[Social Scanning: Improving Detection of Future Trends and Weak Signals Through Web 2.0]

Alex Soojung-Kim Pang, Ph.D.

Future2
askpang@future2.org
http://www.future2.org

making futures more perceptive and persuasive

Summary

I propose to develop an online system to help technology forecasters, strategists, and futurists detect emerging trends, interesting anomalies or "weak signals" of disruptions. It will aggregate content generated by the futures community; filter the content to detect common and uncommon terms; then use longitudinal slices of this data to identify outstanding trends and outlier signals.

Mapping future-related trends and spotting early signs of disruption is much more a craft than a science. Futurists scan the news, scientific literature, and other sources for new trends and discontinuities. If journalism is the first draft of history, scanning is the first draft of the future. Many futurists now share what they're reading on blogs, Delicious, or other online services, but still scanning remains ubiquitous but solitary: futurists have not been able to pool their scanning efforts, generate professional capital from scanning, or develop new insights by working collectively.

I propose to do all three using an approach I call "social scanning." It draws on content futurists and other subject experts are already creating. It takes content that is private but not proprietary and make it shareable and valuable. It creates goods that are useful to the community as a whole and to individual practitioners. Finally, it serves as a foundation for collaborative research in a community of practice that is highly distributed.

Social scanning need not be expensive. The content is already being created by futurists and technology-watchers. A basic working system that could identify trends and disruptions from that content can be created with a few months' development time. Previous online scanning systems I helped develop at the Institute for the Future cost about a million dollars a year; my group estimates we can develop a system, operate it, and improve it for less than half that.

1. Introduction

Scanning is one of the fundamentals of futures practice. Futurists regularly scan the news, magazines, scientific literature, and other materials for new emerging trends, to detect the appearance of potential futures, or to find interesting anomalies or weak signals that herald disruptions or paradigm shifts. Scanning is common but it is not competitive: we don't do it for money,

but to enable us to make money. However, we tend to scan alone: even though we all draw on some of the same sources, scanning is ubiquitous but solitary. Put another way, scanning is immensely inefficient.

For many years, this inefficiency was a necessary evil. Most of our sources were printed, and it was impossible to share our reading in an easy and timely way. Scanning is supposed to be rapid and a bit impressionistic: it's more like note-taking that scholarly writing, and should both generate information and stimulate creative thinking about the future. Scanning, like any exploratory or speculative activity, has a high failure rate: most of what we look at turns out to be of little value.

This can now change. The rise of Web 2.0 systems that make it easy to share links to interesting articles, bibliographic citations, or news stories; the rapid advance of technologies for aggregating and analyzing large bodies of content; the growth of tools and cultural practices that reward information sharing; and experience with a first generation online tools for supporting science and technology forecasting, all make it possible to develop a new approach to scanning. I call this approach "social scanning," for a couple reasons. It would work by drawing on content futurists and other subject experts are already creating. It would take content that is private but not proprietary and make it shareable and valuable. It would create goods that are useful to the community as a whole and to individual practitioners. Finally, it would serve as a foundation for collaborative research in a community of practice that is highly distributed.

This essay describes how social scanning would work: how a basic social scanning platform would work, what functionalities it would need, and what value it would generate for professional futurists and our clients. It begins with a discussion of my own experience developing online tools at the Institute for the Future to scan for weak signals of disruptive changes, analyze emerging trends in science and technology, and publish forecasts. It then surveys the growing use of Web 2.0 platforms among futurists, and how they are generating content that could be aggregated and analyzed for our collective benefit. It then outlines the basic functionalities of a system to do that. Finally, it describes the benefits such a system would bring to individual futurists, and to the discipline as a whole.

