Beyond Positivism in Computer Science

Markus Luczak-Rösch
Freie Universität Berlin
Inst. für Informatik
Netzbasierte Informationssysteme
Königin-Luise-Str. 24/26
D-14195 Berlin
luczak@inf.fu-berlin.de

WebSci'13, May 1 – May 5, 2013, Paris, France.

ACM 978-1-4503-1889-1.....$10.00

Author Keywords
Computer Science; Web Science; Philosophy; Positivism; Modeling; Engineering;

ACM Classification Keywords
H.1.1 Systems and Information Theory, H.1.2 User/Machine Systems

A discourse on modeling and specialization

It is accredited to Marie Curie that she said: "*In science, we must be interested in things, not in persons.*" This seems to be a viewpoint which is very common in Computer Science where a core competence is to model solutions in a rational way as the following two examples illustrate: Algorithms are models that describe the data input, flow, manipulation, and output necessary to compute a terminal deterministic solution in finite time on a computer; Web ontologies are models that represent a consensual and consistent state of domain knowledge conforming to Web standards. *Computer Science teaches humans to model things by standards and technologies which are valid in a digital computer systems*. But how are these "things" different from those in physics meant by Marie Curie?

Comparing the modeling of a pendulum with the facebook "like" button is a very intriguing example of

how modeling in physics is different from today's challenges in Computer Science. While the former refers to quantitatively observable phenomena in nature - even though the phenomena may be microcosmic - the latter is a model that only allows probabilistic predictions about the quality and sustainability of the relation between humans and/or information. Hence, Computer Science is naturally at the interface of technology and society today.

Computer Science in times of the Web - what is it about?

"*Specialization may be a great temptation for the scientist. For the philosopher it is the mortal sin.*", Karl Popper, *Conjectures and Refutations*, p. 184

Positivism in Computer Science - examples and controversies

Positivism is a concept on the meta level of epistemology and scientific theory that is subject to a long lasting discourse, especially in the humanities. Originating from the rather strict viewpoint of philosophers like Compte and Mill that metaphysics do not contribute knowledge on phenomena of general interest, *logical positivism* has important influence on a notion of an - at least temporal - truth in all sciences, assuming that while "scientific laws are often discovered through a process of intuition, this does not mean that they can be intuitively validated" [1]. Furthermore the theory of truth in logical positivism is composed by answering the following question: "With regard to any proposition p, what are the conditions in which p (is true) and what are the conditions in which not-p?" [1] Hence, logical positivism is about "a priori" (e.g. mathematical axioms) and "empirical" propositions that both need to be rigorously and publicly validated.

It is the nature of Computer Science to provide solutions which are valid conforming to the internal conditions of computer systems, which is consequently an admissible positivistic approach. My aforementioned example is the modeling and implementation of psychologically and sociologically complex concepts such as friendship or the expression of esteem in social networks. A software developer creates a solution that exploits the conditions of a computer system and a binary relation between two entities represented in it (be it people and/or information) to express a truth about friendship or esteem. I reason that Computer Science tends to be too positivistic when maintaining this technology-centric direction instead of extending the view to validate the propositions outside of the computer system as well. The following examples are intended to stress this assumption.

Modeling human discourse

IBIS is an approach to model human discourse in information systems [7]. One implementation of this model is the DILIGENT argumentation ontology [9] for the community-driven development of consensual Web ontologies. The argumentation ontology provides the concepts and properties to express atomic parts of a community discourse as facts on the Web. This allows to infer the community consensus or to detect inconsistent argumentations automatically. However, this model is not feasible for the detection of the intricacies of human communication, namely communication strategies or affective behavior which may evolve due to the different nature of different people.

"Mirror with privileged knowledge" or "prediction equals explanation" [4]

Peter Norvig wrote an essay published as a Web page [8] where he discusses a speech of Noam Chomsky which was referenced in an article of the MIT Technology Review [3]. In this article Chomsky is cited

criticizing a trend in Computer Science which he observed that heavily relies on statistical models and "*approximating unanalyzed data*" which, as he argues, fails to result in real understanding of phenomena. Norvig counters Chomsky's rather extreme rejection by consulting the successes of statistical models, most notably applied in search engines, speech recognition, machine translation, and question answering.

Even though they represent antagonistic views within the area of artificial intelligence, discourse on research paradigms (e.g. as in [4]) allows the assumption that both positions are positivistic. One claims for modeling the world objectively a priori - Chomsky's notable work on the principled modeling of linguistic structures - and the other assumes that prediction based on statistics is explanation - Norvig's pleading to accept the success of the probabilistic approaches that purely rely on large and ever growing amounts of raw digital data.

What algorithms do to the financial market
It is a matter of fact that high frequency trading constitutes a large share of stock exchange revenue. Algorithms facilitate trades at nanosecond scales and new business models penetrate the financial market which would not be possible if real human interaction and decision making would be required.

