Alchemist: User Driven Searching in Ubiquitous Networks

Matthew Shields
School of Physics & Astronomy
Cardiff University
Cardiff, UK
matthew.shields@astro.cf.ac.uk

Ian Taylor
School of Computer Science
Cardiff University
Cardiff, UK
Ian.J.Taylor@cs.cardiff.ac.uk

ABSTRACT

This paper describes a generic peer-to-peer architecture, called Alchemist, which provides a peer-to-peer overlay coupled with a data fusion workflow environment for the searching or discovery of distributed resources. The Alchemist is designed to be generic and can host a number of different peer-to-peer applications or other frameworks. We provide an example of how it can be applied to audio and visual searching by using the underlying toolkit to provide workflows for the manipulation of different kinds of audio and visual metadata. This data-fusion capability allows high-level search mechanisms to be built on top of the framework and in addition we demonstrate how other search applications from the scientific community can be implemented.

Categories and Subject Descriptors

C.2.4 [**Distributed Systems**]: Distributed applications, Distributed databases; H.3.3 [**Information Search and Retrieval**]: Search process

1. INTRODUCTION

Internet search engines widely used today are based primarily on crawling the Web to index sites, pushing this data into large databases, and then building sophisticated ranking systems based upon the limited information that can be scraped from individual Web pages. When an end-user search is performed, these massive databases perform an, often cached, textual mapping; relate it to the given ranking system; and return the user with hundreds of thousands of results, ranked accordingly. The Alchemist infrastructure proposed here considers an alternative approach to this paradigm, allowing users to *proactively* push information into the "search database" in a similar fashion to the *publish–subscribe* model, except that queries rather than subscriptions are used to retrieve the data. To enable a completely dynamic, user-driven, scalable, and decentralised database, Alchemist is based on a caching peer-to-peer (P2P) network that scales with growing demand and has no central point of failure or control. In Alchemist, clients push information, from audio, visual, and other files, onto a set of dynamically selected super-peers that collectively form the distributed search database and use idle cycles on participants computers to process information and perform search queries.

Built upon Alchemist's base P2P overlay network is a toolkit, which is designed to allow different search algorithms to be plugged into the network. The algorithms are composed as workflows, allowing developers to extend current search schemes to create finer-grained matching and filtering of results. By allowing this flexibility, search queries will no longer be confined to the lowest common denominator of current ranking and indexing systems, and Alchemist will allow for searches to combine different data sources to create complex workflows to filter results and weigh parameter values.

ADPUC'06 November 27-December 1, 2006 Melbourne, Australia

2. GENERAL OVERVIEW

The Alchemist is a software infrastructure to enable advanced indexing and searching of any digital content, including AV, scientific, etc., in ubiquitous and decentralised P2P networks. By integrating applications, data providers, digital content, and algorithms, the Alchemist toolkit enables the simple composition of mixed-media queries for combinational searches to interpret heterogeneous data sets in a logically defined order. Fusing metadata from distinct, yet related, sources into a contextually aware environment allows Alchemist to multiplex search results and produce rich metadata that goes beyond what can be traditionally harvested from a single object. At the core, Alchemist provides application developers with a framework for specifying complex search algorithms, using a series of logical search steps, within a graphical workflow builder. By building on Alchemist's extensible architecture, developers do not need to write custom software algorithms from scratch, and are able to create complex queries and data fusion techniques in a modular and pluggable fashion. The Alchemist system comprises of a *distributed P2P database framework*; the *Alchemist Integrated Toolkit* which provides the horizontal (algorithm and tool) and vertical (application) integration of components; and *application integration* components that provide a suite of search, indexing and meta-information generation tools as a set of AIT plug-ins.

Alchemist, in the theme of its metaphysical counterpart, is a system that supports the transformation of *raw* data deposited quasi-randomly by users, into complex and composable views from which meaning can be derived. The Alchemist Integrated Toolkit supports the (meta) data fusion from context, social or content analysis for the integrated analysis of AV data, or other digital content. Metadata obtained from various sources is integrated and through morphological derivations forms part of the default search capabilities. The AIT combines content- or index- specific raw data and converts it into a focused search for a specific AV or scientific event.

3. RELATED WORK

Alchemist spans many different technology domains, the following list outlines some of the more general technology categories of related work.

