Turing and the Integration of Human and Machine Intelligence

Susan G. Sterrett
Wichita State University

Turing and the Integration of Human and Machine Intelligence

S. G. Sterrett [1]

Abstract Philosophical discussion of Alan Turing's writings on intelligence has mostly revolved around a single point made in a paper published in the journal *Mind* in 1950. This is unfortunate, for Turing's reflections on machine (artificial) intelligence, human intelligence, and the relation between them were more extensive and sophisticated. They are seen to be extremely well-considered and sound in retrospect. Recently, IBM developed a question-answering computer (*Watson*) that could compete against humans on the game show *Jeopardy!* There are hopes it can be adapted to other contexts besides that game show, in the role of a collaborator of, rather than a competitor to, humans. Another, different, research project --- an artificial intelligence program put into operation in 2010 --- is the machine learning program NELL (Never Ending Language Learning), which continuously 'learns' by 'reading' massive amounts of material on millions of web pages. Both of these recent endeavors in artificial intelligence rely to some extent on the integration of human guidance and feedback at various points in the machine's learning process. In this paper, I examine Turing's remarks on the development of intelligence used in various kinds of search, in light of the experience gained to date on these projects.

---

[1] S. G. Sterrett, Curtis D. Gridley Distinguished Professor of the History and Philosophy of Science, Department of Philosophy, Wichita State University, Wichita, Kansas USA 67260
email: susan.sterrett@wichita.edu

1. Introduction: Isolation, Interference, and Immersion

In his 1948 technical report "Intelligent Machinery: A Report by A. M. Turing", in the course of exploring an analogy between the *education of a human*, where the education is designed to allow a human to achieve his or her intellectual potential, and the *education of a machine*, where the education is designed to result in a machine analogue of it, i.e., in allowing a machine to achieve *its* intellectual potential, Turing makes the observation that "an isolated [human] does not develop any intellectual power." [2] (Turing 2004a, p. 439-440; Sterrett 2012) What he means here by "an isolated [human]" is a human who is isolated from other humans, not a human isolated from contact with everything else in the world.

Turing's remark about the intellectual possibilities of humans isolated from other humans is especially notable, since it is made in a paper that also discusses trial and error approaches to learning. Given the parallels often drawn between trial and error methods of learning and adaptation by natural selection, one might expect that, instead of this comment that a human in isolation of other humans does not develop any intellectual power, an investigation into the potential that trial and error approaches hold. That is, one might expect a more sanguine investigation into what trial and error methods are able to effect (i.e., bring about) for an individual exploring the world on its own, so to speak. Such research projects are concerned with asking: what kind of priming with what kind of examples and what kind of input devices does a machine need [in order to be able to perform a certain task autonomously]? There are approaches to machine learning on which learning of certain kinds can take place due to the interaction of one individual system or machine and its environment, even if the immediate environment with which it interacts happens to be devoid of humans and of others like itself. Some unsupervised machine learning algorithms, such as clustering algorithms, are examples of this.[3] But, upon close examination of Turing's examples, it is clear that the kind of learning under consideration in his investigations is not of this sort.

Rather, in Turing's writings on intelligent machinery, the methods he considers for providing the machine analogue of the education of a human all seem to involve human interaction with the machine in some way or another, and for a significant part of the machine's development process.

---

[2] Where I think Turing was indicating "human" or "humans" in using the term "man" or "men", I may paraphrase or replace words within brackets accordingly.

[3] In *supervised learning*, all of the examples given the computer program during training are labelled; in *semi-supervised learning*, the computer program is provided with labels for only a subset of the examples it is given to deal with, and in *unsupervised learning,* none of the examples the computer program is to deal with is labelled.

Sometimes the human is involved in a teaching role, i.e., actively and intentionally interacting with the child-machine in order to effect behavior of a certain sort: Turing mentions both a method using reinforcement via inputs for analogues of "pain" and "pleasure" (Sections 10 and 11 of the paper (Turing 2004a, p. 433 ff)) and a method using the programming of "principles." (Section 12 (Turing 2004a, p. 438)). However, such explicit training of the machine by humans is not the *only* way of integrating human expertise via the 'education' of a machine. In discussing various stages of the education of a human (and, analogously, how he imagines the education of a machine might proceed), Turing considers the phenomenon of humans learning irregular verbs: "By long experience we can pick up and apply the most complicated rules without being able to enunciate them at all." He suspects some kinds of machines he considers (e.g., the P-type machine, an "unorganized" machine he proposed to train using machine analogues of pleasure and pain inputs delivered by a human), of behaving similarly, due to their "randomly distributed memory units." The key remark he makes here is that "built-in teaching procedures could not help." (Turing 2004a, p. 438) Yet his hope is not empty; it amounts to the hope of there being a machine analogue of the human ability to "pick up and apply" rules -- even "the most complicated" rules.

