Second International Workshop on Knowledge Federation

Noah's Ark for the Exaflood Rushing Upon Us

Version 2 of 3

Mei Lin Fung, Institute for Service Organization Excellence, Palo Alto, California USA

Robert Stephenson, Director of The Tech Virtual, The Tech Museum, San José, California USA

From the beginning of humanity to 1999, a UC Berkeley study estimated that we created 12 exabytes of information. An exabyte is 50,000 years of DVD quality video. As of March 2010 Wikipedia's entry for Exabyte heralds the Exaflood reaching 21 exabytes per month over the Internet. And rising. Entire categories of jobs are changing as print media vehicles jettison their journalists, schools re-think teachers, and everyone tries to keep afloat as the tidal waves of information break on every human endeavor. The Singapore National Library is organizing networked communities inspired by Douglas Engelbart. Social networks are being explored by the US Department of Defense. The US is brimming with collaborative initiatives like Huffington Post (journalism), DreamFish (jobs), TEDxSalonBayArea (conversation), QuantifiedSelf (health) and new grass roots movements are now emerging that span or cross industries like Government, Climate Change, Energy and more. Broadcast TV is being overtaken by YouTube videos. What role does knowledge federation play in nurturing and supporting these emerging human connections?

Upon this gifted age, in its dark hour

Falls from the sky a meteoric shower

Of facts.... they lie unquestioned, uncombined

Wisdom enough to leech us of our ill

Is daily spun; but there exists no loom

To weave it into fabric

- Edna St. Vincent Millay

Upon our gifted age in its new awakening

Falls from the sky a meteoric shower of stories

Of acts of beauty;

Compassion that lies unremarked and unrecognized

Wisdom to nurture life, heal hurts, grow joy, is daily spun

Let us commit our loom of civilized intention

to weave wisdom into the fabric of our souls

- Mei Lin Fung

There is no loom except our own minds: technology can fetch and correlate but it cannot understand or make sense of this shower of facts -- at least not now or anytime soon. The issue becomes one of finding and consulting the few who have the best understanding of that which we seek to "weave." This is actually two problems: 1) what do we mean by "best" and 2) how do we find those who are the best in that domain. Answering the former and succeeding at the latter requires a reputation system. Humans have been dealing with other based on reputation for as long as we have been communicating, and we have developed very subtle systems of reputation and trust that work best in a face-to-face environment, or at least in a restricted group where many people know each other.

The system we propose, the Loom, is intended rather as a distributed system that could work over the Internet among people who are mostly strangers to one another. It is primarily a human system, augmented with a bit of technology. It is, in other words, a "Mechanical Turk."

There are many examples of distributed, computerized reputation systems---for example credit reporting agencies, Amazon's book recommendation system or SlashDot's comment rating system---but they are generally one-dimensional. They treat reputation as a single quantity, whereas it is, in fact, a vector of high dimensionality. For example, I cannot think of anyone I would recommend more for advice on the Java programming language than James Gosling, but if the issue were selecting a suit I would probably prefer to ask George Clooney. We start from the premise that there are as many dimensions of reputation as there are domains of human knowledge where someone might reasonably have a question that needs answering.

The ingredients of the Loom reputation system are six:

1) people with important (to them) questions to ask,

2) a system of communication,

3) a network of people who know stuff,

4) people who know other people who might know stuff,

5) a referral and record-keeping system, and

6) a form of reward or currency.

People with Questions

These questions may require specialized knowledge, particular wisdom or just common sense. Their common denominator is that someone cares especially to get an answer, and that answer should not exceed one page of text. The questions, too, should be expressed in one page or less. Of course, many important questions cannot be posed or answered in the compass of a single page. For these, the Loom reputation system can serve as a referral system: "can someone recommend an architect to remodel a modest private home in Berkeley, California?"

Communication System

By default, all communications will be anonymous. Although this may sound surprising for a system of reputation, it is in general a necessary feature. Many knowledgable people are also busy, and may be reluctant to give out their personal contact information for fear of receiving unwanted contacts. Finally, this helps ensure that answers are judged on their merits, rather than the answerer's authority. To reduce latency, most communications will be via SMS, IM, email or the like, exchanged through a central server to preserve anonymity.

Network of Referrers and Answerers

Each of these people has a profile in the Loom reputation system. The initial group used to bootstrap this system will be the group of KF10 attendees.

Referral and Record-Keeping System

This is the common server that forms the hub of the system, assures anonymity and tracks all interchanges. All messages exchanged, though anonymous, are posted publicly. It is important that it support many modes of communication, as noted above, so that each member of the network can interact with it in the manner she prefers.

Currency

The currency is used to reward successful answerers and referrers. Initially it may be a virtual currency, but for true scalability it will need to be interchangeable with conventional currencies. To provide anonymity, a currency like Bitcoin (http://bitcoin.org) might be appropriate.

How the Loom Weaves

Consider the following scenario:

THE QUESTION

1. George is curious about sustainable living and enters the following question: "is there a design for a compostable toilet that makes sense for my house in the suburbs?"

REFERRAL

2. The communication system checks its logs to see if any similar question had been answered before. None had, so George's question is routed to several people, including Earnest, who have good reputations as referrers on a broad range of issues.

