

Jun 21st, 3:30 PM - 4:30 PM

Data mining, data fusion, and libraries

Katherine Chiang
Cornell University, ksc3@cornell.edu

Follow this and additional works at: <http://docs.lib.purdue.edu/iatul2010>

Katherine Chiang, "Data mining, data fusion, and libraries" (June 21, 2010). *International Association of Scientific and Technological University Libraries, 31st Annual Conference*. Paper 4.
<http://docs.lib.purdue.edu/iatul2010/conf/day1/4>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

IATUL 2010

Data Mining, Data Fusion and Libraries

Katherine Chiang, Head of Services for Academic Programs, Mann Library
Cornell University Library ksc3@cornell.edu

INTRODUCTION

AT&T has a dataset of several hundred million telephone numbers and their call connections. They create massive network graphs of the time of call and the number called then analyze the calling relationships for each number and extract calling patterns using a process called *data mining*. AT&T uses those patterns to identify potential customers for specific services or customers whose calling patterns match those of customers who repeatedly fail to pay their phone bill. (AT&T Research Labs 2010)

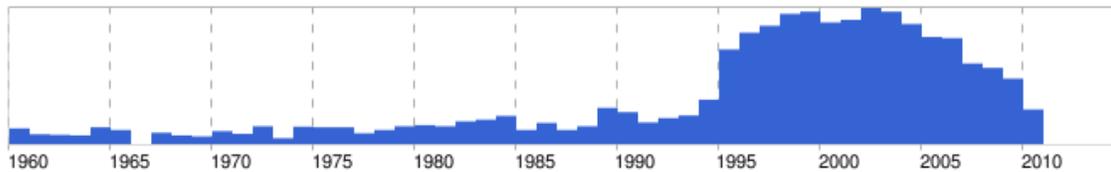
International ports and their harbors are one of the hardest to monitor entry points into a country. The traffic is complex, involving boats, containers, trucks, packages, and people. The Port of Callais processes over twelve thousand vehicles each day. The records to monitor the facility are equally complex. Every object has records associated with it. People have passports, visas, or identification cards. Containers, packages, or vehicles have bills of lading, or other records identifying the source and destination of the object. The records are from many countries and in many languages. Additional data come from sensors, video cameras, or imagery data from flights over the harbor. Managing and extracting useful information from all these data is a challenge. Modern ports are turning to *data fusion* techniques to manage and process these widely disparate sources of data to expedite traffic, and identify smuggled goods or illegal immigrants. (Shahbazian 2005)

Our computerized world churns out data and their analysis is a challenge. Data mining and data fusion are two complementary approaches to processing dynamic, large and/or heterogeneous data. They are already used in various research disciplines and real-world applications and that use will grow.

This paper describes and defines these approaches using examples, and suggests possible services and applications by libraries. The focus is on eliciting the common information concepts that might be of interest, not on understanding the details of the computer/information science work, or the validity of the approach in solving a particular research or commercial problem.

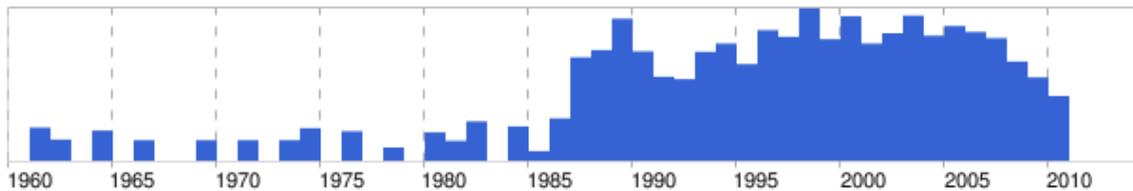
Data mining and data fusion are fuzzy terms, of recent origin. A Google Timeline display yields the following graphic:

1960-2010 [Search other dates](#)



Google search on "Data Mining"

1960-2010 [Search other dates](#)



Google search on "Data Fusion"

Other terms are used for the same broad concept. Instead of data, authors use text, information or graph. Instead of mining or fusion they use mashup, consolidation, integration, aggregation, or merged.