2. Futures Platforms at IFTF

From 2004 to 2009 I worked on a number of content management systems designed to support research at the Institute for the Future. The largest were two public systems: Delta Scan, a project for the British government's Horizon

Scanning Centre, which collected over a hundred long forecasts on the future of science and technology to 2050; and Signtific, a National Academy of Sciences-funded project that collected several thousand signals on trends in global science, technology and innovation between 2007 and 2009. Both projects followed a similar workflow. Experts were contracted to contribute short pieces on current trends or on-the-horizon developments, and longer forecasts that discussed the implications of those trends. In-house researchers then used the content to develop topic maps, and worked with clients or other outside organizations to apply the content to their strategic planning or policy processes. Both systems proved useful, but we also learned some important lessons that could be incorporated into social scanning.

Getting experts to participate for long periods on futures-related projects is hard. In both Delta Scan and Signtific we recruited graduate students and postdocs as contributors, thinking that they would be easier to hire, have a good sense of their fields, and have a strong incentive to think about the future of their disciplines. But personal career interest didn't translate easily into the kind of broad perspective futurists have, nor did it guarantee high participation in the system: thinking about your next professional move isn't the same thing as thinking like a futurist about your discipline as a whole. (It may also be the case that if you're the only one who sees the Next Big Thing, the potential career rewards to keeping that knowledge secret were greater than any incentives we could offer to make it public.) Even throwing more money at the problem wasn't enough to engender investment in and commitment to the project. On Signtific, we had a corps of experts who received a substantial monthly honorarium, who were expected to write a certain number of short pieces and longer forecasts per month. But it proved difficult for busy people with research to conduct, grant applications to write, conferences to attend, and lives to lead to spend a few hours a month writing for Signtific. The problem was not that it was too large a commitment: it was that it was too easy to defer.

It did help to make the contributions less formalized or formulaic, particularly once it became clear that most contributors don't like thinking about or creating metadata. In Delta Scan experts were required to estimate the likelihood, impact, time frame, and geographical scope of each forecast. A number, however, challenged the possibility of forecasting these dimensions. For scientists accustomed to looking for the right answer, talking about long-term trends seemed too much like pure speculation. In a public venue there was no upside to being right, while it would be easy to expose yourself to ridicule. In response, in Signtific we made two changes. First, we reduced the number of factors to two: likelihood and impact. Second, we made it possible for anyone to vote on these factors, much in the same way people can vote on articles on Slashdot. Had it gone well, this system would have let us map

signals or trends that were low-likelihood but high-impact (and thus wild cards), and compare how users in different fields or parts of the world viewed the same trends. Even with the simpler format, however, it proved difficult to get readers to rate content.

Some of the same challenges hindered broader community-building. We gave users the ability to contribute their own content or rate the importance and likelihood of existing forecasts, and assumed they would participate out of intellectual interest or for public recognition; neither was a powerful draw. Experts hired as freelancers or contractors, in contrast, had a clear understanding of both the scope and the limits of their obligations. It's hard to contract out community participation.

On the other hand, we did find other things that worked well, focused expert contributors' attention and labor, and reduced the amount of work necessary to edit and maintain the database. Most notably, we found that workshops, properly structured and supported with the right electronic tools, could yield a tremendous amount of useful content. (As one participant put it, it was easier to get more done in four focused hours than four distracted months.) Some were writing workshops, in which people wrote signals; in other cases I brought together experts to analyze the current state of the database, and develop scenarios or forecasts based on existing signals. Many of these were one-day events, but eventually I was able to design a half-day workshop format that was still quite productive. The key to making them work is to bring together people physically, and provide the group with a good technical framework and process for capturing their insights. Structuring the work this way allowed them to focus their attention, compare their work with others, and get a better sense that they had made a tangible contribution to the project.

But despite our best efforts, we never quite managed to encourage the development of an online community that would create and rate content, update and enrich the database, and help us identify trends or disruptions we never would have found ourselves. But while we struggled with this challenge, futurists discovered Web 2.0. And an unexpected solution to our problem-- and a whole host of new opportunities-- presented themselves.

