Editor: Jimmy Lin, jimmylin@uwaterloo.ca

Searching From Mars

Jimmy Lin, Charles L.A. Clarke, and Gaurav Baruah • *University of Waterloo*

How would you search from Mars? No, seriously.

The *Mars to Stay* concept describes a series of related proposals for establishing a permanent colony on Mars (see marstostay.com and www.mars-one.com).[1] Instead of landing astronauts there with the intention of bringing them home after a short visit, the plan is to send astronauts to become the first Martian colonists. Such missions would be far less expensive, since we wouldn't need to bring along fuel for the return voyage (and manufacturing fuel on Mars is risky). Far from being a cuckoo idea, this approach to exploration has the support of many — including Elon Musk, the founder of SpaceX (and co-founder of PayPal and Tesla Motors)[2] and Buzz Aldrin, the second human to set foot on the moon.[3]

Scientists and engineers have worked out many of the details, and the quite surprising conclusion is that this plan requires no new technological breakthroughs (that is, it's doable with present chemical rockets) and is economically feasible (to the extent that a single wealthy individual could bankroll the entire endeavor). Oh, and there's no lack of volunteers willing to go on this one-way trip.[1]

A survey of mission plans is beyond the scope of this column, as is a discussion of numerous challenges ranging from sustaining life (air, food, shelter, and so on) to maintaining social structures (dealing with conflict and long-term isolation) and even issues such as cost recovery (for example, there are proposals to turn the entire endeavor into a reality show). In short, there are lots of smart people thinking about these issues on which we have no expertise. However, we do know a thing or two about search and Big Data, and on that we will most happily speculate.

How would you search from Mars? And more generally, how would you use the Web from Mars? To make the scenario more concrete, let's look roughly 10 years into the future, and let's assume that "searching" and "the Web" will still feel at least somewhat familiar to a user from today. Of course, technology will have advanced dramatically, but the point is that we'll likely still be searching with something that looks like a Web search engine, engaging with friends on something that looks like a social network, and purchasing items through something that looks like an e-commerce site. Just for rhetorical convenience, we'll refer to brands that everyone is familiar with today, so when we say "Facebook," we really mean "Facebook, or whatever social networking service we'll all be using in 10 years."

Why Search from Mars?

The first question is, why? More precisely, from an information retrieval perspective, what's the task model? Mars missions, at least in the short term, will require substantial ground support on Earth, so our fearless Martian colonists will have access to the best minds from Earth to help with their problems. Plus, the missions will likely have been planned out in sufficient detail that responses to most survival-critical challenges will have been already mapped out. Thus, searching from Mars will likely not be a "Houston, we have a problem" need.

We anticipate that Martian colonists will be using the Web much in the same way we do today — reading the news, interacting with friends, watching highlights from yesterday's game, searching for information related to a leisurely pursuit, accessing adult entertainment, and so on. For convenience, we'll call this casual Web use. One initial reaction might be: What are the colonists doing wasting time on Facebook? Quite the contrary, these activities are critical to the psychological and emotional health of the colonists. They will continue to have strong ties to Earth, having left behind family and friends, and sustaining these connections will be important to overall well-being. It seems silly and a waste of resources to call up ground support to ask for score updates from a football game or to obtain a new vegetable stew recipe. Although intermediated interactions have been the norm in human

 Published by the IEEE Computer Society 1089-7801/16/$33.00 © 2016 IEEE

space missions throughout history, it's hard to imagine how such an approach is sustainable for a permanent colony. Indeed, we're already moving away from such rigid interactions: for example, personal Internet use is possible on the International Space Station today. Thus, we want to be able to search from Mars.

Another category of information needs will likely be scholarly search. An important goal of Mars missions is to advance science, so our colonists will require access to all of the scientific literature on Earth. For example, the colonists might want to publish about breakthroughs in hydroponics, and thus would need the Internet in exactly the same way that an Earth-bound scientist would: looking up related work, reading papers, interacting with peers, and so on. Although it might be possible to have an Earth-side co-author handle all these interactions, this would be awkward and frustrating for the colonists, not to mention contrary to the workflows of modern science. For these reasons, we want to be able to search from Mars.