How are data mining and fusion different from earlier computational data manipulations? The differences are incremental, not distinct. But the increment is as significant as the transportation jump from the horse to the car. Data mining and fusion can handle larger amounts, greater complexity, and dynamic or streaming data.

DATA MINING

Definition.

Data mining is a computational process that extracts patterns that may be significant. It is used when one has lots of data and has the 'don't know where to start' issue. Data mining looks at the data in relationship to the other data within the collection which differentiates it from classic data retrieval where the system is responding to the external criteria of the search term.

Examples.

In the 1850s an enterprising businessman began selling clippings of theater reviews in London's West End. The first clipping service was born and developed into a robust service used by businesses and any organization needing to know how they were being described in the media. Clipping services are now *media monitors* and they have substantially updated their techniques. The Nielsen Company offers a media monitoring service called Buzzmetrics. They monitor 100 million consumer

generated media sites: online communities, discussion boards, blogs, and social networks. The service and products are proprietary, but a sample of what they can deliver is at <http://www.blogpulse.com/>



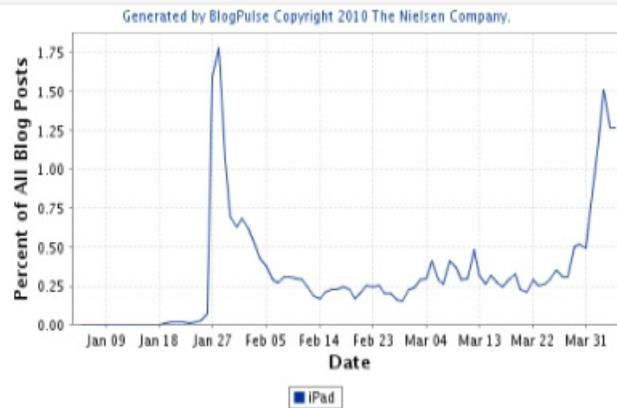
Discover What's Hot...

TODAY'S HIGHLIGHTS



Customers Swarm Apple Stores for iPad

The act of buying the iPad on Saturday creates nearly as much buzz as its announcement on Jan. 27. (4/5/2010)



The service is advertised for those interested in “online reputation management”, or “brand health metrics”. They can identify ‘reputation threats’, daily and trend status, and competitors. Their customers include advertisers, pharmaceutical companies, banks, and the food and telecommunications industries. Cision (the original clipping company, considerably morphed in the intervening years) and Lexis/Nexis advertise similar products. (Cision 2010) (LexisNexis 2010)

These companies use data that available on the Web. If the Blogpulse page is an example of Nielsen’s subscription product, they give their clients graphic visualizations of the data, and probably a series of reports.

A second example of data mining comes from the genomics arena. High speed gene analysis has totally outstripped researchers’ abilities to analyze the data. The response has been an approach called ‘hypothesis free’, or ‘in silico’ analysis. Data mining is a core technology for that approach. Various data mining techniques look for patterns that identify areas for further study.

One approach treats the free text of the PubMed abstract as the data unit, sometimes using co-citation as an additional dimension. The authors look for the co-mention of genes using natural language processing and a computational algorithm that identifies genes that may share a common biological function. (Jenssen 2001) Another group uses a slightly different approach to the same goal. They validate

their process by running the algorithm against known gene associations to see if their process correctly identifies those associations. (Raychaudhuri 2003)

Concepts.

These examples highlight the central role of the *algorithms*. The commercial products are proprietary and generally they are created specifically for the subject or type of data being mined. In the case of the genomics data, the algorithms are created by computer scientists or bioinformaticians and can be invisible to the biologists doing the actual research on the gene associations of interest.

DATA FUSION

Definition.