3. Futurists Discover Web 2.0

Today, futurists using Twitter, Delicious, Digg and other Web 2.0 services publish a flow of content that is probably already too large for any person to follow, and is growing rapidly. For example, Twitter publishes roughly 600-700 tweets per day marked with the #future hash tag. The futurists I follow post 70-80 tweets per day (though some of those posts are personal or auto-

generated by other agents). Futures-oriented lists on Twitter follow anywhere from a dozen to three hundred people, and almost those lists are all available via RSS.

Other systems generate equally substantial bodies of content. Users on Delicious, the oldest social bookmarking service, post about 350 bookmarks per day with the tag "future." My network (which includes a select few futurists) posts about 220 bookmarks per year. That translates into about 1120 separate data-points per day, or over 400,000 signals per year -- just from three services and a relatively small community. Futurists' blogs feature between 100 and 200 posts per week. Casting one's net wider, one can rapidly capture an enormous number of potential signals. Consider Tweet the Future (http://tweetthefuture.com.com/), a Web site that monitors Twitter for tweets containing the word "future. " It finds about 30 tweets every minute--over 40,000 a day-- though the vast majority of these tweets have nothing to do with futures or forecasting.

So many if not most futurists, consulting companies, and futures-oriented nonprofits are using one or more these systems. Most of these datastreams are real time-reflections of what people are reading. These datastreams represent a vast but untapped resource that could be used to build a picture of the collective attention of the futures community, and detect weak signals: indeed, it can largely replace the kind of commissioned content that fed Delta Scan and Signtific. We no longer have to work alone to find interesting things. Instead, we can detect patterns in our and our colleagues' datastreams.

4. How Social Scanning Would Work

How would a social scanning platform work? Its core functionality would be an engine that gathers signals from the free and nearly real-time content produced by futurists and subject-matter experts on blogs, Twitter, and other social media platforms; analyzes this content to find subjects and citations that are of greatest interest to the futures community; and clusters together material that shares unusual terms, keywords, or links to common references. This would let us identify both popular subjects and outlying wild cards, and create a body of data that could support others tools or services.

The system would harvest RSS feeds generated by a list of blogs, Twitter, del.icio.us, Digg and other services generated by the system's managers. The list would have some simple metadata about sources, most notably their authors; it would also record metadata from its sources, particularly the publication date and time of posts and articles, and whatever tags attach to the content.

What would the system it do with this datastream? The first key task would to filter it. By gathering information about the author of each feed, it would be able to associate multiple feeds with the same author. If the same author has several different sources that the system is following, the system would look across those and filters out repeats. For example, if I have a blog and del.icio.us account, and both automatically push updates to a Twitter account, the system knows to look for cross-posts between those services, and count a blog post that generates a Tweet only once.

The second key piece of filtering involves associating multiple hits on the same subject. Different people may talk about the same event but reference articles published in different places, or the same article published in multiple places--a wire service article that appears in several newspapers, or an article that is reblogged. The system would also need to be able to identify different URLs as pointing to the same article—e.g., the full URL or an article and a bit.ly shortened URL. Identifying these sources could be done by software, by users, or both. So while repetition by an individual would be controlled for, multiple citations and references are recorded. The former is noise in the system, but the latter is signal: the more people who tag or blog about a subject, the more important it is. (Citation and referencing also filters out non-professional noise. Many Twitter users combine references to major new articles with announcements like "I am eating a sandwich;" the latter are far less likely to be referenced by others than the former.)

In Delta Scan and Signtific, contributors or community members were supposed to formally rate the importance of different trends. In this system, we can simply assume that if someone takes the time to share a link to an article, they consider that article to be worth their attention. More links, especially links over time, indicate the emergence of a group consensus that a link points to a trend worth watching.

This kind of filtering could be done automatically, and improved by users. People may be able to identify associations between articles that automated systems don't. They could group together content from the data stream by adding tags to specific pieces of content; and they can either tags or identify synonymous terms (e.g., ubiquitous computing, ubicomp, and ubic, and ubiq all mean the same thing, for example). My experience with Delta Scan and Signtific suggests, however, that this system needs to be kept as simple as possible. People generally don't classify things unless there are clear incentives and immediate rewards. Even then there are huge variations in the use of hash tags, keywords, etc. between users and across systems, and little chance that people can be induced to adopt standard taxonomies or ontologies. However, when you're working with high social knowledge, and information that by

nature exists at the boundaries of the human corpus, it's important to maintain a high degree of ontological flexibility.

