Teaching Computers to Read About Mars

Kiri L. Wagstaff
Calvert N. Ellis Memorial Lecture, September 26, 2017

Kiri Wagstaff is a computer scientist and researcher at the Jet Propulsion Laboratory, California Institute of Technology.[1]

I am here tonight to talk about my favorite planet and my favorite project right now at the Jet Propulsion Laboratory (JPL). This work is at the intersection of planetary science and information extraction technology. This unusual convergence has yielded a system that can greatly benefit not only scientists and researchers but also anyone in the public who wants to know more about Mars.

I have several talented colleagues at universities and research labs as well as graduate and undergraduate students who have contributed to this work: Raymond Francis, Thomas Gowda, You Lu, Ellen Riloff, Karanjeet Singh, and Nina Lanza. Our work builds on that of a whole community of scholars, and I am very glad to get to share it with you tonight.

EXPLORING MARS WITH ROVERS

I want to take you on a little journey to Mars and give you an idea of what it looks like from the perspective of our exploration agents, which are robots. Figure 1 shows the Curiosity rover taking a picture of itself. That is why it looks a little distorted, because it gets the same selfie fisheye effect as when you hold up your phone and you are a little too close to it. Here the rover is holding up its "arm" to take the picture. This is the only way that we can see what our rover looks like on Mars, since there is nobody else there to photograph it. I think that is why I love this picture: we can see the rover as it truly is. This is not an artist's conception.

Figure 1. Mars rover Curiosity "selfie" at the John Klein drill site (Feb. 2013). Credit: NASA/JPL-Caltech/MSSS.

How does this rover explore Mars? We send detailed instructions to our rovers that tell them where to go, what pictures to take, what data to collect, what is interesting, and what is not interesting. We compose a list of instructions and the rover faithfully executes them. It collects all the data and sends it back to us here on Earth. We take some time to think about that data and decide what it means, what the next step should be, and where we should go next, before we send new commands to the rover. This is the ideal daily cycle: new commands go up, new data gets collected, new data comes back, and we learn new things.

But this is a simplified view of Mars exploration because it is short-term; it only considers the next step to take. Yet this mission has been on Mars for five years, so you might reasonably wonder what we have learned over the span of those five years. That requires a little more history than just what the rover did yesterday and what it should do next. The larger picture is captured in publications that scientists write. They have had more time to think about the data than just overnight. They digested them to tell us not only what the rocks are made of, but which rocks are unusual and which ones yielded new discoveries. Of course, everybody is interested in finding evidence of past water on Mars, but there are places where we found things we did not even predict would exist, like fluorine. Not every discovery ends up in a press release. You might not know about it unless you dug through the scientific literature, and that is a barrier.

Figure 2. Three images collected by the Curiosity rover showing layered rock, the rover's wheels, and a drill hole. Credit: NASA/JPL-Caltech/MSSS.

Figure 2 shows examples of some of the images that this rover collects. On the left is an image of layered rock that could give us clues about the past history of flowing water. In the middle is an image taken from under the rover to examine the state of its own wheels. On the right is the result of the rover drilling into a rock. Drilling allows us to determine what the rock is made of, not just what the surface dusty layer is made of. After five years of exploration, we have created a lot of graffiti on Mars by making these drill holes. We are perturbing the environment as we are studying it, and we take pictures of the artifacts that we have left behind. The rover also leaves tracks everywhere it drives. You can see these from Mars orbit. You can look down and trace where the rover has driven by finding its tracks.

All of these images are archived by NASA. Anything I show you here today is yours, since all of the data were collected with taxpayer money. There are no premiums or copyrights or access fees for these data. I encourage you, if you are curious, to explore it because it is yours. You can go to the Planetary Image Atlas[2] and do a search, right now if you want, although I hope you'll pay attention here to the talk. Maybe take a look after the talk.

This rover is called Curiosity, but its official name is the Mars Science Laboratory to reflect that it is doing science on Mars. It has collected half a million images with its mast camera, a million and a half images with its hazard camera, and six million images with its navigational camera. That will keep you busy for a while, right? There is such a wealth of data that it can be overwhelming. I can point you at these data and then what do you do with it, how do you search through it, where do you find the good stuff?