The Joint Directors of Laboratories (JDL) authored the most frequently cited data fusion definition. It was first developed in late '80s. Data fusion was described as: "a process dealing with the association, correlation, and combination of data and information from single and multiple sources to achieve refined position and identity estimates, and complete and timely assessments of situations and threats, and their significance. The process is characterized by continuous refinements of its estimates and assessments, and the evaluation of the need for additional sources, or modification of the process itself, to achieve improved results." That definition was subsequently shortened to: "Data fusion is the process of combining data to refine state estimates and predictions." (Steinberg, 1999)

Bostrom reviewed published definitions between 1987 and 2007 and crafted the following synthesis: "Information fusion is the study of efficient methods for automatically or semi-automatically transforming information from different sources and different points in time into a representation that provides effective support for human or automated decision making." (Boström, 2007)

Examples.

Initially data fusion referred to the merging of somewhat homogeneous data. Examples of this level of data fusion are sensor instrumentation or photogrammetric data. (Seinturier 2006) Another example of this sort of fusion is the process used in coastal zone studies to merge datasets coming from the Hydrographic Airborne Rapid Total Survey to produce "higher-level information products in GIS format." (Wozencraft 2007) The subsequent examples are of applications that push the technology and use heterogeneous, dynamic, and complex data.

Intelligent Transportation Systems collect incredible volumes of data. Every vehicle with a cell phone or GPS device is a data producer. Those devices generate what is called 'floating car data'. Data collectors triangulate on cell phone locations and/or

pick up when they are 'handed over' to the next tower on their route. (Dubbert 2007) Sensors in freeway systems collect velocity, location and time data for passing vehicles. Traffic signals, counters and video cameras capture additional data.¹ There are datasets for roadway speed limits, interchanges, and road types. One European dataset has 50 million road segments. (Schäfer 2009)

Advanced Traffic Management Systems (ATMS) and Advanced Traveler Information Systems (ATIS) take these data and use data fusion techniques to generate useful information. The ATIS company TOMTOM provides a subscription traffic update service. "TomTom HD Traffic uses a revolutionary new source of traffic information: the traffic flow of up to 16.7 million anonymous mobile phone users on the road. From this anonymous data, TomTom knows exactly where, in which direction and at what speed all these mobile phone users are travelling throughout the road network. This real-time data is combined with other existing quality traffic information sources, resulting in the most complete and reliable traffic information." (TomTom 2010)

ATMS use fused data to make on-the-fly changes to traffic signals and other traffic control systems including emergency vehicle routing. ATMS can change signals to give emergency vehicles right of way through traffic. At a larger scale, ATMS are being developed for use during large scale disasters such as earthquakes. Traffic and road condition data are combined with information such as hospital loads so drivers transporting the injured can take them to the nearest facility with available beds. (Friedrich 2004) (Shawe-Taylor n.d.) (Hellinga 2000) (Jotshi 2009)

This example illustrates the temporal aspect of this type of information. Data fusion is being used for real world decision making and much of the data being fused is *dynamic*. If the time scales are as small as the 25 minutes of the drive home, the technologies processing the data need to be able to manage volumes of data extremely rapidly.

Data fusion is used in the life sciences: in personal health, genomics, and telehealth. Each person has the potential to accumulate massive amounts of data over the course of a lifetime in the modern health care system. Data from patient tests results range from numeric and image data, to descriptive (patient is a smoker, has moderate alcohol consumption) to acoustic recordings of doctor's examination notes. The data are composed of multiple attributes, at differing scales, come from different underlying disciplines, and span the life of the individual. In 5 days one thousand patients can generate 1 gigabyte of data. Each patient can have over one thousand attributes measured in the course of a year. (Tsumoto 2003) Data fusion is one approach to integrating all the data about an individual patient. It can also contribute to building the *physiome*. In this context, the *physiome* is defined as the comprehensive description of the functional behavior of humans through the integration of information about genomes, cell systems, organs, and organisms.

¹ see <http://service.nyctmc.org/mobile/> for an example of a video camera feed.

