

Information and Communicative Fusion

Kalev Leetaru
September 6, 2006

My research interests lie in the intersection of distributed information and communications systems and the techniques and technologies used to bring them together. This can involve the use of sophisticated information warehouses, Unified Communications infrastructures, and *hypermediums*, situated around common themes of analysis, shared attributes, or communicative modalities.

Communicative or information fusion may be considered broadly as the intersection of a collection of information sources for the purposes of leveraging their collective utility. In communicative fusion, media may be combined in place (Unified Communications) or subsumed into a new media (a hypermedium), allowing a user's assemblage of communicative address to be coalesced into a single "smart" address that automatically routes messages via the appropriate channels. Information fusion involves the compilation of disparate data sources for the purposes of interacting with the data as a single source. While initially developed in the enterprise world, information fusion has become commoditized through the development of consumer products such as Google Desktop.

Information fusion has long been a staple of the enterprise world in the form of data warehouses. These warehouses aggregate data from numerous sources and present it to applications under a single interface, allowing all of the data to be considered together, no matter where each piece originated. A casino, for example, might use its data warehouse to integrate information from every corner of its casino floor with its dining and shopping facilities, and analyze this data in real-time to identify "high rollers." The field of enterprise data fusion is a rapidly growing one, with companies realizing the potential, and driving new technology development:¹

Companies are trying to absorb more information faster and from more far-flung sources so they can understand and adapt to customers' needs more quickly. Yet data remains fragmented at most companies, scattered across enterprise-resource-planning and customer-relationship-management applications, legacy systems, and disparate databases.

While the business world is dominated by large enterprise data warehouses, the consumer world has jumped on the bandwagon with tools such as Google Desktop that use information fusion for search. Email, word processing and spreadsheet documents, instant messenger transcripts, email archives, and other normally disjointed content from a user's machine are brought together into a single data source so that they may be searched in unison, allowing a user, for example, to discover an IM session that mentioned a particular topic, which would have been missed in a standard search of

¹ Integrate or Disintegrate. InformationWeek. Rich Whiting and Charles Babcock. March 21, 2005.

desktop productivity files on the computer. The latest version of Google Desktop, Beta 3, takes this fusion past the threshold of a local user's desktop, and allows users to share their search repositories with Google's central servers, and make those repositories available for search to anyone they chose on the Web. A grass-roots organization, for example, can allow its members to search all business documents and correspondence across all of its member's computers. While large corporations have long used enterprise technology to achieve similar functionality, the commoditized nature of Google Desktop is bringing this kind of information fusion to the masses.

Information fusion has also found a nontraditional role in communications. Rather than bringing together large amounts of information for the purpose of search or analysis, the field of Unified Communications brings together a collection of communications technologies for the purpose of making communication more oblivious to a user's location and communicative modality. These systems work by bridging the gaps between media such as instant messaging, email, voice, fax, and cellular text messaging. In a traditional communicative environment, the initiator of the discourse is responsible for manually traversing the universe of available communications modalities. For example, when attempting to reach a colleague at work, a user might first call that person's office phone number. When that number reverts to voicemail, the user might then try the person's cell phone, mobile IM, and then finally email. Unified Communications systems make use of technologies such as Find Me, Follow Me (FMFM), in which the system "implements a user-configurable list of prioritized contact points to try before going to voice mail when the phone isn't answered."² The unification of a number of communications media together in a single infrastructure enables greater abstraction between message and medium. Similarly, if a fax arrives for a user who is traveling and only available by cellular phone, the system can convert that fax into text and read it aloud to the user. Finally, having a centralized infrastructure makes it easy to archive communications occurring in all of these media into a centralized corporate datacenter.

While Unified Communications seeks to bridge established communications infrastructures in enterprise settings, Instant Messaging is slowly subsuming them into a single medium, representing an emergent form of communications fusion that is resulting in the creation of a *hypermedium*. IM traces its roots back to 1973's PLATO TERM-talk with its text-only "blind calling," but the technology has since evolved in leaps and bounds, with software providers integrating mobile capabilities, audio and video streaming, and file sharing into their offerings. As IM continues to evolve and subsume the capabilities of new media, it provides strong support for those who claim that it is becoming the centerpiece of the new digital society, all the while providing a poignant example of consumer benefit of communicative fusion.³

The field of geography has long led the charge for information unification, bringing together disparate data sources around the common attribute of location. While maps often use multiple geographic data sources to construct a cohesive view of an area's

² "Get The Message Out". Network Computing / InformationWeek. Sean Doherty. October 24, 2005.

³ *A Hypermedium in the Making*. Unpublished whitepaper. Kalev Leetaru. December 9, 2005.

geography, the 1850 publication of *Atlas to accompany the second report of the Irish Railway Commissioners*, solidified the notion of multifaceted information fusion by merging population and traffic data with more traditional topographic sources.⁴ By superimposing a description of man-made behavior over its geologic host, correlations between the natural and man-made worlds could be observed more clearly. The simultaneous consideration of disparate data sources unified around the common attribute of space lies at the heart of Geographic Information Systems (GIS).

The Canada Geographic Information System (CGIS) is recognized as the first true GIS system. Developed in the 1960's to manage and analyze data collected for the Canada Land Inventory (CLI), CGIS pioneered many of the modern GIS techniques of layering data from multiple sources around a common attribute of geospatial information. The founding of Environmental Systems Research Institute (ESRI) in 1969 and its release of Arc/Info in 1982 profoundly influenced the GIS field, bringing these tools to the desktop.

The use of GIS as an example of information fusion is a particularly important one, in that fields such as the humanities, social sciences, and arts, have found established and emergent uses of this technology to unify their understanding of research questions in their fields. History in particular has begun to embrace the use of GIS in exposing unexpected patterns in historical events. For example, I am involved in a model digital history article, in which we are applying a number of information fusion technologies such as GIS to trace the history and evolution of the East St. Louis area. In a particularly poignant example of the power of this technology for exposing patterns, when I was combining various data layers looking for trends explaining the location of landfills in the East St. Louis area, I established a correlation between the racial demographics of each census tract in the area and the number of landfills it contained. This was a rather surprising find, in that it corroborated several other pieces of evidence we had, but the most surprising find of all came when I combined elevation data with the landfill layer, to discover that East St. Louis had the only major cluster of landfills in major municipality in a floodplain in the State of Illinois. This finding has significant implications for examining the city in the context of environmental influences.

I am also involved in a number of other projects across campus involving the unification of disparate data sources and the treatment of those unified data stores for analysis. For my dissertation research, I am interested in examining the multitude of uses of information fusion and the ways in which information resources can be brought together while still preserving the context of domain-specific knowledge they arise from. There has been research in the enterprise world, using elaborate "business intelligence" infrastructures to restore context to data in information warehouses and imbue them with deeper semantic descriptions. I am interested in exploring these kinds of concepts further, especially within the context of scholarly research, and the ways in which information and communicative fusion technologies can advance scholarship in Computing in the Humanities, Arts, and Social Sciences (CHASS) arenas.

⁴ http://www.gesource.ac.uk/timeline_Maps.html