Sol	Activity	Site
1702	DRT Brush "Fern Spring" Target, In-Situ Observations of "Pulpit Ledge" and "Fern Spring" Targets, and Targeted Remote Sensing	63
1701	Remote Sensing	63
1700	Drive Along MSAR, In-Situ Observations of "Ripple Pond" Target, APXS Cleaning, Drill Diagnostics, and Targeted Remote Sensing; SAM Getter Scrubber Cleanup	63
1699	Remote Sensing	63
1698	Drive toward "Ripple Pond" Target, In-Situ Observations of "Woodland Ledge" Target, APXS Cleaning, and Targeted Remote Sensing	63
1697	Remote Sensing	63
1696	Drive Southeast, MAHLI Observations of Calibration Targets, APXS Cleaning, and Targeted Remote Sensing	63

Figure 3. MSL Curiosity Analyst's Notebook list of per-sol activities from sol 1696 to 1702. Screenshot from https://an.rsl.wustl.edu/msl/mslbrowser .

There is a tool called the MSL Curiosity Analyst's Notebook[3] that is trying to help you get one step closer to what the data mean. I did not develop this tool, so I will just make an advertisement for it: our friends at the Geosciences node of the Planetary Data System set this up. You can go there and get a timeline of the mission and see what the science objectives were (see Figure 3). If you click on an individual day (sol), you will see all of the data (pictures, spectra, etc.) collected on that sol. A day on Mars is called a sol since it is not exactly the same length as an Earth day, and we do not want to confuse people. A sol is one rotation of the planet, which on Mars is about forty minutes longer than the Earth day. You can read the narrative that goes along with what these data were intended to study: why these were collected, or why were we looking at the rover's wheel that day. You can find out because it is all annotated by the scientists and mission planners. I highly recommend reading it.

This rover has spent over 1,500 days on Mars. You are probably thinking, "I do not really want to read 1,500 days of descriptions of Mars activities." You might prefer to get a pointer to the good stuff, where we discovered something exciting or new. This is the real question that has motivated our work. What have we learned from five years on Mars? What do we get for all of that money we invested, all of the time, the scientists' and engineers' efforts? In addition, countless student hours of effort in the form of internships and other contributions have gone into making these what they are today. There is a lot of human effort, not just dollars, involved.

To address that question, we turn to information that humans write about their discoveries. One of the things they write about are observational targets on Mars, which are usually individual rocks or soils. Instead of calling them "target one," "target two," "target three," and so on, Mars scientists give them nicknames. You will see things like the John Klein drill site, the Cumberland drill site, the Sheepbed formation. This is a uniquely human thing: we find it easier to talk about something when it has a name. The way we distinguish one thing from another is to give it a name. Papers written about Mars locations do not refer to rock #527; they instead say "John Klein." And that makes it a little confusing for those of us who want to search through written text. Is "John Klein" a person, a rock, a soil, or a sand dune? In this case, John Klein was an actual person who lived on Earth. He was part of this rover's mission, and he contributed to its design and implementation. Unfortunately, he passed away before the rover landed on Mars. The mission, as an homage to him, named the first site that the rover drilled after him. There is a lot of emotion and human investment in these named targets that have meaning far beyond their composition, which is the scientific thing that we seek to discover.

