The Universal Personalized Approach for Human Knowledge Processing

Stefan Svetsky, Oliver Moravcik

Abstract—**The processing of information for supporting human knowledge based processes still represents a big challenge. This is caused both by the abstract concept of the information and by missing a suitable information representation (how to represent human knowledge), which would enable a human – computer communication in natural language. However, any interdisciplinary definition of "information" or even "knowledge" does not exist until now. In the period 2010 to 2012, the authors presented their research approach to the automation of teaching processes as typical human knowledge based processes within the WCE's International Conference on Education and Information Technology. The following research showed that solving the automation of the knowledge processes requires one to solve not only a content knowledge processing but also issues of communication, transmission and feedback. Actually, the research covers some areas of computer science (e.g. computer applications, computing methodologies, theoretical computer science). The question of the information was solved by the design of a specific default data structure so called virtual knowledge which enables an individual to communicate with computer in natural language. In this context, the virtual knowledge unit symbolizes an information representation for human knowledge which is readable both by a human and machine. Such approach enables one to input any kind of information from human resources into the virtual knowledge and to process human knowledge as it is typical for computers. This contribution illustrates how various kind of information from human sources are processed by the in-house software. It also enables one to understand why the software works as an all-in-one tool when processing human knowledge which is inputted into the virtual knowledge. In addition, the information transmission in the knowledge based processes was solved by a specific utility model for processing unstructured data.**

Index Terms—**computer applications, human knowledge processing, human computer interaction, information representation, personal informatics**

Manuscript received March 03, 2017; revised March 09, 2017. Stefan Svetsky. Author is with the Slovak University of Technology - Faculty of Materials Science and Technology in Trnava, Paulínska 16, 917 24 Trnava, Slovakia (corresponding author to provide phone: +421 949 541 835, e-mail: stefan.svetsky@stuba.sk).

Oliver Moravcik. Author is with the Slovak University of Technology - Faculty of Materials Science and Technology in Trnava, Paulínska 16, 917 24 Trnava, Slovakia; e-mail: oliver.moravcik@stuba.sk).

I. Introduction

In the period 2010 to 2012 an empirical research on technology enhanced learning (TEL) was. presented on the WCE's International Conference on Education and Information Technology [1,2,3]. As was presented on the conferences, the batch knowledge processing paradigm and in-house educational software BIKE (WPad) were developed and tested for design of teaching innovative methods within TEL. The software enables individuals to process large amount of data (the human information and knowledge). To program own software was needed because there was an absence of IT tools for TEL. As it is in [4] "a system design on the model basis has been widely ignored by the community" till now, and software engineering is missing in TEL system development.

After this initial phase, the following research was more systematic because the question of knowledge was needed to solve. There was another paradox yet. Despite the fact that each educational book, journal, research papers deals with knowledge, however, however, any interdisciplinary definition of knowledge exists. From terminologically point of view, the definition of information or knowledge has other meaning in computer science, information and communication technology, knowledge management, psychology, teaching and learning, or how laics understand it. Moreover, in computer science there is a big challenge, how to represent "human knowledge". So if one should solve any automation when human knowledge is the key parameter he must design a suitable representation for knowledge or information.

This challenge and considerations were reminiscent of the early period of Cybernetics. For example, in [5] teaching processes were considered as controlled "processes of acquisition, processing and storing information". Between two of its sub-system, i.e. teacher and student, the transmission of information exists on a basis of feedback loop. It is emphasized as well that the "cybernetization " enables one to model a control the teaching process.

In this context, the Cybernetics (within which computers were invented) represents the basic theoretic discipline focused on the automation (regulation, control) of such complex systems, i.e. on the information, information transfer and processing [6]. Teaching and learning processes are considered for cognitive processes, i.e. they are human knowledge based, thus logically connected with thinking and intelligence. Within psychology, behaviorists understand thinking as a process of solving problems (as an adaptation to the changes in environment), while in cybernetics it is a process of processing and using information , as in [7].