Cycle scavenging technologies [2] have become ingrained in current Internet culture through applications like the Search for Extra Terrestrial Intelligence (SETI@Home)[1], BOINC[2], Einstein@Home[3]. In highly-parallel applications similar to SETI, powerful home machines, which are often connected over high-speed residential Internet lines, "donate" their idle cycles to a particular task or set of tasks, e.g. alien hunting, genome matching, high-energy physics. Currently, cycle-sharing technology is limited to specific problem spaces and requires significant work on the side of application developers, as well as a dedicated and relatively altruistic user-community, to succeed. Alchemist uses cycle scavenging on a local basis to process the digital content on a peer, generating the metadata that gets pushed into the distributed database.

P2P networks can currently be broadly classified as using "unstructured" or "structured" approaches to locate resources. Gnutella[4] and Kazaa[5] are examples of unstructured P2P networks, with hosts and resources being made available through dynamic network overlays, and without any global overlay planning. Distributed Hash Table (DHT) based systems such as FreeNet [6] and Chord [7] use a so-called "structured network" overlay of peers. This structuring consists of a logical identifier space to which peers and resources are mapped. Peers maintain neighbourhood state about each other enabling application level routing of messages through the network based on the identifier space.

Alchemist builds upon open frameworks developed by industry (JXTA[6]) and academic institutions (P2PS [21]) to create a specific, yet extensible and broad, P2P network that is tailored toward the searching and processing of digital content. The Alchemist (P2P) Framework uses the latest technologies for developing both structured and unstructured P2P overlays, and then using simulation environments, is able to fine-tune and test both network topologies and algorithms to find best-of-breed matches dependent on network traffic and processing power.

Today's uses for P2P networks range from making telephone calls over the internet, as in the case of the popular and closed P2P network used by Skype[7], to analysing scientific data: BOINC; EinsteinHome; SETI@Home, or making distributed backups of office and home data[4][5]. There are a number of P2P based distributed search engines that can be compared to Alchemist, however these provide simple search mechanisms for locating specific items such as mp3 files, and are often quickly closed down because of illegal file sharing. Alchemist differs from these in that its metadata generation capabilities allow for the location of artefacts within the digital content, specific images within a video, or a phrase within an audio file, for example. Where content is protected under DRM or some other restriction it is only the metadata that is shared in the distributed database and not the digital content itself.

Web service and look-up mechanisms have become popular as ways to expose new business logic and modular systems from previously monolithic, hidden applications. The next stage in the development of Web services is to make them more dynamic [9] and adaptable to different network systems, which will add a new layer of flexibility to current architectures. One current technology investigating this avenue with relation to P2P networks is WSPeer [11]. WSPeer extends P2PS' basic P2P network building tools to enable simple, lightweight Web service management across P2P networks. Using WSPeer, applications can host and invoke Web services through a Service Oriented Architecture using standard discovery protocols such as UDDI. One advantage of WSPeer is that it allows an application to expose functionality as a Web service *on the fly* without requiring the usual infrastructure associated with service hosting, such as a Web server and a service container. Alchemist uses and extends this infrastructure to provide dynamic Web service creation and hosting for different technological aspects of the system such as metadata creation and publishing, exposing these components through standards based Web service interfaces.

Workflow tools and environments provide intuitive graphical user interfaces (GUIs) for the visual construction and integration of multiple services and interfaces. In a workflow, a series of tasks is visually represented to provide the user with an intuitive graphical means of combining separate tasks into a single process. Triana [16] is a popular workflow environment for scientific applications and audio processing that has been used in a number of international projects. Using the Triana GUI, a user is presented with a familiar interface, i.e. toolbars, menus, workspace, that can be used to graphically connect individual tools or algorithms to form a workflow. The Triana environment already contains many standard tools for statistical analysis, signal, visual, and audio processing. Triana also has the capability to use, and incorporate into the workflows, remote services such as P2P, e.g. JXTA and P2PS; WSRF; and both Grid [1]; and Web services, using WSPeer. Within Alchemist, Triana provides a way to develop, link, extend, and provide workflows for complex data search algorithms, which are then pushed out onto the network of peers that comprise the distributed P2P database.