Our goal is to make searching from Mars and Web use in general as close to the experience from Earth as possible. This contrasts with alternative approaches built on the idea of "slow search"[4] and asynchronous search models.[5] It will be a while before we have Amazon Prime on Mars, but it's perhaps reasonable to expect that a colonist could purchase (small) personal items from Amazon and have it delivered on the next supply rocket (estimated delivery time: eight months). Furthermore, the colonists will want to buy presents for friends and family. Although the latest holographic display might be too large for shipping to Mars, it still makes a great Father's Day gift. Such transactions, as well as purchasing the latest Kindle release or the digital plans for the latest gadget (for 3D printing) shouldn't be any more difficult on Mars than on Earth. One possible organization of Mars missions would be modeled after the military, where the colonists are permanently "on-duty" and paid a salary, so having disposable income is entirely plausible.

Big Data Bites

"Big Data Bites" is a regular department in *IEEE Internet Computing* that aims to deliver thought-provoking and potentially controversial ideas about all aspects of Big Data. Interested in contributing? Drop me a line!

—Jimmy Lin

The Constraints

How do we replicate on Mars the complete "Web experience" on Earth? Before sketching out the solution, let's first lay out the constraints and resources. In terms of the latter, the Mars colony likely wouldn't be self-sufficient for a while, so we anticipate substantial ground support and continued investment, including regular cargo supply rockets from Earth. What about the constraints?

Mars is sufficiently far that communication latencies are problematic. It takes radio signals between around 4 to 24 minutes to travel to Mars, depending on the relative positions of the planetary bodies, so we need to cope with a roundtrip latency of between 8 and 48 minutes.[6] That means a Skype call between Earth and Mars is out of the question, and we're not likely to figure out faster-than-light communication anytime soon (perhaps ever).

There exist technologies built around laser-based communication where it's possible to achieve good bandwidth between Earth and Mars. The Lunar Laser Communications Demonstration achieved a 622-Mbps downlink and a 20-Mbps uplink between the Earth and the Moon,[7] so something like this to Mars is technologically feasible. If we need more bandwidth, we simply build more satellites, and thus it's reasonable to count on substantial (but not infinite) bandwidth between Earth and Mars.

Getting physically from Earth to Mars is more complicated than aiming a rocket at Mars: both planets are in orbits around the sun (at different velocities) and we need to take into account the sun's gravitational influence. Transit is accomplished by transfer orbits — injecting a vehicle into an orbit around the sun that intersects the orbit of Mars at the right time. This approach requires two separate changes in velocities (delta-v's): first, from low-Earth orbit into the transfer orbit, and then from the transfer orbit once we arrive at Mars.

There are a variety of options with different tradeoffs: the Hohmann Transfer Orbit is the most efficient in terms of fuel but is based on a particular configuration of the planets such that a launch window only opens up once every two years.[8] Bi-elliptical Transfer Orbits take more time but require less fuel.[8] Conjunction class transfers are our best bet, with current or near-future rocket technology: they'll get us to Mars in between 120 to 270 days,[9] which is in line with the historical average: missions have taken between 150 to 300 days to reach Mars over the last half century.[10] As an aside, if we can overcome the political objections of using nuclear rockets (for example, as outlined in the Orion project), it might be possible to cut the travel time to around two months.[11]

Finally, it's also worth mentioning the concept of a Mars cycler, which is a vehicle in a special orbit around the sun that encounters Earth and Mars on a regular basis.[12] Instead of insertion into Martian orbit, a cycler keeps flying in an endless loop (around the sun) — payloads "hitch" a ride on the cycler and then "get off" at the right time. Cyclers are attractive in that they can rely on gravity-assist flybys to maintain or alter their trajectories, and thus require minimal fuel once the initial orbit is established. Of

course, the downside is that transit is limited to periodic windows.

It's the User Model, Stupid!

So, how do we actually do it?

Before sketching our solution, we note that our focus is on application-level challenges, as opposed to lower-level advances in computing technologies: for example, we simply assume that computing equipment will have been hardened to withstand the stresses of space flight and that technologies such as delay-tolerant networking[13] have already been deployed.