The Cybernetics as a science field deals in general with creation, transmission and processing of information [6]. According to the Shannon's information theory, as a part of the field, one bit represents the basic information unit (with the value 0 or 1). Computers are machines for information processing by using such sequences of numbers 0 or 1. A default block of the sequences then creates data types which enable the machine to process texts, images, video via standardized formats. In [6] Kotek mentions three category of information: syntactic, semantic and pragmatic. In the content of his book, there are chapters dedicated to machine learning, natural language recognition, artificial intelligence. Kotek emphasizes these topics should be considered as informatics constructs derived only from syntactic information, because at this time there was not enough research data affecting semantic information, especially pragmatic information (the pragmatic information should be an useful information within a report).

In other words, computers do not understand human natural language, so any "human information" must be converted to "machine information". It is important for understanding the terminological chaos in the current scientific literature. For instance, in the field of Information and Communication Technology, similarly as in Knowledge Management, there is other hierarchy common: *data* (strings, numbers), *information*, *knowledge* (how it is also understood by laics). This contribution shows on several examples from implemented case studies how the problem of human knowledge processing was solved based on designing a virtual knowledge unit which is readable both by human and machines (computers).

II. Challenges for Human Knowledge Processing

Other obstacles when programming human knowledge based processing are evident from the following schema.

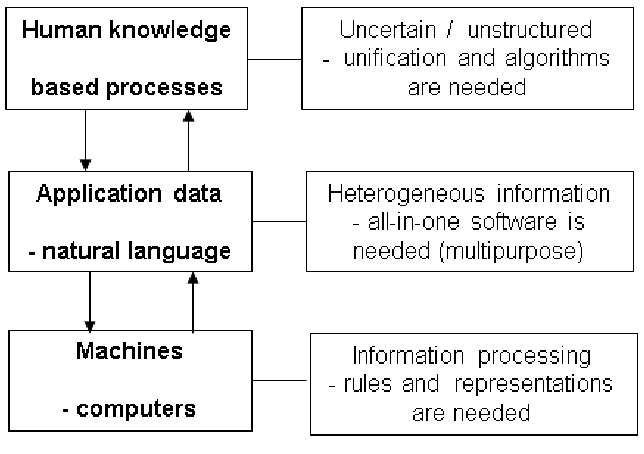

Figure 1. Basic schema for using computers to process human knowledge.

Fig. 1 illustrates that the human knowledge based processes are uncertain, unstructured, thus, an unification and solving algorithms are needed. Moreover, the application data from human resources are in natural language and heterogeneous. Therefore, an universal IT tool is needed which would work like an all-in-one software. Moreover, as was mentioned above, for the processing of information, which is collected from human resources, computers need specific rules and representations in order to process it as the sequences of numbers 0 and 1. In this view, e.g. clustering short texts by their meaning is considered as a challenging task in [8]. Authors even propose a semantic hashing approach for encoding the meaning of a text into a compact binary code. Other similar approaches can be found in literature concerning question of the syntactic or pragmatic information, see e.g. [9, 10]. These approaches are related to the Shannon's information theory, respectively as it was mentioned above in the context of the Cybernetics. However, from our point of view, none of the approaches relate the processing of human knowledge directly in the natural language. It must be emphasized, that all of them are dedicated to automatic information processing. These approaches try to solve how to find or input the meaning of "human information" into the syntactic, semantic or pragmatic "machine information". In other words, one could imagine that all activities are done by a machine.

The same situation could be mentioned regarding Information Technology where ontologies "seem to be right way how to represent knowledge in machine readable form" [11]. It is emphasized that for information and knowledge sharing among users, the web-based systems need machine readable metadata, which define meaning of human readable-data. The so called Semantic Web uses for this purpose ontologies approaches to semantic representation (e.g. abstract representation of real objects, rules of a domain represented by knowledge graphs).