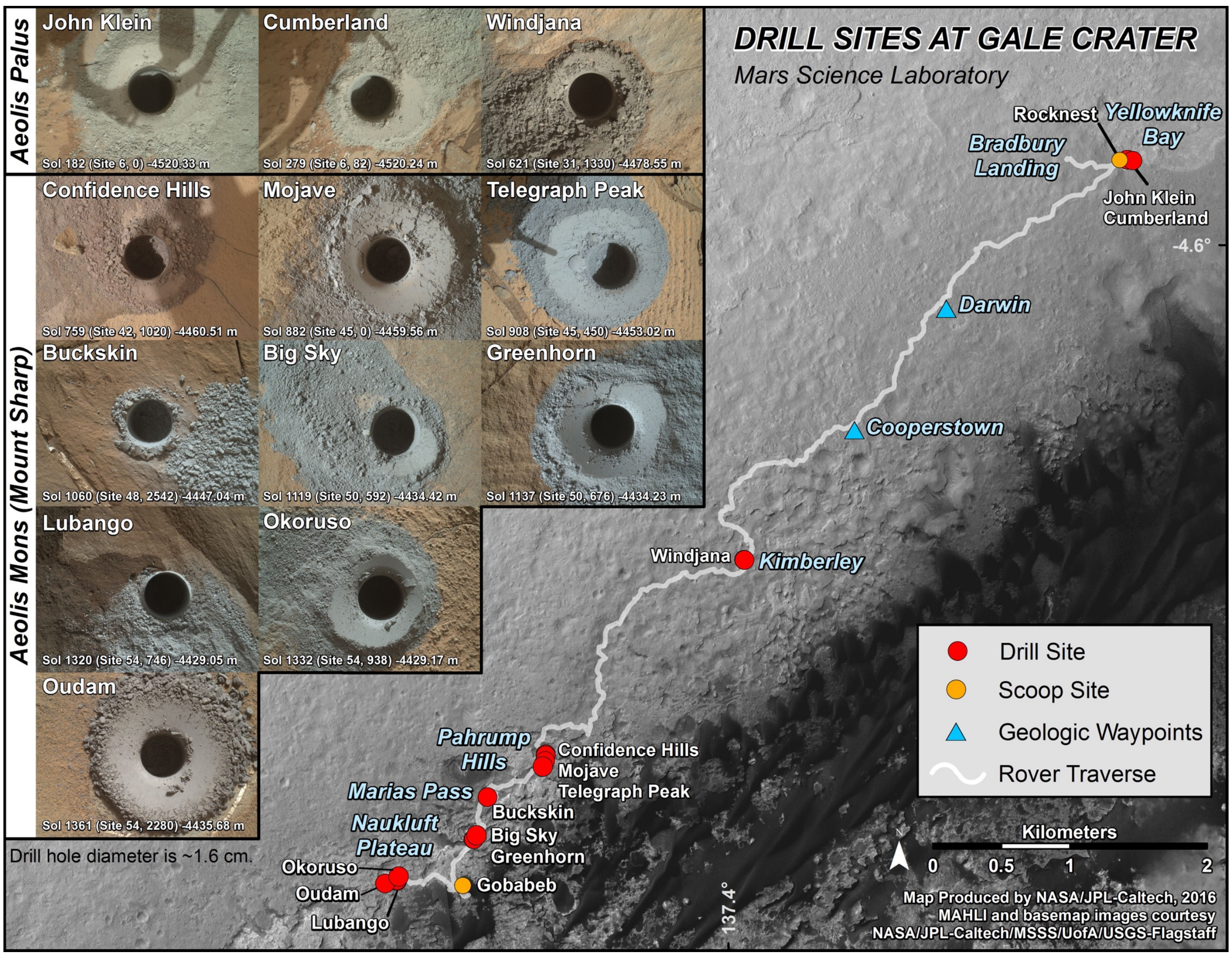

Figure 4. Twelve rover drill sites in Gale Crater, Mars.
Credit: NASA/JPL-Caltech/MSSS/UofA/USGS-Flagstaff.

Figure 4 shows a montage of different drill holes that the rover has created. Mojave, Big Sky, Buckskin: these are all Earth place names. The drill holes do not share properties with the Earth locations;

they are just convenient names. Figure 4 also shows a plot of the rover's traverse. It landed at the Ray Bradbury landing site (many of you will recognize that name), and then it drove off to Yellowknife Bay, which is a place in Australia, and then it drove along to Darwin, also in Australia, and so on. Pahrump is in Nevada. And now Mars as well.

If I wanted to find out what is known about, say, the Buckskin drill site, I can go to Google and search for "Buckskin." As you might predict, it is not going to give me very good results. None of those are relevant to the Mars target. How can I improve this search to find out information about the Mars Buckskin? I need to be more specific in my search. If I search for "Mars Buckskin," I find in the top four results that three of them refer to the Mars target I wanted. Unfortunately, this only works for things where someone has written about them on the internet that Google has indexed. Most of our hundreds of thousands of Mars targets do not have web pages dedicated to them. For example, I tried searching for "Mars Lubango," which is a different drill site, and I could not find anything relevant. In fact, I got all French results, probably because the word "mars" is also the word in French for the month of March. Maybe this target never had a website published about its contents and that is why I cannot find anything. That is disappointing.