2. Teaching Searching

What is the human process for which Turing hopes there is a machine analogue, if not "built-in teaching procedures"? That might depend on the type of intelligent machine behavior under discussion. Turing argues for the generality of search problems, i.e., that a "very great variety of problems" displaying initiative can be characterized as problems whose solutions are solutions to a search problem. To get a general idea of his thinking here; he gives an example and argues:

> ". . . the problem is clearly equivalent to that of finding a program to put on the machine in question, and it is easy to put the programs into correspondence with the positive integers in such a way that given either the number or the program the other can easily be found. We should not go far wrong for the time being if we assumed that all problems [requiring some sort of initiative] are reducible to this form. " (Turing 2004a, p. 438 - 439)

I take it that by "this form" of a problem, he means the form of a problem such that searching for the right program is a matter of finding, or computing, the positive integer that corresponds to it. The details of his argument can be put aside for our purposes; what is of interest for us here is that he goes on to distinguish three types of problem, each of which can be characterized as a variety of search. The three varieties of search he identifies in his discussion, listed below in the order in which he mentions them in that discussion, are:

(a) *intellectual* search,
(b) *genetical* or *evolutionary* search, and
(c) *cultural* search.

Intellectual search appears to be a matter of insightfully reformulating or 'transforming' problems so that mathematical and logical methods can be used to organize and carry out the search for a solution to the problem at hand. Turing gave a brief definition of intellectual searches: "They might very briefly be defined as 'searches carried out by brains for combinations with particular properties'." The kind of problem just referred to in the preceding quote (a problem of finding a program with certain features) is an example of intellectual search.

Now, as to how humans are involved in the development of a machine capable of intellectual searches: It appears that what Turing thinks is that, for a single machine, one good human teacher is all that is needed. Turing speaks of how one would work with a universal machine to get it to the point where it could carry out intellectual searches; " [W]e first put a program into it which corresponds to building in a logical system (like Russell's *Principia Mathematica*). This would not determine the behaviour of the machine completely: at various stages more than one choice as to the next step would be possible." (Turing 2004a; p. 439)

He goes on to explain how a human could interact with such a machine in a way that might be expected to eventually result in a machine capable of carrying out intellectual searches; the details are not important to us here. What is important is this: even if the machine has no other interactions with humans than this one human teacher, there is no question of the entire process occurring completely outside of, or in isolation of, human society: after all, the teacher is not isolated from human society. Turing recognizes the crucial importance of human society in "organizing", or educating a human; he seems to think that it is *in the interaction of a human with other humans*, rather than in the physiology of the brain or the genetics that gives

rise to it, that the organization of the cortex responsible for human intelligence lies. It is in the training of the cortex that what is responsible for human intelligence arises. That one should not look to physiology alone for the key to human intelligence is brought out in a sort of thought experiment:

> ". . . the possession of a human cortex (say) would be virtually useless if no attempt was made to organize it. Thus if a wolf by a mutation acquired a human cortex there is little reason to believe that he would have any selective advantage. If however the mutation occurred in a milieu where speech had developed (parrot-like wolves), and if the mutation by chance had well permeated a small community, then some selective advantage might be felt. It would then be possible to pass information on from generation to generation. However this is all rather speculative." (Turing 2004a, p. 433)

What Turing has done here with the oft-repeated claim that the difference between humans and animals is a matter of being able to speak bears emphasis. In these understated remarks, he imagines what it would take for animals to have speech in a sense that would have significant consequences. It is noteworthy that he distinguishes between a wolf with the physiological ability to speak and a small community of wolves that uses speech to pass information on from generation to generation; only in the latter case would the capability for speech make enough difference for it to be preserved by means of natural selection. Turing does not venture a guess here as to what kind of intellectual abilities would arise in such a community of parrot-like wolves in which speech had conferred some selective advantage but, as wolves live in packs and hunt as a team, it is not hard to imagine the kinds of behaviors in a pack of parrot-like wolves for which speech might confer some selective advantage.

3. Natural Searching

The issue of selective advantage does arise in the next variety of search he discusses, (b) *genetical or evolutionary search*, which is briefly described as: "a combination of genes is looked for, the criterion being survival value." Here, unlike some other commentators on this passage (Koza 1999), I take Turing to be speaking about genetical or evolutionary search *per se*, i.e., literally, rather than conflating the point he makes in this passage with other points in later papers (e.g., his 1950 "Computing Machinery and Intelligence") in which he is explicit

that he is only invoking *an analogy* with natural selection. I take his remark here that "The remarkable success of this search confirms to some extent the idea that intellectual activity consists mainly of various kinds of search." to mean that, since the approach of regarding something as incredibly involved as finding a biological organism with certain features as a kind of search in a combinatorial space had turned out to have been successful in recent scientific research, that his idea that intellectual activity consists mainly in search had gained in plausibility as a result. Thus, "genetic or evolutionary search" may be a sort of model or ideal of the kind of search he means, but it is clearly distinguished from it: when characterized as a search, it is a search for an organism having a combination of features that results in the organism having high "survival value." In the 1950 paper, what we find is an analog of it: there, he explicitly identifies the analog of natural selection as "Judgment of the experimenter" (Turing 2004b, p. 469) and the surrounding text indicates that the experimenter will keep trying different machines to see how well each one learns. As this experimenter would be selecting machines for their value in terms of learning ability, it is not the same kind of search in terms of what is being selected for as we find in this 1948 report, "Intelligent Machinery."