In contrast to these approaches, this contribution presents an entirely different approach which is based on the synergic collaboration between the human and the machine. For this purpose a virtual unit knowledge was designed which simply works as a switch between human mental processes and machine (computer). This enables one to simulate and computerize mental processes, i.e. human knowledge based processes, by using this virtual knowledge. From a human point of view, the content of the virtual knowledge means simply any information or knowledge as it is understood by people, which is written in natural language. From a machine (computer) point of view, the machine understand the virtual knowledge unit as a "user data type", thus, as a kind of information representation which uses ASCII codes. An empty virtual knowledge has ca 10 kilobytes. One could imagine that the empty default data structure is a primitive knowledge (knowledge cell / knowledge table). After inserting the human knowledge in natural language into the virtual knowledge one could measure in bytes how "wise" it is. From another point of view, it is important that the virtual knowledge enables a user to process unstructured data in a user friendly way. In other words, the content of the virtual knowledge is both human and machine readable. Thus such approach seems to be interdisciplinary understandable even from a layman's point of view.

A. Examples: The batch knowledge paradigm

Visual FoxPro database platform seems to be one of the most user friendly software dedicated to Personal informatics Especially, due to the specific programming language FoxPro which enables users to process even large data in batches. In other words, , human information and knowledge can be processed in natural language even by users with lower informatics skills. For example, a user can

use the following simply command to numerate hundred or thousand of records (virtual knowledge units) in the column *txtuni* (“universal text”) :
Replace all txtuni with alltrim(str(RecNo()))+"."+alltrim(txtuni) for not "."$substr(txtuni,1,5)

Fig. 2 shows the result, i.e. all records has been numbered with the exception of records that contain a dot.

332. ??? MYSQL - testy php
331. MYSQL - testy php
330. phpMyAdmin localhost PC svti
329. phpMyAdmin PC škola
328. KZP návrh
327. sessions 09.10.201615:53:08
326. MIX SME

Figure 2. Screenshot of the WPad table column txtuni after numerating.

The batch knowledge processing paradigm was developed by using thousands of such commands. The content of the knowledge tables is controlled and handled by the in-house WPad software (BIKE software is installed only on the computers of the author Svetsky).

B. Examples: The personalized Learning Analytics

The state-of-the-art in the field of Learning Analytics is focusing on statistical monitoring and sampling of simple online data. It is based on the activities of students in online learning environments, but not on they knowledge. In our case, the research approach is focused directly on the knowledge flow because WPad works as well as a convertor of tacit knowledge into explicit knowledge (according to the terminology of the area of knowledge management), so once a student or teacher writes something into the WPad, then this represents the explicit knowledge. Therefore, such approach is closer to human mental processes, i.e. beyond the state-of-the art.

It can be demonstrated how the teacher - designer, monitored his own activities within the research on learning analytics. He wrote a specific programming code which indicated how many time the “knowledge” table was opened by him. In this case, such monitoring can indicate how the personalized data relate not only to his activities but also to the teaching content (knowledge), etc.

Fig. 3 illustrates such personalized data from his home computer. For example, one can see that the knowledge table was opened by the teacher around five thousand times within ten months.

:: 5250 :::: 03.10.2016 || 20:21:06 ::
:: 5251 :::: 04.10.2016 || 11:39:43 ::
:: 5252 :::: 04.10.2016 || 11:41:40 ::
:: 5253 :::: 04.10.2016 || 12:15:42 ::
:: 5254 :::: 04.10.2016 || 12:15:43 ::
:: 5255 :::: 04.10.2016 || 15:30:20 ::
:: 5256 :::: 04.10.2016 || 15:31:06 ::
:: 5257 :::: 04.10.2016 || 15:31:08 ::
:: 5258 :::: 04.10.2016 || 15:32:18 ::
:: 5259 :::: 04.10.2016 || 15:32:20 ::
:: 5260 :::: 05.10.2016 || 12:29:59 ::

Figure 3. Screenshot of the evidence of the personalised data.

C. Examples: The text extraction from large files

The user had at his disposal proceedings with 250 pdf-files from a conference in Villach (Austria). He wanted to analyze research focus of contributions. Firstly he joined the 250 files into one file with Nuance pdf-software and saved as one plain text file. Then inserted the file into the knowledge table, extracted keywords and found the frequency of their occurrence. Fig. 4 illustrates the result in the same knowledge table.

vocational	0.002641	11
virtual	0.005282	31
remote	0.005522	23
enhanced	0.004322	6
mobile	0.005282	22
language	0.002641	11
e-learning	0.002161	9
distance	0.002161	9
control	0.003361	14
component	0.004802	47
assessment	0.002641	11
management	0.002161	9
collaborative	0.002881	29

Figure 4. Screenshot of the text extraction from 250 pdf-files.