This is surely an interesting area for research and development on optimization problems that face global-scale systems and huge amounts of data. It is also no question whether the next, faster, and better (whatever this means) high frequency trading algorithm will be developed. Market rules and infrastructures provide the perfect setup for applying the modeling skills of Computer Science to develop a solution that enables this. However, all this is directly concerned with questions of ethics and of how we understand the problem of stock exchange in its entire shape for the function and evolution of society.

SEO and Twitter bombs - en vogue pollution of the Web
It is a prominent political and societal discourse in Germany about transparency in democracies as well as the necessary openness of public data and the Web. While this discourse promotes important aspects of an enlightened and critical society that acts for a global sustainability and ecology, it also shows how well meant initiatives deteriorate when they begin to act in a way that they have criticized before.

There are parties in Germany that blame other established parties for their conservative organization and political work which they regard as lobbying and information hiding to gain advantages in the next election campaigns. This is a discourse which deserves to be greatly appreciated and which already resulted in changes in the understanding of all political parties and their position towards the citizens. However, one can also get the impression that a party which declares that it is the only one that does not bulkhead itself and which consists of humans, is as an antidemocratic discredit of others [6]. Irrespective of the fact that one can observe a convergence of what is generally treated as online and offline, some parties and their supporters are characterized by their higher online engagement and use of social media, which they exploit to promote topics of interest or to perform discourse with and about the other parties on the Web. That means that those "online parties" exploit the possibilities of the system they obviously know better than other parties, in order to gain an advantage in the next election

campaigns. Finally this is exactly the same behavior that other parties perform offline.

This example from politics shows that it remains necessary to critically question the trends and recommendations generated from massively amounts of raw Web data. Search engine optimization was the beginning of an en vogue corruption of the reputation metrics of the Web and modern forms of crowd-based information sharing continue this. Both phenomena often source from affective behavior of individuals or organizations that target to gain an advantage - be it economically or related to the individual reputation or position.

Thoughts on problems, non-determinism and infinity in Computer Science

I have drawn a picture of Computer Science as an extremely positivistic discipline. In the following I will elaborate on a paradigm shift in Computer Science, focused on socio-technical interferences.

Model the problem instead of the solution

While modeling a solution implicitly contains a model of the problem it does not go far enough to capture the side-effects of technology and the emergent use of it as outlined in this paper. A problem model should not propose any specific standards or algorithms which constitutes a solution. It should rather describe the system where the problem needs to be solved; the intended result; the sources of information as well as computational resources including the trustworthiness and quality of each; the interacting people in the system and their individual goals which interfere with the intended result. This modeling approach is loosely coupled with any concrete technology but it allows to understand socio-technical side-effects of the system where the problem needs to be solved.

Tolerate vagueness, non-determinism, and affective behavior

Already today people rely on services that present probabilistic results. The longer meteorologists analyze the data the better may be the prediction. But, in the end they will hardly ever be able to take the effects of an unforeseen earthquake or other immediate natural phenomena into account. In the context of the Web search engine providers could tell the users that search results are affected by Web publishers' search engine optimization behavior or the bias crowd-sourced content may contain. This could decrease the degree of disappointment about non-sufficient search results because of lower expectations and promote a critical treatment of content with intransparent provenance. For sure, this would thwart various current business models of search engine providers which are in a way intertwined with the corruption of the service for the real end user who has the problem of productive online research.

A well-modeled problem documents its socio-technical boundaries so that any solution to this problem can exploit all available or only selected social, computational, and informational resources of the system it is implemented on, assign trustworthiness and quality scores to partial or end results (if the latter is computable in acceptable runtime), and communicate those scores to the user to allow her to make the final decision about taking a result into account or not. This is a vision of solutions that tolerate vagueness, non-determinism, and affective behavior. The user gets insight about any uncertainties.

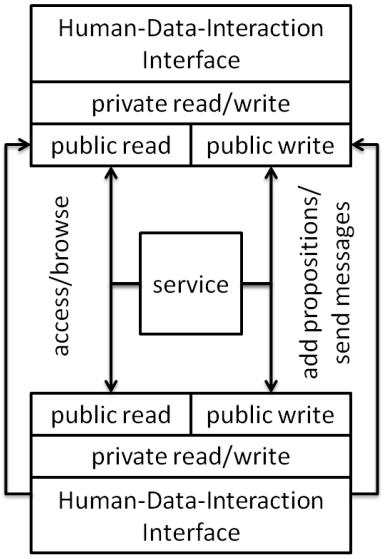

Figure 1: Proposed architecture and interaction paradigm for a new generation of Web servers, providing the human user with a personal data space as well as browsing and messaging capability through a unified Human-Data-Interface.

Accept infinity

Turing's universal machine was designed with a memory of infinite size [10,11]. Based on that, current algorithmic thinking is majorly driven by the theory of finite algorithms and complexity classes that refer to the runtime until a terminal result. This constitutes the common notion of computability. Is there a characteristic subclass of problems of highest complexity which may be solved in an acceptable way from a non-technical perspective, meaning that humans can make a final decision on a partial result that fulfills certain criteria? If this is the case one can imagine a general theory of non-terminating algorithms that require things not feasible or existent to date. That would allow the vision to deliver probabilistic or partial results until some point of time in the future where immediately the final result can be computed because the respective service, resource, data, or person is available.