Current genomics initiatives are fusing data from disparate sources with the goal of building the physiome. One ambitious project is fusing data from literature abstracts, functional annotations, microarray expressions, EST expression data, protein domains, protein-protein interactions; pathway membership, transcriptional motifs, and sequence similarities. (Aerts 2006)

Telehealth, or telemedicine, are the terms used to describe technologies being developed to enable a person with health care needs to stay in their own home for as long as possible. Telehealth monitoring technologies (motion, pressure, and other sensors) are installed in the home and track vital signs and the patterns of daily living: meals, bathroom visits, and sleep cycles.

The data they collect are fused and analyzed to create signature patterns that describe each individual's routine. Violations to the routine generate messages to the occupant to take their medication, or to eat, or send a message to a third party service to check on the safety of the individual. (Alwan 2005)

These examples introduce several information issues. The personal health data and telehealth examples bring in the issues surrounding *privacy* and ownership of personal data. The physiome work illustrates how deeply the *expertise* of the domain is embedded in the fusion process. That work also requires the fusion of data across a huge range of *scales of observation*, from the molecular, to the organism, to the environment. Each of those scales is a fast moving discipline. For the physiome to succeed it will be in a continual state of revision, at all scales of observations, and between observational scales.

LIBRARIES

Libraries could be involved throughout the continuum of the process from the back end (the information miners and fusers) to the front end (the consumers of the information.) Researchers using these approaches work throughout the continuum; they do both the mining and fusing processes and the examination of the results. In the commercial sector there is a more distinct split. The company does the back end work and sells it to the consumer. There is a similar split in government information, most data users are not data manipulators.

The following sections summarize the information issues and discuss the possible library connections. The back end processes are: finding data; finding or building the tools (the algorithms, workflows and data pipelines); doing the actual data process (the mining and fusing); documenting the process; and keeping the whole process current (updating the data, tools, metadata, etc.). At the front end are the use of the processes data, and the use of the technique by libraries.

Data:

The mining or fusing approach has potential application anywhere multiple sources of large amounts of proprietary or free data are available. Many datasets are continuously generated (sensor and other streaming data). Large amounts of data are used in data mining in order for the patterns to be statistically significant while data fusion can be any size. Any processes using dynamic data will have to be updated regularly. Privacy is an issue for any data on individuals. As more data are distributed, and mining and fusion techniques allow researchers to use a wider range of data, libraries could help researchers find data, and alert them to privacy issues.

Tools:

The algorithms identifying the patterns or significances can be predefined, or a product of the data processing. The expertise of the domain is embedded in the process. Subject expertise is required for the selection of the variables as well as the subsequent manipulations of the data. How the algorithm does that identification could be invisible to most consumers of the mined or fused data.

However, the tools could be transferable. An extinction researcher used the Google page rank algorithm to create a model for species extinctions. (Allesin 2009) Other researchers have looked to the biological immune system to create artificial immune systems (AISs) for improvements in computational problem solving. (Twycross 2010) There are efforts within disciplines to catalog the tools. (National Center for Biotechnology Information 2010) (The Arabidopsis Information Resource 2010) Libraries could help researchers find the tools across disciplines. We could find ways of 'cataloging' the algorithms, somehow abstracting them from their content and describing their 'function' so they could be used by another discipline. Perhaps an ecosystem researcher could use, with a few tweaks, an algorithm developed for a harbor monitoring system.

Companies are selling data fusion products. *StreamBase* sells a Complex Event Processing (CEP) product. "a technology for low-latency filtering, correlating, aggregating, and computing on real-world event data. (StreamBase 2010) Their approach is direct processing of the streaming data, unlike classical processes that store the data first. Microsoft has a Fusion Framework and has teamed with ESRI to enable spatial display and analysis. "The ArcGIS Situational Awareness Bundle included in the Fusion Framework provides a complete framework for geospatial intelligence and decision-making." (Microsoft 2010) Perhaps libraries could work with IT units to buy the commercial applications. Eventually the library's workshop series might include: "Data Fusion: Introduction to Applications."

