Response to NITRD RFI
White Paper to Inform the Five-Year Strategic Plan
Collective Intelligence

Internet search technology has created an environment in which any user has access to nearly unbounded amounts of information about any selected topic. Two important aspects of search are that (1) the search engine should consider the largest possible collection of relevant information, and (2) the search engine should select the best match of points in this information space with the search target. Secondarily, the search recommendations should also be of "high quality" (meaning it is significant, verifiable, factual, and correct content). Generally speaking, after searching for information each user has to vet the information recommended by a search engine to determine its *relevance* with respect to then intent of the search, and to evaluate the *veracity* of the information.

The overhead work of vetting the information – both with respect to its relevance and veracity – can be significant. Any such vetting work can be abstracted and saved as *persistent state* in the user's search environment, e.g., bookmarks of locations are persistent state representing the location of pertinent & reliable information. Bookmarks effectively store this state in a simple hierarchical name space; bookmarks can be saved in a root node, or in child subdirectories (folders). For example, the user might be a university teacher, a software researcher, a member of the Chandler family, a tourist, a chess player, a John Mayer fan, and a luthier: the user can create a folder for each of these interests in which to store topic-specific bookmarks.

While it is not commonly done, it would be possible for the system to exploit this state, including its structure, to improve the search function. First, assume that the folder hierarchy is organized so that each folder represents one of the user's *perspectives* of the web (the persistent *substate* associated with each folder represents information about a particular view of the web as, say, a luthier). This substate information can be exploited by having the user select a perspective prior to searching; the overall search process (the search engine, possibly with supplementary software) then uses this filter to locate and discriminate among information normally returned by a search engine.

Bookmarks are only one example of the information that might make up a perspective state, and search is only one example of the NIT capability that uses such information. Any kind of information that the user/system can collect, encode, organize, and save can be part of a perspective state. Such generalization enables the system to use the state as the basis for broader tasks. For example, information stored in cookies could be part of a perspective state; social network relationships can be part of the state; and information extracted from RSS feeds (e.g., photos and videos that friends favor) could be saved in the state. The logical extent of the state is bounded by the nature of the data collection tools, the data organization methods, and the nature of the data analysis algorithms.

Some of these perspectives are unique to the user, but others are likely to be valuable to other users that have similar interests. That is, users with similar perspectives can leverage their effort to refine their perspective by collaboratively constructing and sharing state. The exemplar user might share some state with other luthiers, other state with other *guitar* luthiers, other state with other *acoustic* guitar luthiers, etc. Social network technology provides manual means (friends, groups, and networks) for creating, maintaining, elementary sharing among people with like interests (i.e., with like perspectives). In an attempt to choose terminology that generalizes such

The opinions and positions in the white papers and comments posted on this web site are those of the submitters only and do not necessarily represent those of the Federal government, the NITRD program and its participating agencies, or the National Coordination Office.

relationships, we characterize such shared perspectives as defining *groups*, *teams*, *communities* or *tribes* (all of these terms have been used by people building social network software). I imagine a NIT capability that maintains perspectives with programmable state for each user, that supports networks (or at least hierarchies) of perspectives, and that has a well-conceived state sharing model. Such a system requires a carefully designed and widely accepted structure (e.g., a standard) for representing state, a means of managing communities and their shared states, and a refinement in the way application tools (such as search) use the shared state.

This input focuses on an area of study in which users work in groups to accrue shared value – the implicit knowledge in the shared state – from which each team member derives benefit. This approach correlates with that of *collective intelligence*;[1] in this RFI "intelligence" is described as an evolved state models that incorporates potentially sophisticated data structures to represent structured (possibly hierarchical) and shared knowledge.

I imagine a new set of information processing capabilities in future NIT that facilitates the identification and preservation of group knowledge that represents the group's collective intelligence. I also expect that research in the area of *harvesting* collective intelligence will become an increasingly important component of the new NIT environment. The technology will help to: form groups (as in a social network); to structure and archive information of high interest to all/most group members; assist group members in using the knowledge implied by the collected information, and enable others to provide tailored services and goods to groups. There is already a clear commercial trend in this general direction (e.g., see FriendFeed.com), thus an NITRD effort is needed to coordinate the research frontiers of the work at the same time that it encourages technology transfer and commercial development of the capability.

An ominous aspect of this work relates to cybersecurity: such technologies are also likely to enable (antagonistic) third parties to observe the group information. Because of the implications regarding invasion of privacy, identity theft, etc., it seems clear that while harvesting information can enable highly refined information harvesting and group support, it is likely to be important to have technology, standards, policies, and legislative actions that address security related to the technology. For example, it is important to understand how to create NIT that enables harvesting of collective intelligence in the aggregate without using/storing information about individuals in a group. This aspect of collective intelligence is likely to affect many different NITRD agencies. The evolution/revolution of collective intelligence is deeply intertwined with cybersecurity – what is the correct balance between the ability to exploit collective intelligence without violating social, customary, and legal barriers? I imagine that to the extent collective intelligence emerges as an important NIT, it will necessarily be associated with a carefully thought out cybersecurity strategy.

I believe that the idea of collective intelligence is fundamental in contemporary and evolving NIT; as a consequence, it seems important that part of the NITRD charter should encourage, monitor, influence, and support the evolution of the area as it evolves. Information retrieval for the masses is an important multi-disciplinary application for the NIT infrastructure. Work in this area spans networks, systems, security, software, web programming technology, AI, anthropology, ethnography, and more. Significant research and technology improvements are needed to exploit the notion of collective intelligence.

---

[1] See http://cci.mit.edu/ describing the multi disciplinary M.I.T. Center for Collective Intelligence, formed in 2006 to study the emerging concept of collective intelligence.

The opinions and positions in the white papers and comments posted on this web site are those of the submitters only and do not necessarily represent those of the Federal government, the NITRD program and its participating agencies, or the National Coordination Office.