Human missions to Mars will likely be preceded by multiple robotic cargo missions that transport shelter, equipment, and supplies. A part of the cargo would simply be a copy of the Web — in other words, we start with the first interplanetary sneakernet. Obviously, we can't send everything, but as Andrew Tanenbaum once said, "Never underestimate the bandwidth of a station wagon full of tapes hurtling down the highway." Even with today's technology (16-Tbyte solid-state drives), a petabyte will weigh less than four kilograms. Of course, this is just one copy and we need redundancy (not to mention space-hardening) so there will be added bulk, but physically shipping to Mars a part of the Web is entirely feasible. By the time the mission actually lifts off, we would expect another order of magnitude improvement in storage technology: either 10 times more capacity or a tenth of the weight.

What should we include? A copy of all books and scientific articles that have ever been published would be a good start (that'll be relatively compact). We'd need some webpages too: For this, we'll ask Google to donate a crawl (plus index) of the portion of the Web that would be most valuable to the colonists. How would they know? Well, Google already has the colonists' interaction logs (who have been searching on Earth, no?). This data, coupled with data from millions more users and perhaps some manual curation, could be leveraged to build a long-term predictive model of user interests and information consumption — and from that, Google could extract a portion of the Web that any individual or group is likely to use. This could be set up as a relatively straightforward machine learning problem; something like (digital) bin packing, where the goal is to maximize the pages' overall expected utility.

Next: because the colonists will need entertainment, we do the same for Netflix. That is, based on the viewing history and preferences of each colonist, it's possible to assemble a playlist, subjected to a storage budget, that maximizes "viewing pleasure" (predicted "star ratings" would be a start, but hopefully the recommender systems community will have invented something better). Leaving aside intellectual property issues (with the entertainment content and everything else), gathering all these data is straightforward. We're not lawyers, of course, but even the intellectual property issues might not be that thorny if we simply think of this approach as the logical extension of edge caching — content distribution networks already do something along these lines today.

Okay, we got a rocket hurtling through space with a sizeable chunk of the Web on it. Here's where we encounter the first problem: the voyage to Mars is a long one (especially for cargo ships, which are likely to take longer, more fuel-efficient transfer orbits), and by the time our cache of the Web lands on Mars, it's already stale.

Not to worry, this is already a challenge that all Web search engine companies deal with today: the problem of crawl prioritization. The Web is, of course, dynamic and constantly changing, but some parts more frequently than others. Google must identify those parts and recrawl them, along with new content, subjected to constraints such as trying to minimize bandwidth consumption and not generating excessive load on remote Web servers. Fortunately, this problem is already being solved today.

So, we could (ask Google to) keep track of all the content in transit and capture any updates, continue to refine the predictive user models on Earth with new data, and beam over the "diffs" to Mars. We would send over some empty storage in a robotic cargo mission prior to the large sneakernet delivery, and this could hold the "patches" that are applied when "the Web" lands. The reasonable bandwidth we'd expect to Mars, coupled with this temporary storage (which the colonists might later use to store new data), would ensure that the colonists arrive to a fresh cache of the Web.

Our colonists have arrived on Mars. The biggest enemy now is latency: the worst possible experience is for a colonist to issue a search query or click on a link and have the response be, "Sorry, this content doesn't exist on Mars. Please stand by while it's being fetched from Earth. Estimated time of delivery: 24 minutes."

We envision solving this problem using the same type of technology that was deployed to create the sneakernet delivery to begin with. Each colonist would have an avatar on Earth that represents his or her user model — this avatar would continuously receive a stream of interaction and other data (from Mars), and based on these data, predictively fetch relevant portions of the Web on Earth, package up the content, and beam the material over to Mars to update/replace the cache there. For example, the avatar might observe from the colonist's personal diary that she's contemplating growing the first Martian bonsai, and proactively fetches relevant webpages that are related to the subject. The next morning, when the colonist starts searching about bonsai, the pages are already there — the search experience is seamless and she has no idea that the pages were only delivered last night.

Prefetching by the predictive models could be supplemented by further sneakernet deliveries from Earth, piggybacked alongside regular cargo missions or new waves of colonists — of the updated Netflix catalog, for example. Considering that much content (particularly TV shows) is available on Netflix only after a substantial delay, adding the transit time doesn't seem like a particularly big deal. It seems straightforward to weigh the benefits and costs of beaming versus rocketing bits over to Mars to select the appropriate delivery method.