4. ALCHEMIST TECHNOLOGIES

Alchemist's goals are to create a toolkit for search algorithm deployment and a decentralised P2P network for distributing search databases and processing. This differs from current systems that are centralised, have a finite search functionality, and are costly and prone to attack. The Al-

[1]SETI@Home `http://setiathome.berkeley.edu/`
[2]BOINC - Berkeley Open Infrastructure for Network Computing `http://boinc.berkeley.edu/`
[3]Einstein@Home `http://einstein.phys.uwm.edu/`
[4]Gnutella `http://www.gnutella.com/`
[5]Kazaa `http://www.kazaa.com/us/index.htm`
[6]JXTA `http://www.jxta.org`
[7]Skype VOIP `http://www.skype.com/`

chemist distributed network extends the state of the art for data sharing and searching by providing a defined and scalable infrastructure for hosting advanced search techniques through a pluggable infrastructure for the inclusion of new search algorithms. Rather than relying on hard-coded parameters or simple string comparisons, searches within Alchemist allow mixed-mode queries that can combine different searches and data types through complex logic structures. Alchemist therefore defines the following P2P and Search properties:

P2P middleware that builds a robust super-peer overlay network capable of being deployed on an Internet scale, from mobile devices to cluster computers. The resulting middleware will include a toolkit for simulation, which can be used for stress-testing Alchemist software and algorithms in a number of extreme network scenarios before deployment, helping Alchemist to push the current limits of P2P research.

Performance is a central issue in today's massive P2P networks that involve home users. Incorrect message overheads and frequencies can easily bring a network to a standstill and diminish ever-important user confidence. The pluggable nature of Alchemist allows for new algorithms for searching and indexing to be inserted into the network. To ensure that these new additions are sustainable, Alchemist will make use of prototype systems and simulation environments such as NS-2 [15][8][9] to test, develop, and produce its algorithms and P2P network infrastructure.

Decentralised caching agent overlays enable the design of decentralised layers of super-peers to perform domain-specific caching and processing. Through simple interfaces, an application is able to deploy custom algorithms across the super-peer network, which can be used for advanced caching, sophisticated searches and matching, or to provide agent-type capabilities. In addition the caching super-peers help insulate the network from the effects of the transient nature of the network where peers on users machine come and go. In this system, the cycles of participant computers are used to analyse and gather data, making users both providers and consumers of information and cycles.

Search toolkit allows applications to use the Alchemist P2P network by simply plugging in indexing and search algorithms and new data sources. The modular and future-oriented design of the AIT facilitates Alchemist's use in new application domains beyond those currently addressed. For the scenarios presented here, there could be new algorithms or adding new data sources to provides a core set of AV indexing and search mechanisms that can be extended for any digital data type. However, the framework is designed to be extensible to other domains and we are also particularly interested in certain scientific applications e.g. for template searching for gravitational wave signals (in collaboration with the AEI in Golm) and managing results from distributed astrophysical simulations (in collaboration with CCT in Louisiana State University).

Workflows for creating search algorithms allow developers to move away from the domain-specific and monolithic code structures that typify current search algorithm development and deployment. Workflows enable rapid development in distributed settings through code reuse, componentisation, defined typing, and flexible distribution and packaging. By basing algorithm development on extensible and graphical interfaces for construction, Alchemist workflows can be developed by non-programmers and easily incorporated into the existing network.

[8]The Protean Research Group, SRSS Project, Naval Research Lab. `http://cs.itd.nrl.navy.mil/5522/`

[9]The Ns2 Simulator. `http://www.isi.edu/nsnam/ns/`

5. Alchemist COMPONENTS

5.1 The P2P Framework

The Alchemist architecture is built using P2P technologies for enabling decentralisation in both data storage and processing across the network and to scale in line with the network's size and characteristics. The network is an unstructured centralised/decentralised architecture consisting of a set of super peers that act as the backbone of the network. Super-peers can act as caching gateways to groups of peers that are connected. The information that is cached consists of adverts and the pre-processed metadata. While all peers, super-peers and ordinary peers, expose specific functionality, i.e. services, to the network and communicate using defined message exchanges, the topological details as well as transport, transfer and discovery protocols used, should remain flexible. This encourages diversity, flexibility and inclusiveness in the network.