Further, I want to be able to ask questions that are not even about specific targets. I want to know which of the Mars targets that we have studied contain a specific element, like fluorine, or a specific mineral, like hematite. Where is there hematite on Mars? We have been studying Mars for so many years, and I cannot even ask that question. What about higher-level questions about the scientific community: do we have consensus about what a given target is made of? How does scientific opinion evolve over time? Even if they are looking at the same data, people may perform different analyses and reach different conclusions that supersede the earlier ones. What we really need is to analyze not the Web but the scientific literature that is written about Mars. That is time-consuming for humans and tends to require scientific training. Can computers help us with this problem? Can they analyze and understand written human language?

MACHINE READING

You may be familiar with IBM's Watson system, which was able to beat humans at *Jeopardy!*. Watson is a highly specialized, highly trained, artificial intelligence that could not only read the clues in the game but also read through an enormous amount of information behind the scenes. Watson had to be able to search better than Google, and with more understanding than Google, to assemble structured knowledge about states, capitals, people, history, geography, and all of the other categories that a good *Jeopardy!* player would know. Its success was a landmark moment for artificial intelligence because, like chess before it, we thought this was something only humans could do well. Could we use a similar approach to process a lot of text, pull out the structured pieces, and answer reading comprehension questions for science?

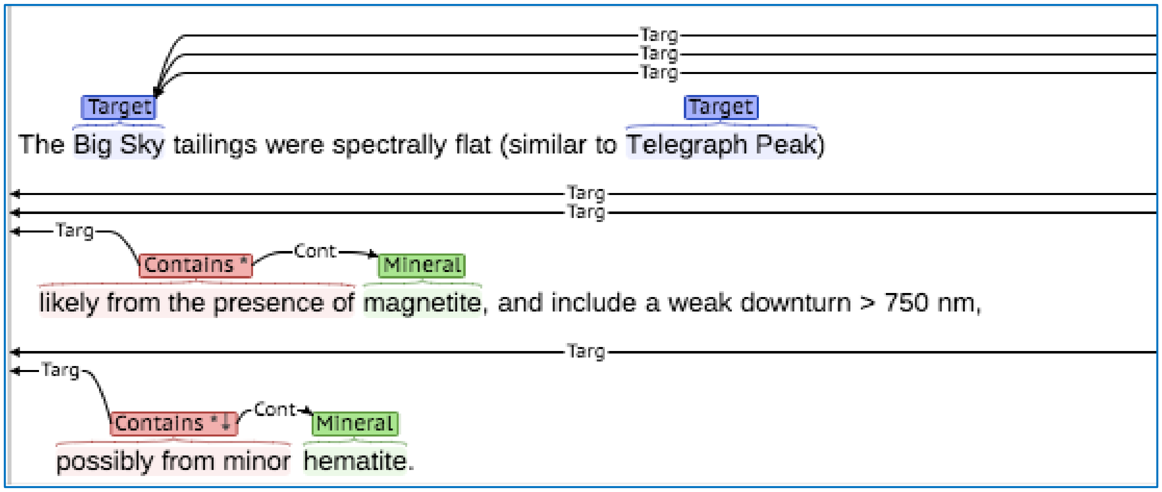

Figure 5. Excerpt from a paper by Jeffrey R. Johnson[4], annotated to show targets, minerals, and compositional relationships ("contains").

It turns out that scientific papers are much more challenging. Figure 5 shows a single sentence from a scientific publication that talks about Mars targets. If I want to extract information about targets that contain specific minerals, then I can ask the reading comprehension question: does Big Sky contain magnetite? "Likely." Maybe. As a reader, I do not know. What about whether it contains hematite? "Possibly." Does Telegraph Peak contain magnetite? I think most of us would be reluctant to conclude that. There is some similarity here, since they are both "spectrally flat," but the author is probably only talking about Big Sky in terms of magnetite. But you really have to think this through to be able to answer the question. This is not like "what is the capital of Tunisia?" That is something I can easily look up in an atlas or a gazette. Even Google could tell me that answer, with probably 100% certainty.