Data producers are developing standards and there are data coordination efforts such as the International Ship and Port Facility Security Code. (International Maritime Organization 2010) The library could help in the standards area. Data fusion initiatives are using semantic approaches to process heterogeneous data. (Boury-Brisset 2003) Libraries are already engaged in semantics ontology development and maintenance, perhaps there are opportunities for us to help here.

There are other data activities where the library could participate. Perhaps we could be partners in projects using mining and fusion methodologies, perhaps in infrastructure projects, such as the Data Mining Grid. (Stankovski 2008)

The library could help with provenance issues. The data trail for a fused data project could be extremely complex. If any of the data are proprietary the copyright and other legal issues could make the Google book process look trivial. In the intellectual property arena they refer to 'patent thickets' and 'royalty stacking.' For example, the Malaria Vaccine Initiative identified a protein antigen possibly critical for a vaccine. A patent thicket of 39 patent families were identified as potentially involved with that antigen. If each each of the 39 patent holders asked for a royalty that royalty stack could exceed the royalties earned by the vaccine. Similar issues could arise in information use. The library might be able to help researchers acquire proprietary data, but would probably refer legal and licensing issues to the university rights office.

Using the data.

Will the products of data mining and data fusion require new library services? Will it be important for patrons to understand where the fused data come from and the complexity of the data input stream? Or will data mining and data fusion become mainstream data manipulation techniques – like linear or spatial statistics? Perhaps the library will need to include data fusion in its literacy instruction.

The products of the mining and fusion can be a piece of information (hospital A has a free bed and roads to it are obstruction free), or reports (product A was mentioned favorably in the last month), or articles, or databases. It is likely libraries will subscribe to commercial databases. Some of the current business and financial databases are probably fusion products. Most non-profit research databases are free, thus research data fusion databases may be free. However, the sustainability of research data is currently under study. Therefore the long term costs of access to research data are unknown.

The evanescence of the fused data creates documentation challenges. Currently many research articles and reports link to their data. But they point to a static dataset. Fused data are used at a point in time. But as the source data change continuously there is no way to 'recreate' those data. How do you cite those data? An additional complication is the data may be displayed in novel ways, such as an animated visualization. Lynch describes the difficulties in documenting this combination of data, software and visualization as the "pressure on the traditional scientific record" (Lynch 2009) (Lynch, 2007)

The cognitive process of extracting meaning from data visualizations is called visual analytics. (Thomas 2005) Librarians often work with faculty to teach sessions on how to 'read' a scientific paper. Will we also help students acquire the skills needed to 'read' visualizations?

Libraries *Using* the technology.

A final suggestion for library involvement, and one we have begun: we should promote the *use* of data mining and fusion techniques on *our* information resources. We offer a confusing array of heterogeneous information resources. Could data mining and fusion approaches solve some of our information delivery challenges? Are Summon and other discovery products data fusion approaches? Endeca describes their products as data mining and the North Carolina State University Library uses Endeca. (North Carolina State University Libraries 2010) Collexis advertises their data mining capabilities. (Collexis 2010) We have data about our patrons and our systems. Could we mine our library systems to better understand our patrons' behaviors?

Data fusion could be a way to improve personalized information retrieval. If we could fuse the disparate databases researchers use in literature searches we could improve the information retrieval process. The literature in relevance research describes several relevance criteria that are dependent on individual differences such as novelty (whether the article author or content is new to the researcher) and time constraints (article or grant proposal due dates.) (Barry 1998) A catalog search should be fused with circulation and personal citation database (such as Zotero or EndNote). Instead of seeing: "item checked out, due 4-15-2010" it should say "you have this item." The literature databases should flag items already in the researchers' citation database, and the system could incorporate personal e-mail lists, so if the article's author was in that list it could flag the article "you have corresponded with this author". To alleviate the time constraint function we could fuse a calendar of deadlines with the literature search. With the due date for a paper being 4-19-2010 the researcher would be given more focused hits as the deadline approached.