Against this backdrop, Alchemist must address how diverse and complex data objects are represented, stored, queried and shared in line with the requirements of the nodes involved and the characteristics of the data. This involves devising strategies to manage different types of data such as protected content, undesirable content and mixed media. The relationship between content, its context: protected; copyrighted etc.; and the queries and algorithms it is subjected to is complex. For example it must be possible to discriminate between a query algorithm that can execute in real time and one that requires pre-indexing; however, the actual data available to make this decision may be context dependent; copyrighted content should be made accessible only as metadata, both for indexing and querying purposes, while large data sets should be referenced where possible in searching rather than transmitted in their entirety, but may be available for indexing purposes. Likewise overlay network decisions need to be made in relation to content and context.

At an interface level Alchemist concentrates on standards used in Web services and Web applications, both established, HTTP, SOAP, WSDL, Ajax; and emerging, WS-*, Comet; which will allow developers to integrate easily with the system. The infrastructure level can have many implementations, allowing Alchemist to span heterogeneous environments and include diverse hosts.

5.2 The AIT

The Alchemist Integrated Toolkit (AIT) is the keystone to the system, it provides a framework into which search and indexing algorithms and data content providers can be integrated. It also provides the high level tools with which application developers can build and deploy complex search algorithms, fusing different data sources and searches, onto the distributed Alchemist framework.

The AIT is used by application developers to simplify the integration of their custom search algorithms, content indexing and metadata generation into a single unified application that can be deployed onto the Alchemist framework. The AIT is accessed by users of the search engine to per-

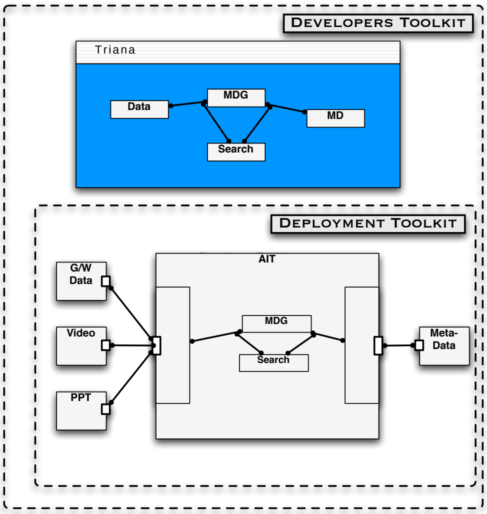

Figure 1: The AIT Toolkit

form smarter, more directed, and distributed searches for multimedia content or any other data for which the application developers have written modules. We envisage multiple clients able to access the AIT for users, from PC based application clients that may also participate as peers in the network to lightweight mobile clients that just provide a query interface and result retrieval mechanism.

The AIT has three main focus areas: *Digital Content Plug-Ins* provide the generic interface, and some specific implementations, for sources of digital content; *Plug-Ins for Index and Search Algorithms* provide the interface, and implementations, for content-specific indexing, metadata generation and searching; *Heterogeneous Data Fusion*data fusion *through Workflow* provides the fusion of content analysis from different media and algorithms through the use of component-based workflows.

At the core of the AIT are two toolkits, the *Developers Toolkit* and the *Deployment Toolkit*; in Figure 1 we can see the developers toolkit with an algorithm represented as a workflow of interconnected components, below is the deployment toolkit where the workflow is "pushed" out to the network, here multiple data source components (gravitational wave data, video and powerpoint files in this case) are processed by the workflow resulting in metadata that is propagated back into the distributed database. The Developers Toolkit is be based around the Triana workflow tool and provides the functionality to easily combine content, individual analysis and search tools into Alchemist's multimedia fusion searches and indexing; using workflow techniques. These search and indexing workflows can then be deployed onto the Alchemist framework. The Deployment Toolkit contains an execution engine that can run a workflow comprising a set of deployed content or search tools on the peers in the network. Data is pulled into the analysis from the local host machine via specific digital content plug-ins tailored to the media type. The metadata generated as a result of this content analysis can then be propagated throughout the network, providing a distributed and scaleable searching and indexing architecture.

Digital Content Plug-Ins provide the ability to access any kind of digital content ranging from Access Grid AV files to binary templates for gravitational wave data. The AIT architecture is designed to be extensible, allowing for the development of new plug-ins for different media types. When a new plug-in is written, it may be seamlessly propagated across the Alchemist framework, with little or no end-user intervention, to provide access to the new data types.