If these papers were written to say "X contains Y," something very concrete, confident, and simple, then this task would be very easy. I could scan through all of these documents, pull out all of those statements, and put them in a database where you could search it. You could answer questions like: "Where do we find fluorine?" You could answer even the high-level question, "Is there consensus about what Big Sky is made of?" because I could give you all the results for Big Sky and you could immediately see whether they agree or do not agree.

Instead, scientific writing is one of the hardest forms of writing to analyze, as you may know if you are student who has to read papers for a class, like papers from the scholarly literature, not just a textbook, which is already hard enough. The writing is dense and it uses complex grammar, a lot of passive voice constructions, and complex noun phrases. Nowhere in this sentence does the word "contains" appear, even though that is the relationship being described.

The authors are not deliberately trying to make reading difficult. The ambiguity inherent in "likely" and "possibly" is being expressed in what is called "hedging language" that allows the writer to avoid full commitment. This is important, because there are many things we do not know with certainty. This sentence is talking about a pile of dust on a planet 140 million miles away. The authors are working with the data available, but no one is there on Mars, and no one can put Big Sky in a lab to do really detailed analysis. It is all very indirect. So the best they can tell us in this case is that it is probably magnetite. There is only a weak downturn in the spectrum, so it could be minor hematite, but this could be wrong. Maybe someone else will analyze more data later and give us a better answer. The uncertain language is difficult but necessary.

Further, it can be a challenge just to recognize real Mars targets in text. Big Sky is a resort in Montana; it is a movie from 2015; it is a lot of other things. Telegraph Peak is a mountain in California, among other things. All of these aspects make it very hard for a computer to read these papers for us.

In our work, we built a system to tackle this problem, despite the difficulties. We took all of the publications from one specific venue, the Lunar and Planetary Science Conference. It generates about 2,000 new papers every year, so after three years, we ended up with almost 6,000 documents. Fortunately, these are not big journal papers. We started here on purpose. Each paper is an extended abstract that is only two pages long. Altogether we have seven million words, and we want to go through and find the ones that teach us something about Mars.

THE MARS TARGET ENCYCLOPEDIA

From those words, we built the Mars Target Encyclopedia (MTE). This is an automatically generated searchable encyclopedia, and our vision was it would have an entry for every target on Mars that collates in one place everything that has been published about that target. What is the target made of and what do we know about it? It would also show you individual excerpts from the source publications to let you immediately check each statement. The MTE would also link to the original publication, so if you are interested, you can click through and read the full document yourself.

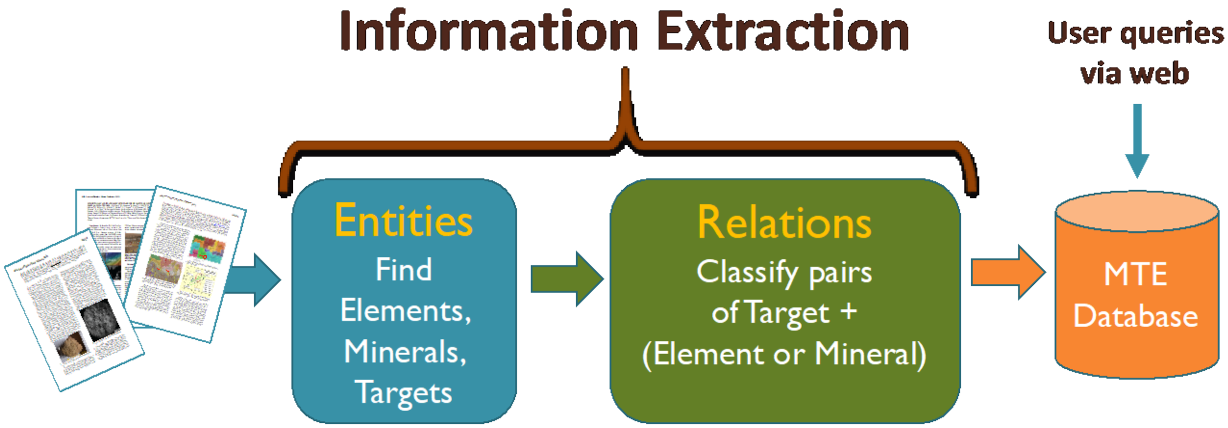

Figure 6. The Mars Target Encyclopedia reads scientific papers to find entities, compositional relations between entities, and then store the result in a database with a searchable web interface.