A technical proposal

The missing piece in current architectures of network-based systems is the human user as an integral component providing tolerance for vagueness and non-determinism by interaction.

Clients as a server and interaction as integral part of network-based systems

In the context of the Web for example, most likely due to the historical importance of Web browsers, the human user is still an information or service consumer browsing Web resources by a client application. Actually each client should be a server at the same time, letting the user decide which interactions with resources are contributed back to the information space. Beside browsing by typing identifiers, following links or requesting search engines, it should be possible to publish a problem or information need as structured data. Other entities - be it human users or services - can write references to solution propositions into a dedicated part of a server. The user can be asked to validate such propositions, decide about their approval, and publish the approval back to the public information space.

The Web architecture already offers the protocols and standards to model and share interactions as messages or problem statements, as well as to run non-terminating algorithms: The Linked Data principles[1] are a means for flexible and scalable data-centric systems; Web servers are machines set up for a literally infinite runtime; the read-write Web[2] idea as well as the WebBox approach [5] and HTML 5 are foundations for personal data spaces and real interoperation of Web resources in an application and protocol independent fashion as envisioned in [2] (Figure 1 depicts a high-level draft of this setup).

A proposal for Computer Science education

A very recent campaign[3] of public figures claims for teaching school children how to code as early and keen as mathematics is taught as one of the most fundamental subjects of our times. Computer Science never was one of the fundamental natural sciences but it was concerned with a fundamental shift for mankind towards the digital. It was, is, and will be responsible

[1] http://www.w3.org/DesignIssues/LinkedData.html as visited on 12-04-2013

[2] http://www.w3.org/DesignIssues/ReadWriteLinkedData.html, as visited on 12-04-2013

[3] http://www.code.org/, as visited on 12-04-2013

for automation and optimization in various areas. Hence, the campaign is precious without exception.

Teach how to code but also to criticize technology
Since the digitalization enfolds a massive societal impact, most prominently in shape of the Web, lecturers and researchers in Computer Science should not only focus on modeling and implementing the technologically feasible but also stimulate discourse about the effects the progress may have. What's technically feasible does not need to be good or may emerge worse than expected.

Conclusion

Marie Curie's discovery of radium had massive impact on our world but today we also know about the danger of radioactivity. Computer Science is not going to discover the next phenomena of nature that may be unmasked to be physically disruptive. It aims at solving problems of mankind in the digital space and it may be, and to some extend already was, socially disruptive. Hence, the discipline should evolve towards modeling these problems in the digital space and a critical review of the socio-technical effects of this limited extract of human reality. This does not necessarily require an entirely non-positivistic Computer Science nor disruptively novel components. It asks for a critical thought about current positivistic practice and an extended regard of interaction.

References

[1] Ayer, Alfred J.. *Language, Truth, and Logic*. Dover, New York, (2nd edition, 1952)

[2] Berners-Lee, T., Fischetti, M. *Weaving the web - the original design and ultimate destiny of the World Wide Web by its inventor.* Harper Business, (2000)

[3] Cass, S. *Unthinking Machines*. http://www.technologyreview.com/news/423917/unthinking-machines/?a=f, as visited on 12-04-2013

[4] Gephart, R. *Paradigms and Research Methods*. http://division.aomonline.org/rm/1999_RMD_Forum_Paradigms_and_Research_Methods.htm, as visited on 12-04-2013

[5] Van Kleek, Max, Smith, Daniel Alexander, Shadbolt, Nigel and schraefel, m.c. *A decentralized architecture for consolidating personal information ecosystems: The WebBox*. In, PIM 2012, Seattle, Washington (2012)

[6] Koschmieder, C. *Dauermeckern der Piraten gegen "die da oben"*. http://www.tagesspiegel.de/meinung/gastbeitrag-dauermeckern-der-piraten-gegen-die-da-oben/6622950.html, guest article in German newspaper Der Tagesspiegel (online), as visited on 12-04-2013

[7] Kunz, W. & Rittel, H. W. J. *Issues as elements of information systems.* Working Paper 131, Institute of Urban and Regional Development, University of California, Berkeley, California, 1970.

[8] Norvig, P. *On Chomsky and the Two Cultures of Statistical Learning*. http://norvig.com/chomsky.html, as visited on 12-04-2013.

[9] Tempich, C., Pinto, H.S., Sure, Y. and Staab, S. *An argumentation ontology for DIstributed, loosely-controlled and evolving engineering processes of oNTologies (DILIGENT)*. Proc. of the 2nd ESWC, Springer-Verlag (2005), 241-256

[10] Turing, Alan M. *On Computable Numbers, with an application to the Entscheidungsproblem* in proceedings of the London Mathematical Society 2(42): 230-265 (1936)

[11] Turing, Alan M. *Intelligent Machinery* reprint in Machine Intelligence, 5, chapter 1, Edinburgh University Press, Edinburgh, UK, *(*1969).