D. Examples: The use of digital pen

The use of digital pen was also tested in order to automate transmission of handwritten schemas and chemical formulas to computer. Fig. 5 illustrates personal notes of the designer made by the digital pen (in relation to the principle how human knowledge from various human sources is processed via WPad for knowledge based processes).

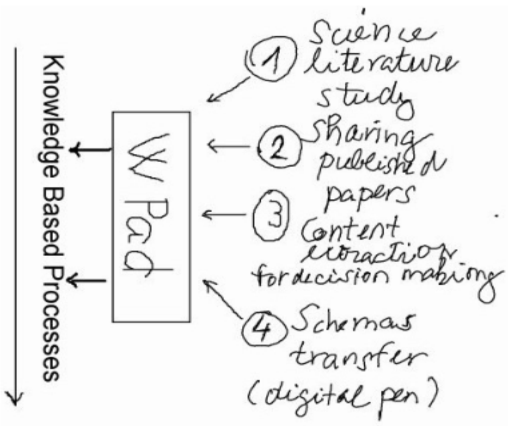

Figure 5. Example of an image made by the digital pen.

Despite the fact that the image seems to be quite good, it took more time when one was using a paper and pen. So it has been used only once, as a signature list of students (there was also a problem with the ink replacement, incl. the price).

E. Examples – The modelling collaborative activities

The design of the universal virtual knowledge unit which is controlled by the in-house software BIKE (WPad) as a knowledge table (by using the Visual FoxPro database platform) was a first step to complex human knowledge processing.. The second step was a solution how to transmit the virtual knowledge between off-line and online environments in order to solve various collaborative activities. This was possible after implementation of the Utility Model 7340/2014 (Slovak Patent Office, registration on December 2015) which relates to processing and transmission of unstructured data. The question of the data transmission was important for solving collaborative activities on a shared virtual environment.

In this view, some applications were developed with focus on the so called Human Language Technology and Computer Supported Collaborative Learning (CSCL).

There was a project idea to create personal corpuses which would contain multilingual human expert knowledge from the field of machinery engineering [12]. The CSA proposal objectives were focused on cooperative benchmarking for the design of the pilot human processes-driven system of automated Machine Translation consisting of research infrastructure of recommended multilingual repositories, specialized corpora, ontological batch knowledge sets, aMT software and a federated trans-lingual pan-European portal. In other words, the expert human knowledge should be shared and exchanged via the shared virtual environment within the portal.

The developing CSCL application represents progress within the participation research on Technology-enhanced learning. The actual state of the research is illustrated by Fig. 6. Within the supportive IT infrastructure a shared virtual environment works for educational knowledge exchange.

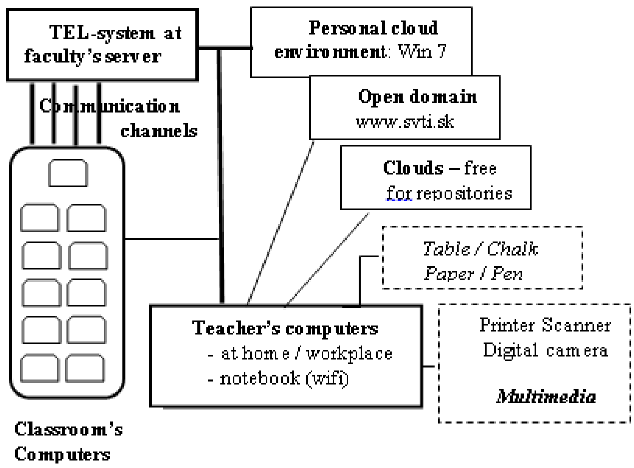

Figure 6. Schema of the developed supportive IT infrastructure.