Besides the reference to selective advantage in the pack of parrot-like wolves in the 1948 paper, and the other references to genetics ("It clearly would not require any very complex system of genes to produce something like the A- or B- type unorganized machine." Turing 2004a, p. 433) in the 1948 paper, I think there are other good reasons to take Turing literally here, and these are based on looking at the historical context of the work. The historical context with respect to genetics is that, in the summer of 1948, when Turing's "Intelligent Machinery" paper was written[4], the scientific research community in genetics was attempting a return to normalcy following the war. In 1948, the eighth meeting of the International Congress on Genetics was held, and it was a huge affair, both in size and in significance.

It was the first time the Congress had convened since 1939, the delay being due not only to the interruption of the war, but also to political issues in the science of genetics. It was the first such meeting that had been held since the "evolutionary synthesis" of the science of evolution

[4] The dating of the work as composed in the summer of 1948 is per Jack B. Copeland, footnote 53 on page 409 of *The Essential Turing*. Copeland notes errors by others in stating the date of this report. He notes that the phrase "Manchester machine (as actually working 8/7/48)" appears in both the finished NPL [National Physical Laboratory] report and in the draft typescript. Thus the draft report was completed sometime after July 8. The 1948 conference of the International Congress of Genetics was held July 7 - 14, 1948 (Bengtsson and Tunlid, p. 709).

by natural selection and the science of genetics had been fulfilled, and the potential that being able to examine and manipulate genetic material might hold in light of the synthesis was just being sketched out. At that 1948 conference, Muller, the president of the 1948 conference, spoke (among many other things) about the ability to extract and study in vitro genetic material from bacteria. Concerning genes themselves, Muller suggested that all genes probably "have the same essential composition, inasmuch as they require a combination of some kind of nucleic acid or nucleic acid prototype with some kind of protein or protein prototype." [5] Muller's speculation about the composition of genes themselves was framed in terms of combinations: Combination, that was the key in 1948.

Once Turing's comment about "genetical or evolutionary" search here is placed in its historical context, there does not seem to be any reason not to take at face value this comment by one of the most outstanding cryptographers of the time, to the effect that he regards the task of finding the genetic code associated with features of an organism that increase its "survival value" to be a type of search [among different possible combinations]. I think his point in identifying genetical or evolutionary search as a distinct kind of search is that, in light of how well the evolutionary synthesis had turned out by 1948 (from the standpoint of those judging it in 1948), genetical or evolutionary science seemed finally to conclusively support the idea that natural selection, too, can be seen as a kind of search, a search for the right combination of genes. It seemed to portend great advances, too. It seems to me to be something of a precursor of a way of viewing natural selection now known as the "selfish gene" view. Writing over 65 years later, such a "selfish gene" view is now no longer novel, nor is it now unquestioned. The current status of this view in biological science does not matter to our discussion in this paper, though, which is understanding Turing's points about machine intelligence.

The point about genetic search relevant to us here, and to the point of the essay on intelligent machinery in which it occurred, is that *survival value* and *intellectual power are not the same thing*. That is why the genetical or evolutionary search, item (b) on the list, is a *different, distinct, variety of search* than either intellectual search (item (a) on the list) or cultural search (item (c) on the list). In intellectual search (item (a)), it is brains that carry out the search; in genetical or evolutionary search (item (b)) it is natural selection, or nature, carrying out the search (for a combination of genes.") [6] And, in cultural search (item (c) on the list) it is "the

[5] quoted in Bengtsson and Tunlid (2010; p. 712).
[6] I think that Turing is very clear about the fact that he is drawing an analogy when he lays out the

human community as a whole" that carries out the search, rather than either a human (brain) or nature. Thus, these varieties of search can be distinguished by the *agent to whom Turing attributes the action of carrying out the search,* when the process is regarded as a search of that kind.

Turing notes that for a human to develop "intellectual power", he or she must "be immersed in an environment of other [humans], whose techniques he or she absorbs during the first twenty years of his [or her] life. He may then do a little research of his own and make a very few discoveries which are passed on to other [humans]." It is in this passage that he makes the distinction between the agency of the individual being trained and the agency of a human community of which the individual later becomes a part, writing that "from this point of view the search for new techniques must be regarded as carried out by the human community as a whole, rather than by individuals." (Turing, 2004a, p. 440) The importance of being immersed in an environment of other human is emphasized at some other places in the 1948 "Intelligent Machinery" report for NPL, too. When pointing out that "although a [human] when concentrating may behave like a machine without interference, his behavior when concentrating is largely determined by the way he has been conditioned by previous interference", he notes not only that this interference includes "contact with human beings for twenty years or more", but also that a human is "in frequent communication with other [humans]." (Turing 2004a, p. 430)

