

# From Community Discussion to Knowledge Distillation: The I-DIAG Research Project

Mark S. Ackerman<sup>1,2</sup>, Stephen Cotterill<sup>2</sup>, and Kevin Nam<sup>1</sup>

<sup>1</sup>Department of Electrical Engineering and Computer Science

<sup>2</sup>School of Information

University of Michigan, USA

{*ackerm, scotteri, ksnam*}@umich.edu

Imagine the following scenario: The president of a large public university in the US asks a blue-ribbon panel of his highly regarded faculty to reflect upon the future of their university. The president wants to keep their university not only in the forefront of similar universities but also in front of basic societal pressures and opportunities. However, the faculty are also admonished to consider the various often-overlooked stakeholders – the university’s staff, undergraduate students, graduate students, alumni, non-tenured instructors, state legislature members, and local community residents. A large US state university may have several thousand faculty members, and the various concerned stakeholders might include 50 thousand or more people. Of course, the faculty committee could do as a typical blue-ribbon panel often does, going into their respective rooms to inscribe their already acquired expertise. But if they wished, how might they reach out to these stakeholders, include their perhaps divergent opinions, and search for new and interesting opinions and options?

We know that Internet-scale systems can provide forums for large groups ( $> 10^5$  people) to gather, discuss, and trade ideas. Within a corporate setting, these systems can be used for brainstorming, new produce ideas, quality circles, and the like. Governments, institutions, and universities can discuss such issues as organizational change and future plans in order to come to a “shared mind”.

Yet all too often problems arise in these attempts. People do not come to the site, or do not stay on topic. More importantly, once use has finished (either by deadline or by neglect), the site is often a bramble of ideas and topics, too large and unwieldy for its information to be successfully reused.

Our system, I-DIAG<sup>1</sup>, investigates how to garner and then distill this valuable community knowledge. It is part of a larger project to investigate how to maintain and reuse informal information within organizational and Internet-scale settings.

---

<sup>1</sup> The main quad of the University of Michigan campus is called the Diag. I-DIAG is also short for Interactive Diagenesis. Diagenesis is “the recombination or rearrangement of constituents (as of a chemical or mineral) resulting in a new product, or the conversion (as by compaction) of sediment into rock (Webster's 1986)”.

## **I-DIAG Goals**

We created the I-DIAG environment to consider several general research problems as well as provide a concrete application with which to examine these problems. Overall, we are investigating:

- New models for refinement and distillation. Our primary interest is in finding social and technical mechanisms to facilitate the distillation of knowledge from large amounts of informal information, such as bulletin-board messages, chat messages, e-mail, or quickly written brief documents. Our argument below is that previous mechanisms have failed because of the social barriers. Accordingly, our emphasis is less on the technical mechanisms for doing textual summarization or knowledge elicitation than on finding social models with augmentative technical mechanisms to foster the creation of material and then “boiling down” of that material into something that will be subsequently useful to others.
- These “boiled down” repositories are the distilled and refined versions of many people’s thoughts about a subject, mostly likely specific to a particular socio-technical environment. We are also investigating mechanisms to foster the sustainability of this distilled repository over time.
- I-DIAG, accordingly, has three sets of users. The first user group consists of the people entering their comments and discussing appropriate topics. In general, these people will be from a specific organization, institution, geopolitical community, or scientific community. The second user group consists of the moderators, editors, and wizards who control the interactive discussion. The final set consists of the people distilling the archived materials, either for an external report or to create a more concise site.
- In any social space, mechanisms must exist to foster social regulation and sustainability over time (as in Ackerman and Palen [1996]). While social regulation can have pejorative connotations for computer people, some amount is necessary to continue any collectivity’s activities. It seems as though there are always problem or abusive users in online spaces. We also wish to prevent or ameliorate unproductive or hateful exchanges. As we will see, the duration for I-DIAG is very short – our vision is a few weeks of intense discussion. Nonetheless, there are still social regulation and maintenance issues to be resolved; indeed, some may be exacerbated by use assumed to be brief. Through I-DIAG we are investigating collaboration-centric, agent-based mechanisms to quickly move users into an understanding of the system and its uses, enable productive exchanges, and control potentially unruly users and problematic exchanges.
- Since we hope that use is rapid and the corpus of information is constructed very quickly, we are investigating interface mechanisms to allow users to return to the space and understand what is new quickly and effectively.
- Overall, we see ourselves as investigating new forms of knowledge production. I-DIAG forms an interactive or dynamic “book”, where the corpus is constructed iteratively and collaboratively by people with different opinions, types of expertise, and varieties of experience and viewpoints. This “book” is a living document – not only is it constructed

by people in terms of their own interests and knowledge, but it can be maintained over time in the same manner.

Our major goal, then, is to understand how to iteratively construct a refined knowledge repository (probably less than completely formalized but more distilled than raw messages). To do so, we must necessarily also investigate what technical *and* social mechanisms we need for sustainability, social regulation and maintenance, navigation and return, and interface metaphors.

Space limitations prevent from describing the components of the I-DIAG system – I-DIAG/CyberForum for the discussions themselves and I-DIAG/Consolidate for distillation, as well as a Social Agents environment for examining social regulation. See [Ackerman et al. 2003] for a technical description.

## **References**

Ackerman, Mark S., and Leysia Palen. 1996. The Zephyr Help Instance: Promoting Ongoing Activity in a CSCW System. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI'96)* : 268-275.

Ackerman, Mark S., Anne Swenson, Stephen Cotterill, and Kurtis DeMaagd. 2003. I-DIAG: From Community Discussion to Knowledge Distillation. *Proceedings of the International Conference on Communities and Technologies*