Plug-Ins for Index and Search Algorithms give the developers using the AIT an easy way to integrate disparate code and algorithms, that perform specific search or analysis tasks, into Alchemist. The AIT provides an interface to which plug-ins can be written and a simple mechanism to package and deploy them to the Alchemist framework.

Heterogeneous Data Fusion through Workflow is the ability to combine index and search plug-ins in new ways using the workflow paradigm to simplify the process. Fusing metadata into new forms is key to returning more exact search results from search engines, especially when searching complex multimedia data. Alchemist does this by combining content specific indexing and metadata generation with social indexing for relevance and quality. The AIT provides the melting pot in which data from social and content-based sources are mixed and then used in smarter searches.

Through the integration of digital content providers with indexing and search algorithms the AIT provides a flexible toolkit for performing searches. The data fusion through workflow aspect adds a powerful extra dimension to this, by allowing application developers to combine heterogeneous index and search algorithms in new ways, through the intuitive workflow paradigm, to create complex smart search mechanisms.

5.3 Metadata Searching

The AIT is based on a component architecture which allows chaining of algorithms into workflows as well as hierarchical composition allowing groups of algorithms to be subjected to further processing. Two classes of algorithms are defined, those that work directly with raw input data, and those that perform processing based on the output of other algorithms. Members of the first class of algorithm extract low-level features from raw data such as AV material. In terms of the visual content this includes measurements such as the amount of motion and the density of editorial cuts, and SIFT descriptors (Scale-Invariant Feature Transform), which are used to extract distinctive features from images which can then be used in algorithms to match different views of an object for object recognition. In terms of the audio content this may mean measurements such as voice pitch and recording soundtrack melody, timbre and rhythm features, speech velocity and acceleration, jitter, and other features drawn from standard audio analysis techniques such as LPC (linear predictive coding), MFCC (Mel frequency spectral coefficients).

These components represent the automated indexing capabilities of Alchemist, i.e. the classification of data according to data dependent primitives. Members of the second class of algorithm are based on higher level parameters and analyse the low level extractions made by other components to build more sophisticated views of the content. These higher level parameters allow the content to be mapped to the semantic requirements of domain specific query and search modules. In AV searches, two particular areas of research are the focus of of higher level algorithms; analysis of *affect-*

ive content and social indexing.

The affective content of a video clip or piece of music can be seen as the amount and type of affect (feeling, emotion, and mood) that characterises the multimedia content. A human affective response is usually represented on multiple dimensions or scales such as valence (type of emotion) and arousal. Arousal and valence values can be plotted together to generate a 2D affect curve, which can be seen as the most complete representation of the affective content of an AV file. This affect curve can be used by client modules to map their domain specific material against semantically defined requirements such as user preferences for news information or the extraction of significant events from sports or surveillance material.

6. CURRENT ALCHEMIST USERS AND SCENARIOS

The novel and flexible elements of Alchemist offers the opportunity to develop a number of new, sophisticated, highly desirable and yet simple media search strategies, which could allow the input of aspects of complex but common human knowledge in specifying a search.

One current project that is using the Alchemist framework is DART (Distributed Audio Retrieval using Triana) [14] that is being researched jointly between Cardiff University and the Laboratory for Creative arts and Technologies (LCAT) in Louisiana State University. DART capitalises on these Alchemist developments to provide a decentralised overlay for the processing of audio information for application in Music Information Retrieval (MIR). Audio analysis algorithms and frameworks for MIR are expanding rapidly, providing new ways to garnish information from audio sources well beyond what can be ascertained from ID3 tags. Modest successes have been made in audio-based musical genre classification [19][3], beat detection and analysis [8], similarity retrieval [18], and audio fingerprinting [10]. This work uses Short-Time Fourier Transforms to track the means and variances of the Spectral Centroid, standard deviations of the spectrum around its centroid, spectral envelopes, and signal power to represent sound textures, beat and pitch content [17]. These values are then transformed into attribute-value pairs for pattern matching and semantic retrieval. The analysis component of MIR requires extensive computational resources. Distributed environments and P2P networks are already being used for this purpose [20], and Alchemist's P2P architecture builds upon these developments, providing an advanced platform for developing, testing and deploying new search and analysis algorithms.