4. Being in a Search Party

The capability to carry out intellectual search (for combinations with particular properties), then, is ascribed to a brain, and the development of an individual into a human with intellectual power is dependent not just on having a good human teacher, but on being surrounded by and interacting with lots and lots of other humans, over the course of many years, and continuing to do so throughout one's adult life. In fact, the kind of interaction that continues to be frequent is "*communication* with other [humans]." (italics added) I believe that it is this "frequent *communication*" aspect that is crucial to being able to make sense of Turing's ascription of the capability to make new discoveries to a human community, but not to an individual. He grants

---

analogy between natural selection and the analogue he proposes be tried (in his later 1950 paper *Computing machinery and intelligence)*. In the analogy in *that* paper, which he explicitly lays out there, he says that "the judgment of the experimenter" is analogous to Natural selection. In that analogy, "how well [the machine] learns" *would* be analogous to how well an animal form does in terms of survival.

that individuals do conduct research and make discoveries, but he does not consider the "search for new techniques" to be something that is carried out by an individual human. Were the human interference with the machine limited to the educational phase of the machine (which occurs prior to the individual's research-and-discovery phase), it would not make a lot of sense to distinguish between the aspects of the individual's researches and discoveries that were to be ascribed to him as an individual, and the aspects that could only be seen as a contribution to a joint action carried out by "the human community as a whole." When, however, we consider the model of human researcher on which it is not just during an initial educational period that the researcher is in contact with others, but that he or she is "in frequent communication with other [humans]", it makes sense that, although some of the individual's research and discovery could be regarded as carried out by the individual, that he or she is also participating in a joint action of discovery of which the individual is not the agent carrying out the joint action of discovery.

Turing predicted that "Further research into intelligence of machinery will probably be very greatly concerned with 'searches' of [the intellectual] kind' " but the portion of his paper about "the cortex" should be seen as only part of the larger inquiry whose lines the paper sketches out. I think we can see where he means the activity of intellectual search fits into this sketch with respect to his comments on "the cortex" when he writes about the role of "definite nerve circuits required for quite definite purposes": "all the reflexes proper (not 'conditioned') are due to the activities of these definite structures in the brain. Likewise the apparatus for the more elementary analysis of shapes and sounds probably comes into this category. But the more intellectual activities of the brain are too varied to be managed on this basis." (Turing 2004a, p. 432). I want to focus on how the issues brought out above bear on this observation. That is, how the significance that Turing placed on the human's advantage over a computer in being immersed in a community of humans, of having had a lot of contact with other humans over the course of many years, and, finally, of being in frequent communication with others bears on the observation that the more intellectual activities of the brain are too varied to be managed by definite structures in the brain. That means they are due to interference, or contact, with other things; Turing's view seems to be that to develop true intellectual power, that interference involves frequent communication with other humans. The question for a machine is: How much, and what kind of contact does a machine need in order to carry out various kinds of search well?

5. Human Help & The Human Brake

That Turing cites the *value* of a machine's contact with (or interference from) humans, especially frequent communication with others, in order to gain "intellectual power" in this 1948 paper in which he is exploring how to educate a machine on analogy with how humans are educated so as to come to have the intellectual capabilities that they do is noteworthy. Interaction with humans carried another association in other studies of the capabilities of computing machinery; in a 1945 investigation into building a computer, Turing had been concerned about the *negative* effects of involving humans in a computing machine's operations. Marie Hicks notes that Turing coined the term "human brake" and explains: "The human brake slowed computing processes by delegating actions to human operators that should ideally reside within the capability of a machine and its programs." (Hicks 2008)

In terms of speed, we don't want our machines to be bogged down by having to interact with humans; as Turing wrote in "Proposed electronic calculator" in 1945: "Once the human brake is removed the increase in speed is enormous." (Turing 2005, p. 371) At that time, he stressed the crucial importance of a large memory for information storage in order for the machine to be able to carry out complicated things; as for human interaction: "How can one expect a machine to do all this multitudinous variety of things? The answer is that we should consider the machine as doing something quite simple, namely carrying out orders given to it in a standard form which it is able to understand." (p. 372)

When it came time to consider the possibility of intelligence, there was the theoretical issue that exhibiting intelligence required permitting the machine to be fallible (since "if a machine is expected to be infallible, it cannot also be intelligent." (Turing 2004c, p. 402)) After dealing with that theoretical point, though, when it came to achieving intellectual development of a machine, interaction with humans was almost always cited as necessary for achieving parity with human displays of intelligence, and he proposed concrete, practical examples of rudimentary interactions by which a machine could improve: playing games with humans. Copeland refers to the 1948 "Intelligent Machinery" report for the NPL as "The First Manifesto of Artificial Intelligence" [7] and we have seen how salient the points about interaction with humans are in that paper. In his 1947 "Lecture on the Automatic Computing Engine", too, he had stressed the same point we see in the 1948 "Intelligent Machinery" report about the crucial nature of *interaction of the machine with humans*. He makes it, explicitly, in the closing. Arguing from a comparison with how humans come to gain the intelligence they do, he argued from the