Rewarding people for doing this kind of tagging and associating would send the important signal that community-oriented work deserves to be recognized and encouraged. This kind of work has traditionally been essential for high-quality scholarly and professional activity (think of the legal profession's vast corpus of precedents and codes, the medical profession's reference works, the scientific world's gigantic structures for sharing everything from raw data to polished research) but has either been done largely by professionals-- librarians, catalogers, and others-- with little professional visibility, or by organizations that extract high rents for their work. By rewarding users for improving the system and contributing to the professional good, we can harvest some of the benefits of that organizational work without incurring its costs.

5. Intellectual Benefits

Heat Maps of the Future

This content could be presented in a variety of ways, at several time scales. A list of most popular subjects or citations from the last 24 hours, akin to the defaults lists on Technorati or Digg, would have the virtue of simplicity and familiarity. Citations and references in today's datastream can tell you what futurists think is interesting right now; but looking at the datastream over longer time periods-- weeks or months, say-- would give users a clearer picture of what issues are of enduring interest. New product announcements, elections, or disasters generate a frenzy of postings and repostings that die off quickly. In contrast, articles that are still cited after weeks or months are likely to deal with issues of more enduring importance. Looking at a longer stretch of the datastream will also help identify people who are good at spotting important trends early, and who can do so consistently. It will note who first identified the event, who subsequently picked it up, and what chains of influence connect people together.

Weak Signals

These heat maps would provide the background for what many people are really interested in: weak signals of disruptive change. Embedding the search for weak signals in social scanning would improve it greatly, by providing a standard against which the uniqueness of any signal can be measured. Today, the search for weak signals is pretty intuitive, and what counts as a weak signal is personal and subjective: my weak signal may be someone else's conventional

wisdom, and vice versa. Aggregating signals from across the futurists' community would help individuals tune their intuition by letting them see when their weak signals are genuinely novel, and are actually well-known to people in other countries or experts in other specialties; and it would help the discipline as a whole by nudging the search for weak signals into something more rigorous and systematic.

Building Additional Functionalities

Identifying heat maps, trending topics, and weak signals would be basic functions of a social scanning system. Of course, it would be possible to develop additional functionalities based on this content. You could create tools for professional forecasters tools to benchmark and improve their practice, by showing users how their interests compare to those of the field as a whole; how often they identified weak signals that later were cited by others; and how important things they rated highly turned out to be over time.

Other tools could be used by groups. Top-rated topics could be flagged in a prediction markets system whose participants could more explicitly bet on the importance or timing of disruptions or future developments. Yet others could be used with clients. For example, interactive roadmaps based on content material from the system into an online presentation software system Prezi could be used in strategic planning workshops.

6. Professional Benefits

Social scanning would offer benefits both for individual futurists, and for the field as a whole. It would facilitate better scanning by converting private work into public goods. Social scanning would provide a social platform connecting the field together. The system would identify people who are good broad scanners, who are good at seeing trends early, who can spot weak signals, or who don't know each other but share research interests. Finally, social scanning could improve the profession of futures by giving practitioners incentives to share their work and systematically improve their forecasting.