Each colonist's avatar could also proxy websites for transactions such as Amazon purchases: from the Martian point of view, the experience would be seamless, but a final confirmation would arrive only after the actions have been relayed and applied "for real" on Earth. This would take some amount of software engineering, but seems eminently doable. It might be interesting to consider how you might trade equities from Mars, but let's leave aside asynchronous transactions in the remainder of this piece.

Back to search: the predictive user models in the avatars would have a different objective than the ones used to "bin pack" the initial cache. Whereas the latter are optimized to maximize expected utility per unit storage, the avatar's primary goal is to hide latency, and hence it might be more profligate in its use of bandwidth. However, the underlying principles are the same — we'd probably still be using some form of machine learning. All of this can be accomplished with today's technology.

It's interesting to speculate what sources of data the avatars could bring to bear on the prediction problem. Naturally, we would expect all the types of interaction data that Web search engines already capture today: queries, clicks, dwell times, and so on (yawn). Our colonists likely would be under constant (non-intrusive) physiological monitoring (heart rate, cortisol levels, amount of physical exertion, and so on) for health and safety reasons, which might provide interesting sources of signal. The avatar would also have access to all personal communications (such as voice/video messages and email), other personal files (for example, diaries), as well as official mission logs and reports. In addition to human-generated data, we would expect a multitude of sensor data, ranging from environmental monitors to the output of scientific experiments.

In short, it would be relatively easy to capture all data coming in, going out, and being generated on Mars — thus creating the opportunity for the ultimate big brother. We can justify gathering all these data, but this also creates interesting data privacy issues. Just because an avatar *could* take advantage of all these data doesn't mean it *should*. The avatar of each colonist should have tight security safeguards and be kept logically distinct — from other colonists and from eavesdroppers on Earth. However, it isn't hard to construct scenarios where information inadvertently leaks across avatars.

The scenario described above isn't fanciful science fiction, but already here today. Every user interaction with an online service is already being logged. Personal communication is already being monitored and captured, for example, by cell phone companies and Web-based email services. Online calendars, airline e-tickets, and GPS keep track of where we are and what we're doing nearly all the time. With the advent of cloud storage, our personal files are already "out there," and the same goes for personal physiological data through the proliferation of fitness and wearable devices. The only significant difference today is that all these data are gathered in silos, whereas on Mars, all data would be conveniently accessible (an important distinction that probably leaves many companies today salivating).

In terms of machine learning, the biggest difference between what's being done today and the rich user models necessary to support a seamless Web experience from Mars is the optimization objective. Somewhat oversimplifying, the focus today by most Internet companies is on short-term prediction — query typeahead and ad prediction are two great examples. The first strives to save users a few keystrokes and the second attempts to predict the next interaction (clicking an ad). In contrast, user models to support searching from Mars need to capture longer-term user interests, potentially over months and even years (in the case of scientific research being conducted on Mars). This represents an interesting direction in information retrieval and machine learning research, and recent work on modeling longer-term user engagement suggests that industry has begun to move in this direction.[14]

In summary, what does it take to make searching from Mars work? It's the user model, stupid!

Despite the technological and economic feasibility of colonizing Mars, there are presently no concrete plans to actually do it. This doesn't mean, however, that we should just sit idle waiting for Elon Musk to deliver. In fact, searching from a rural village in the developing world exactly parallels searching from Mars: instead of a cache of the Web on Mars, we have a cache of the Web at the Internet access point shared by the villagers.

Internet connectivity in the developing world is often intermittent and poor in quality: we can use the proposed techniques to hide latency from Mars to create a more seamless user experience for rural villagers in India (for example). We can even substitute a "robotic cargo ship" with a "Fedex delivery of hard drives" and the sneakernet concept is still applicable.[15,16]

Yes, we can build the search from Mars experience today to improve search on Earth. Let's do it!

Acknowledgments

We thank Doug Oard and Stephen Green for their helpful comments on earlier drafts.