[7] In his *The Essential Turing*, p. 409.

observation that "A human mathematician has always undergone an extensive training" to "the machine must be allowed to have contact with human beings in order that it may adapt itself to their standards." In fact, he said there that "One must . . . not expect a machine to do a very great deal of building up of instruction tables on its own."
(Turing 2004c, p. 403)

The suggestion that human cognitive abilities can not develop in isolation was not new with Turing. Lev Vygotsky's theories [8] about the role of social interaction in the development of an individual human's cognitive abilities were published, although it is hard to tell how well known they were at that time. Turing's views that the development of "intellectual power" depends upon contact with other human beings is in sympathy with Vygotsky's views, but Turing's suggestion that it was the best way to approach machine intelligence, as opposed to using principles and rules, went beyond Vygotsky's views, which were limited to humans. [9]

Turing was certainly right in predicting that much research into machine intelligence would be concerned with searches of the intellectual kind -- but how has his view about the value of interactions with humans fared? Let us examine two recent projects: IBM's Question-answering machine, Watson, and Carnegie-Mellon's language learning machine, NELL.

6. NELL - "Reading" the Web

According to Tom Mitchell, "The idea of NELL ("Never Ending Language Learning") is to capture a style more like the on-going learning of humans." By "style", though, he meant that humans are continuously, rather than intermittently, learning. So, NELL operates continuously, 24/7, "reading" the Web; the overall goal is to acquire the ability to extract structured information from unstructured web pages. However, unlike Turing's idea of having a human teacher of the child-machine, NELL was left to operate autonomously for the first six months, after being given a list of several hundred categories and relations, and about "10 to 15 seed examples" of each. [10] What NELL is supposed to do is to form beliefs, and to determine how strongly to believe (what confidence level to assign to) them.

---

[8] Vygotsky's *Thought and Language* was published in 1934, and he had traveled to London before that.

[9] In a suggestive paper "Social Situatedness of Natural and Artificial Intelligence: Vygotsky and Beyond", Jessica Lindblom and Tom Ziemke discuss how Vygotsky's views might apply to designing human-robot interaction. (Lindblom and Ziemke, 2003)

[10] "NELL: The Computer that Learns", downloaded 25 January 2014. http://www.cmu.edu Web.

The approach taken is called "Macro-Reading"; the motivation for the idea is to tackle a problem that is much more tractable than the problem of natural language processing. Mitchell explains that "the problem of automatically populating large databases from the internet can be formulated so that it is much easier to solve than the problem of full natural language understanding." Macro-reading (in contrast to micro-reading) is defined "as a task where the input is a large text collection (e.g., the web) and the desired output is a large collection of facts expressed by the text collection, without requiring that every fact be extracted." This is supposed to be simpler because the redundancy of the facts on the web allows the reader to ignore complex statements and yet still pick up on the main facts available on the web. Other features of NELL that make the task of "reading" the web more tractable are that NELL aims for something rather streamlined rather than embracing all the bricolage surrounding a category or relation on the massive web. NELL doesn't try to understand the text it encounters. NELL's task is to "populate" a portion of the semantic web by "populating", i.e., filling in, an ontology that is given as input. Finally, NELL does figure out some ways to extract information from the text based on patterns; Mitchell's example is " 'mayor of X' often implies X is a city." [11]

NELL comes up with beliefs and these are listed on the web. Help from humans in scoring confidence levels for these "beliefs" is crowd-sourced; for each "belief" listed, a visitor to NELL's website can click a "thumbs up" or "thumbs down" to help NELL correct, or fine tune, the confidence level associated with that "belief." On August 27th, NELL learned that "belfast_zoo is a zoo in the city belfast" and that had a 100 percent confidence level. But "family_animals is a mammal" did not rise so high, nor did "savior is the parent of trusting jesus." The visitor cannot interact with NELL to edit, revise, or ask questions about, these beliefs; interaction is limited to a "correct" or "incorrect" response. I think it is clear that what NELL is going to gain from interacting with humans is only going to take it so far. [12]

With respect to our question, what I would say about NELL is that its creators aimed for a system that "learned" (in the sense of improving its performance in extracting true statements from web pages) autonomously, and because they were interested in making it as autonomous as possible, they were not thinking in terms of how best to incorporate interactions with humans into its learning process. As it turned out, human help was later added, both in the form of additional constraints on the relations NELL worked with, and also (via crowdsourcing) with confidence

[11] In Mitchell, Betteridge, Carlson, Hruschka, and Wang, 2009.

[12] Checking NELL's "Recently-Learned Facts" on January 26, 2014, I find that "h_ross perot is a politician who holds the office of president" is held with 99.2 % confidence!

levels of the beliefs it generated. So, NELL does make some (slight) use of ongoing interaction with humans. However, it is not discriminative with respect to which humans' input it uses; this may account for its acquisitions of some strange "beliefs" that are scored with high confidence ratings from time to time.