Under the DART scenario, users provide CPU cycles to provide metadata to the network by analysing the audio files located on their individual hard drives. The local processing on the network participants would be achieved though the use of Triana workflows to provide pipelines for processing the data in specific ways. Triana enables extremely rapid prototyping and code reusability of analysis algorithms exposed by a graphical interface, where units (representing audio analysis tools) can be dragged, dropped and connected together in order to create the algorithms. Such algorithms can be deployed i.e. uploaded to the peers in the network, through the use of the Alchemist decentralised P2P structures that would enable caching and file replication for speed and efficiency. The metadata created from the users would be stored across the network and used within specific searches such as searching for new music suggestions based on a user's current playlist i.e. in an "if you like this, then try this scenario", by performing statistical analysis on user trends.

Another more complex example could be for the searching of *sound effects* by a combined use of verbal and contextual information as well as acoustic properties. This could allow user requests automatically to utilise prior searches and subsequent classifications, to make sense of a complex search specification request; "a door less squeaky and somewhat heavier than this one" or a "thud with a swish at the end". There are also opportunities to create Human/Machine interfaces for such search engines which allow moving through N-dimensional space towards or away from sounds with specific characteristics once the data has been searched, analysed and structured. Music searches for professionals could be made in a similar way looking for similarities which are not necessarily verbal but could be defined spatially by multi-dimensional analysis and input of "musical hints" or "sounds-like" suggestions.

For film productions, *dialogue searches* could search amidst the sometimes hundreds of hours of recorded sounds that are often made during film shooting, for words, phrases or even sounds, without necessarily depending on a specific language, but optionally in conjunction with a film script input. Such opportunities to find characters' voices quickly for alternative takes can be a real time saver and enhance the overall editing performance.

7. CURRENT STATUS

The core framework is being built using the P2PS binding for WSPeer, which allows SOAP messaging with standardised Web services interfaces over ah-hoc super-peer networks. The WSPeer binding has been written and released [12] and we are currently working on the extensions for the caching overlays and the basic set of services, with a working prototype expected by the first quarter of 2007. In addition a mobile version of WSPeer is being developed that will enable the inclusion of mobile clients in the Alchemist network. We have already began simulations in collaboration with ICAR-CNR for the simulation of the decentralisation of data within BOINC-style networks [13]. A set of simulation runs were performed to evaluate the impact of the caching and replication mechanism on a set of performance indices, such as overall time to execute the jobs, throughput, mean time to download a data file, and load experienced by data centers and worker nodes. Results showed that indeed such a caching mechanism significantly improves the performance of such networks.

8. CONCLUSIONS

In this paper we have described Alchemist, a decentralised P2P framework for novel digital content analysis, search and retrieval. The framework based on open standards and proven technologies could be run on hardware from a mobile user device up to dedicated data and compute clusters, making the network truly ubiquitous. Caching mechanisms and network super-peers allow for network transience and partial failure, and the context aware nature of the algorithms can make use of metadata about DRM or data volumes before deciding whether to share digital content or not. Through the use of audio/visual examples we illustrated how Alchemist could be used and how it might be applied to different domains with different digital data content. Building on top of existing prototypes we hope to apply the Alchemist search paradigm to domains as diverse as e-Health and relativistic astrophysics.

9. ACKNOWLEDGEMENTS

Many thanks to the programming team working on various aspects of the Alchemist project including Eddie Al-Shakarchi, Tom Goodale, Andrew Harrison, Ian Kelley and Ian Wang. The Alchemist is built on previous work on P2PS, WSPeer, which have been funded in part by the EU (in Gridlab), PPARC (in GridOneD) and EPSRC (in EP/C006291/1).

10. REFERENCES

[1] G. Allen, K. Davis, K. N. Dolkas, N. D. Doulamis, T. Goodale, T. Kielmann, A. Merzky, J. Nabrzyski, J. Pukacki, T. Radke, M. Russell, E. Seidel, J. Shalf, and I. Taylor. Enabling applications on the grid: A gridlab overview. *International Journal of High Performance Computing Applications: Special Issue on Grid Computing: Infrastructure and Applications*, 17(4):449–466, November 2003.