Social scanning would be better scanning. It would generate a continuously-updated, community-wide and collective view of what trends are shaping the future, and what signals suggest the emergence of new trends. We can see what various futurists (somewhat independently) consider important, by comparing input from multiple sources. In other words, our collective reading patterns may reveal some insights that we could not create individually. At the organizational level, it would reduce the work of starting new scanning

platforms for projects; instead, researchers could draw on existing, automatically-updated scans, augmenting them with additional work when necessary. It would make scanning more efficient at an individual level, too. Today there's a lot of repetition in scanning, since futurists don't have a way to systematically share the work of scanning. If we could pool the results of our work, and trust the whole community to keep up with the most popular (and, one hopes, most critical) trends, individuals would have more time to spend looking through specialized or offbeat sources-- a diversification which would enrich the discipline as a whole-- as well as working on synthetic, interpretive activities. To draw a parallel to the academic world, most scholars focus their own energies and writing on specialized subjects, and work with colleagues to evolve new approaches, schools of thought, etc. This latter work doesn't always happen formally: it emerges through literature reviews, thematic essays, conferences, and conversations-- a whole infrastructure for producing collective knowledge that futurists haven't really replicated.

Social scanning would allow practitioners to build professional reputations for more kinds of work and insight. Today the fastest way for a futurist to build professional capital is to make flamboyant public pronouncements; doing the more mundane work of identifying less flashy trends, or assembling evidence that others can use, receives virtually no credit. There are currently no mechanisms for recognizing researchers who are terrific scanners but lousy forecasters, or who have a brilliant eye for weak signals but no public presence. By awarding users points for each item them contribute to the datastream (i.e., writing posts on their blogs, adding bookmarks to their del.icio.us account, etc.) and additional points for work they do within the system (e.g., tagging content, associating different pieces of content, or rating contributions), it would quickly become possible to identify people who are community-minded and generous with their ideas. Some of these users may turn out to be well-known names in the field; others may not. (Because the system can also analyze the importance of contributions, it could distinguish people who's work is defined by quantity rather than quality.) But by making it public, the system would give scanning and sharing the recognition they deserve.

This in turn will enrich the professional ecology, by making it possible to practitioners to build social capital from a wider variety of intellectual and professionally constructive activities. This would make futures more like better-developed and -organized disciplines like physics, where people can specialize in particular subjects (high-energy physics, cosmology, condensed matter, etc.), but also make careers as theorists, experimentalists, instrument designers, or computational experts. This is not to say that some of these specialties aren't higher-profile than others, but what matters is that the field has mechanisms for recognizing and rewarding all kinds of contributions to

science. This is missing in futures, but there is an opportunity here, thanks to the fact that very few futurists make any money from scanning, but instead make money from the things that scanning enables. Turning this largely invisible private activity into a public good would raise the overall quality of scanning, and recognize and reward good scanners for their contributions to the field.

Social scanning could bring gentle coordination to the discipline. The field lacks the centralized, gatekeeping institutions-- a few dominant graduate training programs, a strong professional society, government certification-- the give shape to other professions like law and medicine. Nor does it have the canonical literature, moral codes, and daily practices that define members of religious orders. Futurists are spread in corporations, government agencies, consulting companies, one- or two-person groups, and academia, and most of us spend much more time talking to clients than to each other. As a result, the field is physically dispersed and intellectually decentered. Social scanning would help build a more cohesive sense of identity by making the community's interests visible to itself; allow far-flung practitioners who share common interests to find each other, and let them build on each other's work in ways we cannot now.

Social scanning would also raise the quality of the discipline. It would provide clear benchmarks for practitioners: it would let me compare what I've been reading to my colleagues. Social scanning would also contribute to the development of more solid and rational professional standards. Today, the market rewards the most public futurists for being provocative more than for being useful or right. The upside to analytical rigor and correctness is low, and the downside to being wrong is even lower. Social scanning would begin to shift the economics of professional reputation, and provide a system that ignored flamboyance, gave less credit to single dead-on predictions, and rewarded less spectacular but more consistent performance.

Finally, a social scanning platform would do all this without requiring something as elaborate as a World Brain (appealing though that idea might be), or requiring all futurists to adopt common software packages. Like all good knowledge tools, it lets people do what they're best at, and computers do what they're best at. It can be easily adapted by users and integrated into their existing workflows and habits. We can harvest work that people are already sharing. Nobody who already has a blog or thousands of del.icio.us bookmarks has to switch systems, learn a new tool, or abandon legacy content. They just keep doing what works for them.