In a deeper sense, though, I would say that NELL did not really take advantage of opportunities to learn from humans *via* the web. This failure is inherent in the "macro-reading" approach, since its modest aim is to extract some simple relations that appear frequently on the web (and, later, to mine its set of these relations for rules it can use to conclude more simple relations (Lao, Mitchell, and Cohen, 2011) . In contrast, I find that IBM's Watson, in part because its task was different, was perfectly situated to learn from humans via the web, as well as via more direct kinds of "interference" (i.e., programming). As I'll explain, understanding how Watson made use of human contributions via the web makes Watson in some ways more impressive as a learner, but makes Watson's performance on the game show *Jeopardy!* somewhat less impressive (than it appears if one is in a state of ignorance about how Watson achieved its success.)

7. IBM's Question-Answering Champion, Watson

IBM's Watson is a question-answering system that was specially designed, constructed, and tested to compete in the television game show *Jeopardy!* It is very sophisticated. It won against the reigning human champions of the game. Watson uses a software architecture called DeepQA. It employs a variety of languages and methodologies, and it uses parallel computer architecture to provide the necessary speed. It performs an analysis of the question to determine what is being asked, then generates many candidate answers. Each candidate answer is checked against many sources for consistency and other kinds of checks, and, eventually, confidence levels are computed and candidate answers are ranked. Watson then uses these ranked candidate answers from DeepQA, along with other information it generates, to play the game (e.g., to choose clues, answer questions, decide how much to wager). The confidence levels are important to playing the game well, since wrong answers incur substantial penalty.

As answering the questions to *Jeopardy!* requires knowledge of specific details about a wide variety of publicly available information, Watson, too, was taught to use the unstructured information on the web. However, in complete contrast to NELL, Watson does not avoid natural language processing, nor look for simpler tasks such as "macro-reading."

The approach to making use of the vast amount of information on the web that was taken for building Watson was to be discriminating about the quality of the sources acquired, and to be practical about the amount of text involved in a source.

> "A key element in high-performing question-answering (QA) systems is access to quality textual resources from which answers to questions can be hypothesized and evaluated. [. . .]
>
> We developed three procedures to obtain high-quality textual resources, i.e., source acquisition, source transformation, and source expansion. When developing a new QA system or adapting an existing system to a new domain, relevant sources need to be identified to cover the scope of the task. We refer to this process as *source acquisition*, which is an iterative development process of acquiring new collections of documents to cover salient topics deemed to be gaps in existing resources. The acquired sources are examined with respect to characteristics of system components, as well as to the nature of the questions and the answers in the new domain to ensure that they are represented in the most effective manner. Some acquired sources go through a process that we call *source transformation*, in which information is extracted from the sources, either as a whole or in part, and is represented in a form that the system can most easily use. Finally, whereas source acquisition helps ensure that the system has coverage in salient topics of the domain, *source expansion* attempts to increase the coverage of each known topic by adding new information, as well as lexical and syntactic variations of existing information. We believe that the methodology that we developed for source acquisition, transformation, and expansion is crucial for providing Watson with the necessary resources to achieve high QA performance." (Chu-Carroll 2012a.; pgs. 4:1 - 4:2)

Commentators on Watson's performance often make much of the fact that the *Jeopardy*! game requires knowledge about a very wide range of topics in a very wide range of disciplines: history, literature, sports, popular culture, science, mathematics, geography, politics, and so on. It wasn't just that Watson was able to outperform humans in a question-answering task, but that Watson was able to outperform humans in a question-answering task that was (seemingly) unrestricted with respect to topic, that was so remarkable.

However, Watson's designers share something discovered in the course of source acquisition that changes one's perspective on this feat a bit: "Fortunately, although the questions cover

many domains, for the most part, they cover popular topics within those domains and represent information of interest to the general audience." (Chu-Carroll 2012a; p. 4:2) Investigating further, they found something astounding: "on a randomly selected set of 3500 questions, all but 4.53% of the answers were Wikipedia titles." [13] What to make of this? It didn't really matter what this said about the game show or about Wikipedia, it was a useful conclusion that made the task of building Watson much easier. Wikipedia article titles "can serve as an excellent resource for candidate-generation." (Chu-Carroll 2012b 6:8) Wikipedia articles provided useful metadata, too. Metadata (such as what text in the article links elsewhere, and what it links to) helped determine what was salient to a topic and what was not. Watson's designers found "We observed that plausible candidates typically satisfy two criteria. First, they represent salient concepts in the passage. Second, the candidates have Wikipedia articles about them." (Chu-Carroll 2012b 6:8)

By itself, just using the corpus of Wikipedia to generate candidate answers wasn't enough. But Wikipedia articles could help with more than just candidate generation. They may count as unstructured text, but there is a lot more than simple text in them. Chu-Carroll 2012a explains how the team was able to make use of the "title-oriented" character of Wikipedia entries. Other articles (Chu-Carroll 2012c, p. 12:3) show how the links in Wikipedia articles provide help in identifying implicit relationships, to help Watson build up its store of knowledge about concepts that were closely related, which can come in handy when figuring out what a clue is asking for.