To achieve this goal, the MTE must be able to read thousands of papers and find compositional statements about what elements or minerals are in a given target. Figure 6 shows the process of going from a PDF document to a searchable database. We break this into two stages. The first step is to find all of the "entities," which are places where a target, element, or mineral is mentioned. For elements, we have the periodic table. For minerals, you might be surprised to learn that there are over 5,000 different minerals defined by the International Mineralogical Association.

However, the most challenging group is the targets themselves. I hope you have already gotten a sense of that, with Big Sky and Telegraph Peak and John Klein. Target-ness is not obvious from the name like it is for a mineral, where most of them end in '-ite'. The Curiosity mission can provide a list of target names, but that list is continually growing, unlike the periodic table. We identify new targets almost every day from Mars. The list is always out of date. Therefore, we wanted to go beyond blindly applying lists to find terms and instead use machine learning to help automatically learn the new names. This gives the MTE the kind of adaptability that humans have. Even if I only gave you a few examples, you probably would be able to find many of these new targets inside the text. You see something like "the Big Sky tailings were spectrally flat." Even if you don't know what "spectrally flat" means, if you saw another sentence that said "the John Klein tailings were spectrally flat," you would easily conclude that "John Klein" is also a target. You are very good at pattern recognition.

We want the computer to learn to do this as well. There is a system called CoreNLP from Stanford that allows you to train a custom model to do exactly that.[5] Stanford does not care necessarily about Mars targets, so their system does not yet know about them. We trained it with our own data, by going through and marking them up the way I showed you in Figure 5. CoreNLP reads through those manually annotated documents and learns a model to find elements, minerals, and targets in new texts.

This is a pretty standard approach in information extraction or machine reading to help find the entities of interest. And how well does it work?

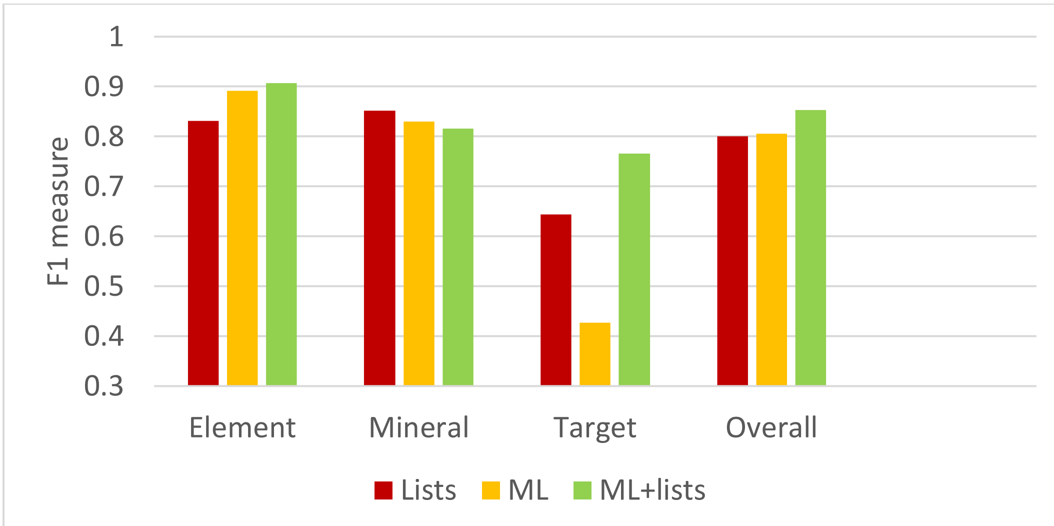

Figure 7. Entity recognition performance using lists, machine learning (ML), or both.

Figure 7 shows that just using the lists (the periodic table, the mineral list, and our list of known targets from the mission) works pretty well to find elements (about 83%), better for minerals (85%), and not as well for targets (64%). Using machine learning (CoreNLP) on its own improves element detection, does worse on minerals, and does abysmally on targets (43%). However, using the lists and the machine learning in concert yields the best overall performance, and it dramatically improves the performance on targets (77%), which is exactly what we were hoping for..