CONCLUSION

Data mining and data fusion could help tame the data deluge. The techniques are proliferating and applied to more and more datasets. Our researchers will be, or are already, using these approaches and our students will be using the results. Libraries and commercial information vendors should, and will, use these data mining and fusion techniques to enhance the delivery of heterogeneous information. It will be interesting to see how this all develops.

REFERENCES

- Aerts, Stein et. al. "Gene prioritization through genomic data fusion." *Nature biotechnology* 24, no. 5 (May 2006): 537-544.
- Allesin, Stefano, Mercedes Pascual. "Googling Food Webs: Can an Eigenvector Measure Species' Importance for Coextinctions?" *PloS computational Biology* 5, no. 9 (September 2009). doi:10.1371/journal.pcbi.1000494 Accessed 4-22-2010

Alwan, Majd et. al. "A Rule-Based Approach to the Analysis of Elders' Activity Data: Detection of Health and Possible Emergency Conditions." *Telemedicine and e-Health* 11, no. 5 (October 2005): 594-599.

AT&T Research Labs. *Graph Mining*.

<http://www2.research.att.com/~stat/research/graphmining.php> Accessed 4-21-2010, 2010.

Barry, Carol L., Linda Schamber. "User's criteria for relevance evaluation: a cross-situational comparison." *Information Processing and Management* 34, no. 2-3 (1998): 219-36.

Boström, Henrik, et. al. *On the Definition of Information Fusion as a Field of Research*. Technical report, University of Skövde, School of Humanities and Informatics <http://his.diva-portal.org/smash/get/diva2:2391/FULLTEXT01>, 2007 Accessed 4-21-2010.

Boury-Brisset, Anne-Claire. "Ontology-based Approach for Information Fusion." *Sixth International Conference on Information Fusion*, 2003. 522-529. <http://www.isif.org/fusion/proceedings/fusion03CD/special/s31.pdf> Accessed 4-22-2010.

Collexis. www.collexis.com/products Accessed 4-23-2010, 2010.

Dubbert, Jörg, Urban Stenberg,. "Approaches to Floating Car Data Collection in the VIKING Area." *i2TERN 2007*. Aalborg:, 2007 <http://i2tern.plan.aau.dk/doks/paper/paper107.pdf> Accessed 4-21-2010.

Friedrich, Bernhard. "Traffic Monitoring and Control in Metropolitan Areas." *2nd International Symposium "Networks for Mobility"*. Stuttgart, 2004 <http://www.ivh.uni-hannover.de/peb/de/Mitarbeiter/friedrich-Dateien/veroeffentlichungen/FOVUS%20-%20Traffic%20Monitoring%20and%20Control%20in%20Metropolitan%20Areas.pdf> Accessed 4-21-2010.

Hellinga, Bruce, Rajesh Gudapati. "Estimating Link Travel Times for Advanced Traveller Information." *Annual Conference Canadian Society for Civil Engineering (CSCE) 2000*, 2000. <http://www.civil.uwaterloo.ca/bhellinga/publications/Publications/CSCE-2000-Data%20Fusion.PDF> Accessed 4-21-2010.

International Maritime Organization. 2010 <http://www.imo.org> Accessed 4-22-2010.

Jenssen, T.K., A.Laegried, J.Kmorowski and E.Hovig. 2001. "A literature network of human genes for high-throughput analysis of gene expression." *Nature Genetics* 28, no. 1 (2001): 21-28.

Jotshi, Arun, Qiang Gong, Rajan Batta. "Dispatching and routing of emergency vehicles in disaster mitigation using information fusion." *Socio-Economic Planning Sciences* 43, no. 1 (March 2009): 1-24.

Lynch, Clifford A. "Jim Gray's Fourth Paradigm and the Construction of the Scientific Record." In *The Fourth Paradigm. Data-intensive scientific discovery*, by Tony Hey, Stewart Tansley, Kirstin Tolle. Redmond: Microsoft, 2009.

