page 184, Washington, DC, USA, 2003. IEEE Computer Society.

[17] Erietta Liarou, Stratos Idreos, and Manolis Koubarakis. Evaluating Conjunctive Triple Pattern Queries over Large Structured Overlay Networks. In *International Semantic Web Conference*, volume 4273 of *Lecture Notes in Computer Science*, pages 399–413. Springer, 2006.

[18] Paolo Bouquet, Luciano Serafini, and Stefano Zanobini. Semantic Coordination of Heterogeneous Classifications Schemas. In Steffen Staab and Heiner Stuckenschmidt, editors, *Semantic Web and Peer-to-Peer*. Springer-Verlag, Berlin Heidelberg, 2006.

[19] Zharko Aleksovski and Warner ten Kate. Semantic Mapping By Approximation. In Steffen Staab and Heiner Stuckenschmidt, editors, *Semantic Web and Peer-to-Peer*. Springer-Verlag, Berlin Heidelberg, 2006.

[20] Marc Erhig and Steffen Staab. Satisficing Ontology Mapping. In Steffen Staab and Heiner Stuckenschmidt, editors, *Semantic Web and Peer-to-Peer*. Springer-Verlag, Berlin Heidelberg, 2006.

[21] Serge Abiteboul, Angela Bonifati, Grégory Cobéna, Ioana Manolescu, and Tova Milo. Dynamic XML documents with distribution and replication. In *SIGMOD '03: Proceedings of the 2003 ACM SIGMOD international conference on Management of data*, pages 527–538, New York, NY, USA, 2003. ACM Press.

[22] Wolfgang Nejdl, Wolf Siberski, and Michael Sintek. Design issues and challenges for RDF- and schema-based peer-to-peer systems. *SIGMOD Record*, 32(3):41–46, 2003.

[23] L Gong. JXTA: A Network Programming Environment. *IEEE Internet Computing*, 5(3):88–95, 2001.

[24] Wolfgang Nejdl, Martin Wolpers, Wolf Siberski, Christoph Schmitz, Mario Schlosser, Ingo Brunkhorst, and Alexander Löser. Super-peer-based routing and clustering strategies for RDF-based peer-to-peer networks. In *WWW '03: Proceedings of the 12th international conference on World Wide Web*, pages 536–543, New York, NY, USA, 2003. ACM Press.

[25] Mario T. Schlosser, Michael Sintek, Stefan Decker, and Wolfgang Nejdl. Hypercup - hypercubes, ontologies, and efficient search on peer-to-peer networks. In Gianluca Moro and Manolis Koubarakis, editors, *First International Workshop on Agents and Peer-to-Peer Computing*, volume 2530 of *Lecture Notes in Computer Science*, pages 112–124. Springer, 2002.

[26] Igor Tatarinov, Zachary Ives, Jayant Madhavan, Alon Halevy, Dan Suciu, Nilesh Dalvi, Xin Dong, Yana Ka diyska, Gerome Miklau, and Peter Mork. The Piazza Peer Data Management Project. *SIGMOD Record*, 32, September 2003.

[27] Gerome Miklau and Dan Suciu. Controlling Access to Published Data Using Cryptography. In *in Proceedings of Very Large Databases Conference* , pages 898–909, USA, September 2003. VLDB Endowment.

[28] Ryan Huebsch, Joseph M. Hellerstein, Nick Lanham, Boon Thau Loo, Scott Shenker, and Ion Stoica. Querying the Internet with PIER. In Johann Christoph Freytag, Peter C. Lockemann, Serge Abiteboul, Michael J. Carey, Patricia G. Selinger, and Andreas Heuer, editors, *VLDB 2003, Proceedings of 29th International Conference on Very Large Data Bases, September 9-12, 2003, Berlin, Germany*, pages 321–332. Morgan Kaufmann, 2003.

[29] Seth Gilbert and Nancy Lynch. Brewer's Conjecture and the Feasibility of Consistent, Available, Partition-Tolerant Web Services. *SIGACT News*, 33(2):51–59, 2002.

[30] Sylvia Ratnasamy, Paul Francis, Mark Handley, Richard Karp, and Scott Schenker. A Scalable Content-Addressable Network. In *SIGCOMM '01: Proceedings of the 2001 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 161–172, New York, NY, USA, 2001. ACM Press.

[31] Wee Siong Ng, Beng Chin Ooi, Kian-Lee Tan, and Aoying Zhou. PeerDB: A P2P-based System for Distributed Data Sharing. In Umeshwar Dayal, Krithi Ramamritham, and T. M. Vijayaraman, editors, *Proceedings of the 19th International Conference on Data Engineering*, pages 633–644. IEEE Computer Society, 2003.

[32] Wee Siong Ng, Beng Chin Ooi, and Kian-Lee Tan. BestPeer: A Self-Configurable Peer-to-Peer System. In *Proceedings of the 18th International Conference on Data Engineering, 26 February - 1 March 2002, San Jose, CA*, page 272. IEEE Computer Society, 2002.

[33] Karl Aberer, Philippe Cudré-Mauroux, Manfred Hauswirth, and Tim Van Pelt. GridVine: Building Internet-Scale Semantic Overlay Networks. In Sheila A. McIlraith, Dimitris Plexousakis, and Frank van Harmelen, editors, *International Semantic Web Conference*, volume 3298 of *Lecture Notes in Computer Science*, pages 107–121. Springer, 2004.

[34] Karl Aberer, Philippe Cudré-Mauroux, and Manfred Hauswirth. The Chatty Web: Emergent Semantics through Gossiping. In *WWW '03: Proceedings of the 12th international conference on World Wide Web*, pages 197–206, New York, NY, USA, 2003. ACM Press.

[35] S. Rhea, T. Roscoe, and J. Kubiatowicz. Structured peerto -peer overlays need application-driven benchmarks, 2003.

[36] Thomas Neumann, Matthias Bender, Sebastian Michel, and Gerhard Weikum. A Reproducible Benchmark for P2P Retrieval. In Philippe Bonnet and Ioana Manolescu, editors, *ExpDB*, pages 1–8. ACM, 2006.

[37] Daniel J. Abadi, Adam Marcus, Samuel R. Madden, and Kate Hollenbach. Using The Barton Libraries Dataset As An RDF Benchmark. Technical Report MIT-CSAIL-TR-2007-036, MIT, 2007.