3. Earnest refers the question to Sophia, since he knows she is interested in sustainable architecture.

4. Sophia forwards the question to John, because she knows he has just completed building a sustainable house.

NEGOTIATING

5. John replies that the question is incomplete: George needs to specify how much he is willing to pay for installation and maintenance. George adds these details to his posted question.

6. George and John negotiate a price for the answer. George offers 10 BTC (Bitcoins) for a complete answer and John accepts.

7. George receives several other offers. He chooses John's, based on Sophia's recommendation, and puts the 10 BTC fee in escrow.

ANSWERING

8. John forwards his answer to the question.

SETTLING

9. George reviews John's answer and, after consideration, decides that, although it did not completely fit his situation, John's answer satisfied 80% of his question. John agrees to 80% and is satisfied. Two BTC of the fee is returned to George and the remaining 8 BTC is removed from escrow.

10. Earnest and Sophia both receive 5% of that for their referrals. The Loom system fee is 1%. John receives 89%, or 7.12 BTC.

REMEMBERING

11. The system logs the following information:

   George: his question; that it had to be clarified, that he was 80% satisfied; how much he paid.

   Earnest and Sophia: the question; that John bid on it; that his answer was 80% satisfactory. That the answerer

   agreed that was fair.

   John: the question; the clarification; the price; the answer; that it was 80% satisfactory, that he was

   satisfied.

Should another questioner ask the same question, she would immediately receive John's bid at the original price.

Should she ask a different but related question, or balk at the fee, her question would be relayed directly to John for him to bid on.

In either case, if John's bid was accepted and his answer satisfactory, the original referrers (Earnest and Sophia) would each receive 4%; the Loom system 1% and John 91%. As John answered subsequent questions related to the first one, the referrers' share would decline to 3%, then 2%, then 1%, then 0%. The Loom system always gets 1%.

NOTES AND COMMENTS

----

feedback from PFTF Meetup

- rate & refine the question

- payment may be 0-100% of fee, depending on quality of answer

----

procedure for CtC meeting:

exchange: some food (cookies, jelly beans, etc, that can be cut (qty:~40)

rules: raise your hand to talk - don't speak until called on by moderator

roles: loggers (1-3), moderator (communication hub) (1), the others (questioner, referrer, answerer) will evolve spontaneously.

topic: the CtC world of discourse - local businesses, economy, environment, health, facilitating green building

some feedback from the meeting participants:

Bill: clarify that this is for a *distributed* reputation system; in a f2f situation there are better ways to do this.

Howard: your hypothesis, that reputation is the ark, not explicitly stated.

Maurey: teams might want to answer a Q, providing better answers than an individual and sharing the reputation.

my remark: best to bootstrap the system with a limited domain of QQ, like e.g. knowledge representation/storage/retrieval/federation.

Mark: system is flawed ... (I agree in general, but didn't understand his reason)

Jack Park comments:

Rob,

I understand where you are going much better now. This is cool stuff!

I have a few comments, none of which is more than half-baked.

First, let me point [1] to the sorts of gaming that can go on in some

kinds of reputation systems.

I spent some time looking at [2] for papers on a product that was

created for CALO that operates very much like what you were

demonstrating at PFTF last night. It's a system that takes a given

question and tries to figure out two whom to send it. It tracks

keywords, recommendations, etc, with a bayes network. Could not find
such paper. During the CALO trials, we ran it, tossing in questions,
answering questions, and recommending others to answer them. There is
a system that does just that which was a startup that just recently
got bought by Google --aardvark--(after which I stopped answering
questions). Can't remember its name, but I think that it was, in fact,
a spinoff (like Siri) from CALO--looking closely at aardvark, I think
not. I can find no papers on the CALO artifact (yet), but they do
exist. "bayesian expertise finder" is a good google query but I still
don't see the papers.

I see a tie that binds the ideas evolving in this paper to knowledge
gardening in a big way, especially since, IMHO, gardening is the
social component of any federation. I think that Dino's Value Matrix
might play a role somewhere even in the structure of those ideas, even
though I envision a value matrix assigned to every topic in the topic
map (to make reference to my particular implementation of a federation
platform).

A recent book on these topics is the O'Reilly book _Building Web
Reputation Systems_.

Miscellaneous links below...

Jack

[1] http://blogs.alternet.org/oleoleolson/2010/08/05/massive-censorship-of-digg-uncovered/

[2] http://caloproject.sri.com/publications/

http://sprouter.com/blog/sprouter-debuts-answers-site/

http://www.google.com/Top/Reference/Ask_an_Expert/

http://www.fiercecontentmanagement.com/story/google-aardvark-purchase-has-expert-finder-potential/2010-02-17

The name I was trying to resurrect is iLink, the name is jeffrey davitz

Do a google on him and, among other things, you land on:

http://videolectures.net/jeffrey_davitz/

http://www.faqs.org/patents/app/20080275849

the names on that patent app are the prime players from SRI

http://portal.acm.org/citation.cfm?id=1281292 available at

http://www.ai.sri.com/people/davitz

http://techcrunch.com/2010/02/02/aardvark-research/

http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.95.2851

http://battellemedia.com/archives/2010/02/the_anatomy_of_a_large-scale_social_search_engine

SB Shum preso on social learning & sensemaking: http://people.kmi.open.ac.uk/sbs/2010/08/social-learning-sensemaking-ci/