The first results when modelling the CSCL-knowledge exchange and sharing shows, that the processing of the human knowledge in a knowledge based process requires one to solve three categories of algorithm – *pedagogical* (didactic methods), *informatics* (adaptation to the operating system) and *integrated* (CSCL -communication and human knowledge transmission). In this case, a design of templates for sharing the human knowledge is needed. Fig. 7 shows such template for sharing annotation in virtual environment. The lower part of the figure represents annotation data in WPad and the upper part data after the conversion to html.

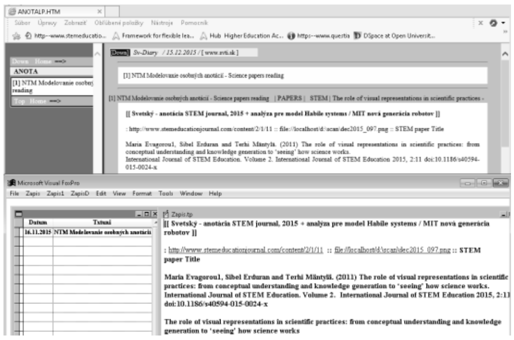

Figure 7. Scrrenshot of the annotation template.

III. The Future works

The above mentioned examples demonstrated universal "all-in-one" function of the in-house software BIKE (WPad) when processing human knowledge. Batch internet retrieving, extraction data from WEB databases, text processing, collaboration with internet search or translation services (these are also used by bachelor students) can also be mentioned as examples. One could speak about a system "evolution". Namely, each software has a user menu with ten to twenty items divided into certain categories. By solving a tailor made modification of the in-house software, when modelling the support of knowledge based processes, a combination of menus had to be used. This resulted in a new category of applications. In the future, each of them could be developed as an autonomous software, or solved as classes in object oriented programming.

These all-in-one activities are similar to the ones made by so called multi-agents or robots. For example, at this time a possibility to extract data from bibliographical MARC format was tested by the BIKE software. Format MARC (MAchine-Readable Cataloging), is a data format and set of related standards used by libraries to encode and share information about books and other material they collect. In future works, there would be a possibility to build an application for bibliographical purposes (a citation or annotation system). In this case also programming a personal multilingual support system for writing science papers could be planned (even this idea was worked out when writing this paper). In an education environment specific applications could be developed as well (see for inspiration in [14]).

In this context, authors of this contribution see a certain analogy to Nilsson's vision of "Habile Systems" in the field of Artificial Intelligence [15]. He wrote "Rather than work toward this goal of automation by building special-purpose systems, I argue for the development of general-purpose, educable systems that can learn and be taught to perform any of the thousands of jobs that humans can perform". Fig. 8 illustrates how authors were inspired by the Nilsson's vision. In other words, the BIKE(E) should work as a black box for human knowledge processing. The outputs would be thousands of applications, which are useful for various human "jobs". So future works leads to a vision of an "educational robot" or "multipurpose intelligent system".

ISBN: 978-988-14047-4-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

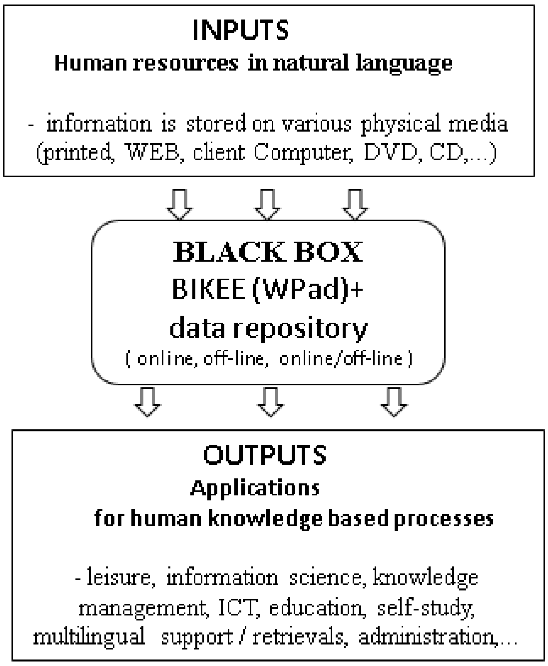

Figure 8. Schema of function in-house software as a black box.