References

1. N. Angier, "A One-Way Trip to Mars? Many Would Sign Up," *New York Times*, 8 Dec. 2014; www.nytimes.com/2014/12/09/science/a-one-way-trip-to-mars-many-would-sign-up.html.
2. E. Howell, "SpaceX's Elon Musk to Reveal Mars Colonization Ideas This Year;" *Space.com*, 9 Jan. 2015; www.space.com/28215-elon-musk-spacex-mars-colony-idea.html.
3. B. Aldrin, "The Call of Mars," *New York Times*, 13 June 2013; www.nytimes.com/2013/06/14/opinion/global/buzz-aldrin-the-call-of-mars.html.
4. J. Teevan et al., "Slow Search: Information Retrieval without Time Constraints," *Proc. 7th Ann. Symp. Human-Computer Interaction and Information Retrieval*, 2013, article no. 1.
5. W. Thies et al., "Searching the World Wide Web in Low-Connectivity Communities," *Proc. 11th Int'l World Wide Web Conf.*, 2002; www2002.org/CDROM/alternate/714/.
6. T. Ormston, "Time Delay between Mars and Earth," blog, *Mars Express*, 5 Aug. 2012; http://blogs.esa.int/mex/2012/08/05/time-delay-between-mars-and-earth.
7. B.S. Robinson et al., "The Lunar Laser Communications Demonstration," *Proc. Int'l Conf. Space Optical Systems and Applications*, 2011; doi:10.1109/SMC-IT.2009.57.
8. D.A. Vallado, *Fundamentals of Astrodynamics and Applications*, Springer-Verlag, 2007.
9. P.D. Wooster et al., "Mission Design Options for Human Mars Missions," *The Int'l J. Mars Science and Exploration*, vol. 3, 2007, pp. 12–28; doi:10.1555/mars.2007.0002.
10. F. Cain, "How Long Does It Take to Get to Mars?," *Universe Today*, 9 May 2013; www.universetoday.com/14841/how-long-does-it-take-to-get-to-mars.
11. G.R. Schmidt, J.A. Bunornetti, and P.J. Morton, "Nuclear Pulse Propulsion — Orion and Beyond," *Proc. AIAA/ASME/SAE/ASEE Joint Propulsion Conf. and Exhibit*, 2000; http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20000096503.pdf.
12. R.P. Russell and C.A. Ocampo, "Systematic Method for Constructing Earth-Mars Cyclers Using Free-Return Trajectories," *J. Guidance, Control, and Dynamics*, vol. 27, no. 3, 2004; pp. 321–335; doi:10.2514/1.1011.
13. K. Fall, "A Delay-Tolerant Network Architecture for Challenged Internets," *Proc. 2003 Conf. Applications, Technologies, Architectures, and Protocols for Computer Comm.*, 2003, pp. 27–34.
14. M. Lalmas, "Evaluating the Search Experience: From Retrieval Effectiveness to User Engagement," keynote talk, Conf. and Labs of the Evaluation Forum, 2015; http://clef2015.clef-initiative.eu/CLEF2015/keynotes.php.
15. A. Pentland, R. Fletcher, and A. Hasson, "DakNet: Rethinking Connectivity in Developing Nations," *Computer*, vol. 37, no. 1, 2004, pp. 78–83.
16. A. Seth et al., "Low-Cost Communication for Rural Internet Kiosks Using Mechanical Backhaul," *Proc. 12th Ann. Int'l Conf. Mobile Computing and Networking*, 2006, pp. 334–345.

Jimmy Lin holds the David R. Cheriton Chair in the David R. Cheriton School of Computer Science at the University of Waterloo. His research lies at the intersection of information retrieval and natural language processing, with a particular focus on Big Data and large-scale distributed infrastructure for text processing. Lin has a PhD in electrical engineering and computer science from MIT. Contact him at jimmylin@uwaterloo.ca.

Charles L.A. Clarke is a professor in the David R. Cheriton School of Computer Science at the University of Waterloo. His research interests include information retrieval, Web search, Web data mining, text data mining, and software tools. Clarke has a PhD in computer science from the University of Waterloo. Contact him at claclark@plg.uwaterloo.ca.

Gaurav Baruah is a PhD student in the David R. Cheriton School of Computer Science at the University of Waterloo. His research interests include information retrieval, search engine evaluation, and data mining. Baruah has a master's degree in computer science and engineering from the Indian Institute of Technology, Guwahati. Contact him at gbaruah@uwaterloo.ca.