This leg-up in generating candidate answers had numerous benefits, but there is one that deserves special mention: a better way to handle assigning a type to the answer that is sought by a particular question. The usual order of things in state of the art question-answering programs of the time was to first figure out what "answer type" the answer to the question would have:

> "Many open-domain question-answering (QA) systems adopt a *type-and-generate* approach by analyzing incoming questions for the expected *answer type*, mapping it into a fixed set of known types, and restricting candidate answers retrieved from the corpus to those that match this answer type (using type-specific recognizers to identify the candidates.)" (Murdock et al 2012, p. 7:1)

[13] Of the 4.53 % of answers that were not Wikipedia titles, "some are multiple answer questions (e.g., "Indiana, Wisconsin, and Ohio" and "heat and electricity"), some are synthesized answers to puzzle questions (e.g. "TGIF Murray Abrahams" and "level devil") and a small number are verb phrases (e.g., "get your dog to heel") (Chu-Carroll 2012a; 4:2)

The Watson team switched the order of things, using instead a *generate-and-type* framework. Types matter, but checking that the type of the candidate answer fits what the question asked is done much later, as part of the task of assigning confidence scores to each answer and ranking them according to confidence level. This permitted many more types, and it allowed for much more flexibility about types. The designers explain that early on, in analyzing the domain of questions from the TV quiz show *Jeopardy!*, they found the type-and-generate approach "to be problematic." They found, in fact, that they could not reliably predict "what types the questions will ask about and what their instances are." The number of type words that could occur in a *Jeopardy!* style question was, practically speaking, at least, unbounded. A survey of 20,000 questions yielded "roughly 5,000 different type words." (Murdock 2012, p. 7:3) In explanation of this striking situation, they give some examples:

> Human language is remarkably rich when it comes to assigning types; nearly any word can be used as a type, particularly in some questions.
> - Invented in the 1500s to speed up the game, this *maneuver* involves two pieces of the same color. (Answer: "Castling")
> - The first known airmail service took place in Paris in 1870 by this *conveyance*. (Answer: 'hot-air balloon")
> - In 2003, this Oriole *first sacker* was elected to the Baseball Hall of Fame. (Answer: "Eddie Murray")

An answer type, they concluded, ought to be treated as a property of the question and answer combined. The approach they developed differed from the usual QA systems approach of looking for candidate answers of the right type; instead, it was to "find candidates (in some way) and judge whether each one is of the right type by examining it in context with the answer type from the question. " (Murdock 2012, p. 7:3)

I suspect this approach, which was in part possible because of the leg up on generating candidate answers provided by Wikipedia entries, was important to Watson's stellar showing in a game that required not just a basic competence in natural language processing, but lots of savvy about that language.

Isn't Wikipedia really just another unstructured text source, though? Yes, it is an unstructured text source, but I also think that Wikipedia is special. Wikipedia entries are written to be read by someone with a question, and Wikipedia text is constructed by humans who enjoyed writing and

sharing the information -- and were joined by others who added to and revised it. The article titles are constructed for the web visitor looking for an answer, and the information, though unstructured text, is still organized in predictable ways (i.e., the links it contains will not be arbitrary words in the entry, but will be especially relevant to the title topic.) It is generally targeted towards the general, i.e., non-expert reader. And, not only is it comprehensive in time backwards, but it is constantly being updated, so that it can always be up to date -- on everything it covers. In some ways, having access to Wikipedia articles on a continuous basis resembles being in frequent communication with humans.

I think it speaks well of the Watson team that their methodology quickly uncovered the relationship between Wikipedia entries and *Jeopardy!* answers, and that their approach makes such rich use of the various aspects of the Wikipedia entries. Were the Watson technology to be adapted to another use, i.e., as a collaborator in some other professional field, the kinds of documents that are important might be, in fact probably will be, different ones. More generally, Watson receives training from humans, who tell it which sources to prefer over others. One of the many things the humans told Watson for its *Jeopardy!* contestant "job" was to use Wikipedia -- and how to use it. Watson still generates candidate answers and metadata on its own, still performs lots of checks and evaluations on candidate answers, and still comes up with confidence rankings on its own, but a good part of the reason it does all these things so well is the help from humans it receives in terms of getting quality sources and being told how to make good use of them. Thus it seems to me that the case of Watson bears Turing out; the machines that are closest to developing intellectual power on parity with humans are those that are not only trained by humans, but that are in frequent communication with them.