A limitation of the lists is that they can also refer to things that are not targets. Anytime I encounter the word "hydrogen," it is probably talking about the element. But it is not the case that every time I see the word "Ithaca," it is a Mars target. In fact, when I say "Ithaca," your first thought is probably not "oh yeah, that one on Mars," right? There are Earth Ithacas in New York and in Greece. We reuse names a lot on the Earth, and now we are doing it on Mars. If you see "Ithaca" in one of these documents, it could mean the Mars target, but it could also mean the one in New York because Cornell University has some very active planetary science researchers that publish a lot of papers about Mars. They include their affiliations in the papers, so the same paper could include Ithaca, New York, and also the Ithaca target on Mars. The authors know that, as a human, you can tell the difference, so they will not say "Ithaca on Mars," they will just say "Ithaca contains high levels of calcium" and you will know that they are not talking about New York. It is much harder for the machine to do that. But we can improve with machine learning, and even though it is not perfect, it is getting us most of the way to finding what we want to find.

The second step in the MTE is to decide when there is a relationship between targets and elements and minerals. For example, consider the sentence "the mineralogy of Confidence Hills is dominated by plagioclase, augite, and hematite." This is one of the clearer examples of a compositional relationship, yet it still does not use the word 'contains' or 'composition' or anything like that. The goal is to make a yes or no decision about whether a given pair has a compositional relationship. A simple rule that votes "yes" whenever a target and a mineral appear in the same sentence does pretty well (59%). However, it makes a lot of mistakes, especially false positives in which it claims there is a relationship when there isn't one. To reduce this kind of mistake, we use a machine learning method called jSRE.[6] We gave it many positive and negative examples, and it learned the relevant patterns. This approach is a bit lower on overall performance (53%), but it makes far fewer false positive mistakes. That is the right direction to go, because we would rather have our database miss some relationships than include ones that are incorrect. This is still an open area of research.

After this testing, we turned the computer loose on all 6,000 documents. While it took us humans about half an hour to read each document, the MTE processed each one in about five seconds. Compared to the contents of the 118 documents that we had manually annotated, the computer found many more elements, minerals, and targets and more than doubled the number of relations. You might be wondering: if we went from 118 to 6,000 documents, how come the number of relations only doubled? The answer is that we had already annotated the most relevant documents that discussed Mars targets. The Lunar and Planetary Science Conference also includes papers about the moon, other planets, comets, and more.

I will show you a couple of examples of what the MTE found.

- "<u>Link</u>, which was one of the first <u>K</u>-rich conglomerate targets observed with ChemCam…" The MTE concluded that target "Link" contains "potassium." By the way, please never name anything Link; it is far too generic. However, Link is a real Mars target. As you can see, the MTE has to be able to find abbreviations like "K," not just the fully spelled out element names.
- "The RN crystalline component is depleted in MgO and FeO relative to <u>JK</u> and <u>CB</u> because of the absence of <u>olivine</u> and enrichment of <u>magnetite</u> in the latter." "JK" and "CB" were not in our target list, because people made them up. They got tired of typing "John Klein," so they used "JK." "CB" is "Cumberland," and "RN" is "Rocknest." You would have to read the entire paper to see early on where they define those abbreviations. The MTE was able to find them due to the surrounding language.

There are also some mistakes:

- "The results indicate that the dip of the Shoemaker Fm impactite section…" The MTE concluded that "Shoemaker" contains fermium, but instead this is a reference to the Shoemaker formation. There is no compositional relationship.
- "Finally, the Bilanga diogenite has a model age that seems older but still similar within the error than basaltic and cumulative eucrites." In this case, there is a Bilanga target on Mars, but this sentence is talking about a meteorite on Earth also called Bilanga. In addition, "diogenite" is a type of meteorite, not a mineral. These are subtle challenges.

However, most of the MTE's findings are correct, and you can now search the resulting database. You can search for a target named Dillinger and find that the MTE knows about seven properties and three different publications about Dillinger.

