



---

**TITLE:** Concept feasibility analysis

**AUTHORS:** Gunnar Forsén, Thomas Lundin, Jonas Löwgren

**DATE:** October 13, 2008

**REVISION:** 1

---

In the following pages, we analyze the feasibility of the Pinpoint concept in terms of **technical** feasibility, **organizational** feasibility and finally **integrity and privacy** issues. We conclude that the concept is feasible for organizations where overall goals are furthered by openness and exchange across internal boundaries.

## TECHNICAL FEASIBILITY?

The key question is, of course, whether Pinpoint's envisioned data-mining engine would be able to provide subject distance and communicative distance matrices that would characterize the people in a large organization appropriately.

One aspect of the question is whether semantic analysis and clustering is technically possible to the level needed for **computing sensible subject distances** and **choosing sensible system-generated tags**. Technologies for analyzing and clustering concepts in text documents have been researched and published for more than twenty years, including landmark algorithms such as latent semantic indexing, multidimensional scaling and self-organizing maps.

In recent years, consumer applications are starting to emerge in the wake of the Semantic Web hype, oriented towards automating the creation of metadata for web contents. Most of the interest has been focused around news, where some of the more successful examples are Daylife [www.daylife.com] and Opencalais [www.opencalais.com]. These examples typically draw on one or more of the generic semantic-analysis algorithms mentioned above, complemented with heuristic methods for identifying genre-specific concepts and labels. In the case of news, such genre-specific concepts include actors (people mentioned in the news reports) and formal relations between actors. A similar approach of combining generic analysis with genre-specific heuristics is straightforward to adapt to the domain of organizational skills and competences.

**Computing communicative distance** is a comparatively simple task of combining communication frequencies based on email, IM and possibly internal phone logs with communicative preferences as expressed in address books and buddy lists.

A final aspect of the question has to do with the quality of the raw data. When a person starts in an organization, there are only a few data isolated data points for the data-mining engine to analyze. Consequentially, Pinpoint will draw mainly on organizational role and user-provided tags to compute subject distance; communicative distance will be pointless

at this initial stage. As the new employee works on projects, writes documents, uses the intranet, takes part in documented meetings, and communicates with colleagues, the data set grows and enables the data-mining engine to relate her in more meaningful ways to others in the organization.

Our assessment is that **the technology for semantic analysis and clustering is sufficiently mature** to build and deploy a data-mining engine that would deliver distance matrices with a perceived relevance of 80 percent if the raw data is rich enough. In other words, when a person has worked in an organization long enough to leave significant amounts of manifest data in the internal-public domain, she would judge four presumed close people out of five proposed by the data-mining engine to be truly relevant in terms of subject or communicative proximity. We find this to be strong enough to pronounce the concept technically feasible, since the goal of Pinpoint is to inspire through serendipity rather than giving exhaustive and »correct« answers to search queries.

## ORGANIZATIONAL FEASIBILITY?

What is required of an organization in order to deploy a system like Pinpoint? There appears to be at least two issues involved here; one of **organizational policy** and one of **information system federation**.

Pinpoint builds on the assumption that it is more feasible to leverage off work that is done for other reasons than to collect specific information in separate information-gathering activities. Another assumption is that it is beneficial for a group within an organization if the group members establish contacts with colleagues in other groups of the organization. Both of these assumptions can be at odds with existing practices and norms within an organization, and if so, an attempt to deploy a system like Pinpoint is likely to fail.

Moreover, if a person becomes exposed to the rest of the organization as a key competence in a certain area, a situation of competition for that person's resources might arise which can be harmful to the productivity of the group where the person belongs. Whether the greater good for the organization compensates for the local losses is a question of organizational policy.

The question of information-system federation simply concerns the fact that Pinpoint is intended to mine several separate systems (file servers, databases, intranet logs) for data. In an organization where such mining is considered harmful for policy or security reasons, deployment of a system like Pinpoint is likely to fail.

## INTEGRITY AND PRIVACY?

The most obvious aspect of integrity and privacy concerns the way that Pinpoint potentially exposes an individual to the entire organization. The concept has been designed with this consideration in mind: The only information that the visualization gets from the data mining engine, and hence the only information that a Pinpoint user can access, is two numeric matrices of subject distance and communicative distance. Still, it should be noted that the idea of mining internal-public data behind the scenes has undesirable traits of surveillance and Big Brother scenarios.

Similar concerns should be raised against the mining of communication patterns and preferences. Even though data mining and surveillance is already very much a reality, both on the open web (e.g., Google's personalized search results and advertising) and within organizations (e.g., automated web-browser log monitoring), the Pinpoint concept makes it much more visible and tangible. We suggest that while Pinpoint might be acceptable in certain corporate and national cultures, it might be an impossible proposition in others.

There are also concerns about managerial misuse – in other words, critics have pointed to the danger that Pinpoint may be used by managers to analyze the work habits of their employees and use the insights in ways that are potentially harmful to the employees. This is certainly a valid concern and it cannot be easily dismissed by referring to the fact that all the sensitive information is already available to the manager. After all, the strength of Pinpoint in pulling the information together and presenting it in a manageable way is also what makes it vulnerable to this kind of scenarios. The risk for managerial misuse will be contingent on the managerial policies prevalent in the organization where a system like Pinpoint is deployed.

A final issue to do with integrity and privacy is the situation where a person's attractive skills and knowledge are disseminated throughout the organization to the degree that new Pinpoint-mediated contacts start occupying too much of the person's work time. This indicates the need for individual availability-status settings as part of the contact information visible in Pinpoint, and possibly the organizational implementation of Single Point Of Contact entities for different competence areas that emerge as popular targets for Pinpoint-mediated contacts.

## CONCLUSIONS

- To summarize, Pinpoint appears to be a feasible concept for organizations where the overall goals are furthered by openness and exchange across internal boundaries, and where integrity and privacy are founded on mutual respect.
- Moreover, it is technically feasible even though a full-scale implementation would be moderately demanding.
- Implementing the data-mining engine is a comparatively small, albeit somewhat advanced technical task (on the level of computer science M.Sc. or equivalent).
- The largest costs in implementation and deployment would arise from relatively unsophisticated work on federation of legacy systems, and from organizational initiatives required to roll out the system and fine-tune its genre-specific data-mining heuristics.