Lynch, Clifford A. "The Shape of the Scientific Article in The Developing Cyberinfrastructure." *CTWatch Quarterly* 3, no. 3 (August 2007)
<http://www.ctwatch.org/quarterly/articles/2007/08/the-shape-of-the-scientific-article-in-the-developing-cyberinfrastructure/> Accessed 4-23-2010

Microsoft. 2010 <http://download.microsoft.com/download/1/5/C/15C33BAA-1CEE-431A-A8C0-8D521BDA6B9A/FusionFrameworkBDMDDeckFinal.pdf> Accessed 4-22-2010

National Center for Biotechnology Information. 2010
<http://www.ncbi.nlm.nih.gov/guide/data-software/> Accessed 4-22-2010

North Carolina State University Libraries. 2010 <http://www.lib.ncsu.edu/endeca/> Accessed 4-20-2010

Raychaudhuri, S R.B. Altman. "A literature-based method for assessing the functional coherence of a gene group." *Bioinformatics* 19, no. 3 (2003): 1-6.

Schäfer, Ralf-Peter. "IQ routes and HD traffic: technology insights about tomtom's time-dynamic navigation concept." *Foundations of Software Engineering* 2009. 171-172. <http://doi.acm.org/10.1145/1595696.1595698> Accessed 4-21-2010

Seinturier, Julian, Pierre Drap, Odile Papinis, Guido Vannini, Michele Nuccioti. "Knowledge Representation and Data Fusion for Archaeology: The case study of the castle of shawbak." *3IA International Conference on Computer Graphics and Artificial Intelligence*. 2006
http://3ia2008.teiath.gr/3ia_previous_conferences_cds/2006/Papers/Full/seinturier20.pdf Accessed 4-21-2010

Shahbazian, Elisa, Galina Rogova, Michael J. DeWeert. *Harbour Protection through Data Fusion Technologies*. Dordrecht: Springer, 2005.

Shawe-Taylor, John, Tijn De Bie, Nello Cristianini. *Data mining, data fusion and information management*. Foresight Project on Intelligent Infrastructure Systems, London: n.d.
http://www.foresight.gov.uk/Intelligent%20Infrastructure%20Systems/Data_mining.pdf Accessed 4-21-2010

Stankovski, Vlado et al. "Digging deep into the data mine with DataMiningGrid." *IEEE Internet Computing*, November/December 2008: 69-76.

Steinberg, Alan N., Christopher L. Bowman, Franklin E. White. "Revisions to the JDL Data Fusion Model." *Sensor Fusion: Architectures, Algorithms, and Applications*. SPIE 1999 <http://www.dtic.mil/cgi-bin/GetTRDoc?Location=U2&doc=GetTRDoc.pdf&AD=ADA391479> Accessed 4-21-2010

StreamBase. 2010 <http://www.streambase.com/complex-event-processing.htm> Accessed 4-22-2010

The Arabidopsis Information Resource. 2010 http://www.arabidopsis.org/portals/metabolome/metabolome_tool.jsp Accessed 4-22-2010

Thomas, James J. and Kristin A. Cook. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. National Visualization and Analytics Center 2005 <http://nvac.pnl.gov/agenda.stm> Accessed 4-23-2010

TomTom. 2010 <http://www.tomtom.com/services/service.php?id=2&tab=4> Accessed 4-22-2010

Tsumoto, Shusaka. "Discovery of Temporal Knowledge in Medical Time-Series Databases using Moving Average, Multiscale Matching and Rule Induction." In *Information fusion in data mining*, by Vicenç Torra, 449-459. New York: Springer, 2003.

Twycross, Jamie, Uwe Aickelin. "Information Fusion in the Immune System." *Information Fusion* 111 (2010): 35-44.

Wozencraft, Jennifer M., Christopher L. Macon, , W. Jeff Lillycrop. "CHARTS-enabled Data Fusion for Coastal Zone Characterization." *Coastal Sediments '07 Proceedings of the Sixth International Symposium on Coastal Engineering and Science of Coastal Sediment Processes*. American Society of Civil Engineers 2007 [http://dx.doi.org/10.1061/40926\(239\)143](http://dx.doi.org/10.1061/40926(239)143) Accessed 4-21-2010