8. Closing Thoughts

Both NELL and Watson have access to the same unstructured text on the web. If Watson (when connected to the internet) can be seen as having interaction with humans that approximates communication with them via the web, why not NELL, too? Well, communication takes two. The way NELL learns from Wikipedia pages is very different from the way Watson does. NELL sees the text on the web, not as text to be processed in order to understand what its author meant to communicate, but as a landscape studded with a plethora of simple facts caught in a complicated matrix of other text. The landscape is to be mined by ignoring all of the complication in that complicated matrix, and paying only scant attention to the context surrounding the simple fact. Watson, on the other hand, sees the text on the web as much more variegated; it uses only quality text sources and (via an iterative process) ensures it has enough

of the right kinds of them to address its anticipated needs. The approach Watson's designers took appreciates that the text is authored by someone who meant to communicate something, and so designed Watson to analyze the text accordingly, not only to understand what its author meant to communicate, but also to get the most information out of the text that it can. That means, loosely speaking, that Watson attempts to understand the text as natural language, of course, but also that Watson uses other features about the text (e.g., which string of text is designated as the title, which strings of text are anchors for weblinks) that yield information on how things and concepts may be related, and that Watson uses metadata drawn from links and statistical data.

We may suffer a bit of disillusionment upon finding out just how important the work of all those human Wikipedia contributors was in Watson's *Jeopardy!* win, but, in a sense, Watson is impressive for knowing how to make such good use of what the Wikipedia contributors provided. On this view of things, it is Watson whose approach might someday lead to real intellectual power, and for reasons akin to Turing's emphasis on the importance of human contact: Watson knew how to listen to and learn from what humans were telling it.

References (to Sterrett chapter)

Bengtsson, Bengt O., and Anna Tunlid. 2010. The 1948 International Congress of Genetics in Sweden: People and Politics. *Genetics* 185: 709–715. Also available at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2907196.

Chu-Carroll 2012a. Chu-Carroll, J., J. Fan, J., N. Schlaefer, and W. Zadrozny. 2012a. Textual resource acquisition and engineering. *IBM Journal of Research and Development*, Vol. 56, No. 3/4, Paper 4 May/July 2012.

Chu-Carroll 2012b. Chu-Carroll, J. et al. Finding needles in the haystack: Search and candidate generation. *IBM Journal of Research and Development*, Vol. 56, No. 3/4, Paper 6 May/July 2012.

Chu-Carroll 2012c. Chu-Carroll, J., E. W. Brown, A. Lally, and J. W. Murdock. Identifying implicit relationships. *IBM Journal of Research and Development*, Vol. 56, No. 3/4, Paper 12 May/July 2012.

Hicks, Marie. 2008. Repurposing Turing's "Human Brake." *Annals of the History of Computing*. Vol. 30, issue 4.

Koza, J. R., F. H. Bennett III, D. Andre, and M. A. Keane. 1999. Genetic Programming: Turing's Third Way to Achieve Machine Intelligence. EUROGEN workshop in Jyvdskyld, Finland on May 30 - June 3, 1999.
Available at http://www.genetic-programming.com/jkpdf/eurogen1999turing.pdf

Lao, Ni, Tom Mitchell, and William W. Cohen. 2011. "Random Walk Inference and Learning in a Large Scale Knowledge Base." in EMNLP '11 Proceedings of the Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA: Association for Computational Linguistics. p. 529 - 539.

Lindblom, Jessica and Tom Ziemke. 2003. Social Situatedness of Natural and Artificial Intelligence: Vygotsky and Beyond. *Adaptive Behavior*, Vol. ii (2) 79 - 96.

Mitchell, Tom M., Justin Betteridge, Andrew Carlson, Estevam Hruschka, and Richard Wang. 2009. "Populating the Semantic Web by Macro-Reading Internet Text." In *Proceedings of the International Semantic Web Conference* (ISWC), 2009.

Murdock, J. W. et al. 2012 Typing candidate answers using type coercion. *IBM Journal of Research and Development*, Vol. 56, No. 3/4, Paper 7 May/July 2012.

NELL: The computer that learns. Accessed 25 January 2014.
http://www.cmu.edu/homepage/computing/2010/fall/nell-computer-that-learns.shtml

Sterrett, Susan G. 2012. "Bringing up Turing's 'Child-Machine'." in Cooper, S. Barry, Anuj Dawar, and Benedikt Lower (Eds.) *How the World Computes: Turing Centenary Conference and 8th Conference on Computability in Europe, CiE 2012, Cambridge, UK, June 18-23, 2012. Proceedings*. Lecture Notes in Computer Science Volume 7318 (pp 703-713)

Turing, Alan M. 2005. Proposed electronic calculator (1945) In *Alan Turing's Automatic Computing Engine: The Master Codebreaker's Struggle to Build the Modern Computer*. Edited by B. Jack Copeland. Oxford and New York: Oxford University Press.

Turing, A. M. 2004a. "Intelligent Machinery (1948)" in Copeland, J. B. (Ed.), *The Essential Turing* (pp 418 - 440). New York: Oxford. 2004.

Turing, A. M. 2004b. "Computing Machinery and Intelligence." in Copeland, J. B. (Ed.), *The Essential Turing* (pp 418 - 440). New York: Oxford. 2004.

Turing, A. M. 2004c. "Lecture on the Automatic Computing Engine" in Copeland, J. B. (Ed.), *The Essential Turing* (pp - 403). New York: Oxford. 2004.

Vygotsky, L. S. 1986. *Thought and Language*. Cambridge, MA: MIT Press. (Original work published 1934)