[2] D. P. Anderson, E. Korpela, and R. Walton. High-Performance Task Distribution for Volunteer Computing. In *First IEEE International Conference on e-Science and Grid Technologies*, 2005.

[3] J. J. Aucouturier and F. Pachet. Representing musical genre: A state of the art. *Journal of New Music Research*, 32(1):83–93, 2003.

[4] D. Brookshier. BitVault Security. Technical report, 321 Products Inc.

[5] A. L. Chervenak. Applying Peer-to-Peer Techniques to Grid Replica Location Services. *Journal of Grid Computing*, to be published.

[6] I. Clarke, S. G. Miller, T. W. Hong, O. Sandberg, and B. Wiley. Protecting free expression online with freenet. *IEEE Internet Computing*, pages 40–49, January, February 2002.

[7] F. Dabek, E. Brunskill, M. F. Kaashoek, D. Karger, R. Morris, I. Stoica, and H. Balakrishnan. Building Peer-to-Peer Systems with Chord, a Distributed Lookup Service. In *Eighth Workshop on Hot Topics in Operating Systems*, page 0081, 2001.

[8] J. Foote and M. Cooper. Audio retrieval by rhythmic similarity. In *Proceedings of the International Conference on Music Information Retrieval*, volume Paris: IRCAM, pages 265–266, 2002.

[9] I. Foster, C. Kesselman, J. Nick, and S. Tuecke. The physiology of the grid: An open grid services architecture for distributed systems integration. *Open Grid Service Infrastructure WG, Global Grid Forum*, June, 2002.

[10] J. Haitsma and T. Kalker. A highly robust audio fingerprinting system. In *Proceedings of the Third International Conference on Music Information Retrieval*, volume Paris: IRCAM, pages 107–115, 2002.

[11] A. Harrison and I. Taylor. Wspeer - an interface to web service hosting and invocation. In *HIPS Joint Workshop on High-Performance Grid Computing and High-Level Parallel Programming Models*, 2005.

[12] A. Harrison and I. Taylor. WS-RF service and resource mobiltiy in a peer-to-peer context. *Accepted for publication to the Journal of Grid Computing*, 2006.

[13] D. T. Pasquale Cozza, Carlo Mastroianni and I. Taylor. A Super-Peer Protocol for Multiple Job Submission on a Grid. In *Proceedings of the CoreGRID Workshop on Grid Middleware Workshop in conjunction with Euro-Par, Dresden. To be Published.* Springer, August 28-29 2006.

[14] I. Taylor, E. Al-Shakarchi, and S. D. Beck. Distributed Audio Retrieval using Triana (DART). In *International Computer Music Conference (ICMC) 2006, November 6-11, at Tulane University, USA.*, 2006. to appear.

[15] I. Taylor, I. Downard, B. Adamson, and J. Macker. Agentj: Enabling java ns-2 simulations for large scale distributed multimedia applications. Penang, Malaysia, 14th to 17th May 2006. Second International Conference on Distributed Frameworks for Multimedia DFMA 2006.

[16] I. Taylor, M. Shields, I. Wang, and O. Rana. Triana Applications within Grid Computing and Peer to Peer Environments. *Journal of Grid Computing*, 1(2):199–217, 2003.

[17] G. Tummarello, C. Morbidoni, P. Puliti, and F. Piazza. Semantic audio hyperlinking: a multimedia-semantic web scenario. In *Proceedings of the 1st International Conference on Automated Production of Cross Media Content for Multi-channel Distribution*, pages 111–115, 2005.

[18] G. Tzanetakis and P. Cook. Beyond the query-by-example paradigm: New query interfaces for music information retrieval. In *Proceedings of the 2002 International Computer Music Conference*, pages 177–183, 2002.

[19] G. Tzanetakis and P. Cook. Music genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, 10(5):293–302, 2002.

[20] G. Tzanetakis, J. Gao, and P. Steenkiste. A scalable peer-to-peer system for music information retrieval. *Computer Music Journal*, 28(2):24–33, Summer 2004 2004.

[21] I. Wang. P2PS (Peer-to-Peer Simplified). In *Proceedings of 13th Annual Mardi Gras Conference - Frontiers of Grid Applications and Technologies*, pages 54–59. Louisiana State University, February 2005.