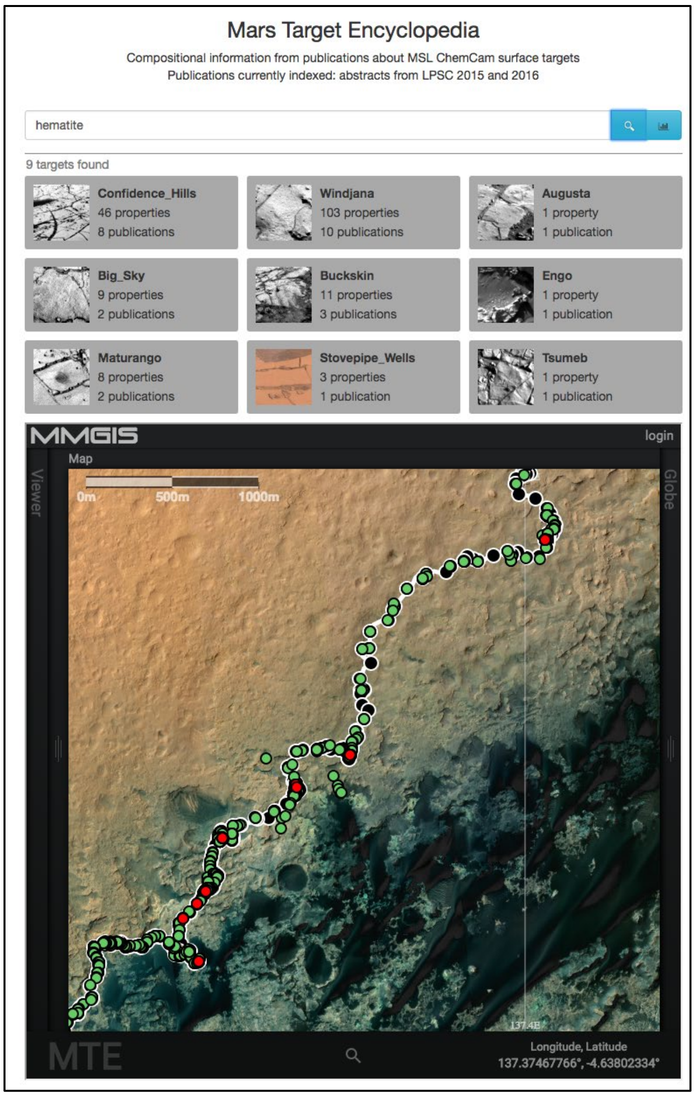

Figure 8. MTE search for "hematite" showing search results and spatial distribution on Mars.

You can also create queries that were never possible before, like "Where have we found hematite on Mars?" Figure 8 shows nine search results and highlights in red every Curiosity rover site from which a finding about hematite was published.

To summarize, the goal of this project was to connect Mars rover data to what has been published about Mars and be able to answer questions about what we have learned from the mission. We want everyone to be able to easily search and access this knowledge. The Mars Target Encyclopedia is also enabling us to ask questions that could not be previously answered without a lot of human effort. What is most exciting to me personally is that this is an example of AI helping us do better science. It is helping us push the boundaries of what science is and what is accessible beyond what one human mind or one pair of eyes can read. You do not have to be a scientist anymore to access this information. Anyone can type in a search box and find out where the hematite is in these publications.

NOTES

1. The work described in this talk was carried out in part at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. I want to thank the JPL MGSS Program, the NASA Planetary Data System, and the Mars Science Laboratory project for funding this work. © California Institute of Technology. Government sponsorship acknowledged.

2. "PDS Image Atlas." NASA Jet Propulsion Laboratory. https://pds-imaging.jpl.nasa.gov/search/ (Accessed June 3, 2019).

3. "MSL Curiosity Analyst's Notebook." NASA Planetary Data System GeoSciences. https://an.rsl.wustl.edu/msl/mslbrowser (Accessed June 3, 2019).

4. Johnson, J. R. et al. (2016) "ChemCam passive reflectance spectroscopy of recent drill tailings, hematite-bearing rocks, and dune sands," *Proceedings of the 47th Lunar and Planetary Science Conference*, Abstract #1155.

5. Finkel, J. R., Grenager, T., and Manning, C. (2005) "Incorporating non-local information into information extraction systems by Gibbs sampling," *Proceedings of the 43nd Annual Meeting of the Association for Computational Linguistics,* pp. 363–370.

6. Giuliano, C., Lavelli, A.,and Romano, L. (2006) "Exploiting shallow linguistic information for relation extraction from biomedical literature," *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics,* pp. 401–408.