IV. Conclusion

In this contribution, the universal personalized approach for human knowledge processing within knowledge based processes was described. A specific focus was given on explaining this approach to the information representation of human knowledge by a specific default data structure - so called virtual knowledge unit. Such approach seems to be related to "pragmatic" information because the human knowledge is useful from the users point of view (e.g. a user can make decisions).

This enables one to simulate and computerize mental processes, i.e. human knowledge based processes, by using the virtual knowledge. From a human point of view, the content of the virtual knowledge means simply any information or knowledge as it is understood by people, which is written in natural language. From a machine (computer) point of view, the machine understands the virtual knowledge unit as an "user data type", thus, as a kind of information representation which uses ASCII codes. In addition, the virtual knowledge enables the user to process unstructured data in a user friendly way. In other words, the content of the virtual knowledge is both human readable and machine readable. Any other rules, representations, languages, semantics graphs, etc. are not needed. It all works as an all-in-one IT tool, a black box, because if the user finds any way how to write human knowledge into the virtual knowledge, it automatically means, that he embedded into the data structure his own semantics and ontology.

This was illustrated via several examples from human knowledge processing (in each screenshot, a different kind of human knowledge was used). Thus, the inputs were human knowledge (information) in natural language with user's embedded semantics and ontology. The outputs, which are produced by the computer, were also in natural language. This enables user to "automate" his mental processes.

Moreover, if one defined a set of algorithms for processes where the human knowledge is a basic parameter, many applications could be developed in the future. This seems to be a certain analogy to Nilsson's vision of "Habile" systems. The analogy lies in the idea that the applications for human knowledge processing could be performed by a "Virtual Educational Robot" functioning as a "multi-purpose intelligent system" (see the Partner Search section within the IDEAL IST web-page: http://www.ideal-ist.eu/ps-sk-99947).

References

[1] S. Svetsky, O. Moravcik, J. Stefankova, and P. Schreiber, "The Implementation of the Personalized Approach for Technology Enhanced Learning," in Lecture Notes in Engineering and Computer Science: World Congress on Engineering and Computer Science 2010, pp. 321-323.

[2] S. Svetsky, O. Moravcik, P. Schreiber, and J. Stefankova, "The Informatics Tools Development and Testing for Active Learning," in Lecture Notes in Engineering and Computer Science: World Congress on Engineering and Computer Science 2011, pp. 265-268.

[3] S. Svetsky, O. Moravcik, P. Schreiber, and J. Stefankova, "The Informatics Tools Development and Testing for Active Learning," in Lecture Notes in Engineering and Computer Science: World Congress on Engineering and Computer Science 2011, pp. 265-268.

[4] A. Martens," Software Engineering and Modeling in TEL," in Book: The New Development of Technology Enhanced Learning Concept, Research and Best Practices. Editors: Ronghuai Huang, Kinshuk, Nian-Shing Chen. Springer, 2014, p. 27-40.

[5] L.Habinak, "Využitie výpočtovej techniky pri riadení výchovno-vzdelávacieho procesu (The use of computation technique for the controlling educational processes)," in Conference: The use of computation technique in Education, DT CSVTS Bratislava,1984, pp. 11-21.

[6] Z. Kotek, P. Vysoky, Z. Zdrahal, Kybernetika (Cybernetics).Praha: SNTL, 1990.

[7] J. Boros, Zaklady psychologie (Basics of Psychology). Bratislava: SPN, 1976.

[8] Zheng Yu, Haixun Wang, Min Wang, " Understanding Short Texts through Semantic Enrichment and Hashing," IEEE Transactions on knowledge and data engineering, vol. 28, no. 2, February 2016.

[9] Choi, Freddy Y. Y., Peter Wiemer-Hastings, and J.Moore, "Latent semantic analysis for text segmentation," in n Proceedings of the Conference on Empirical Methods in Natural Language Processing. Pittsburgh: PA, Carnegie Mellon University,, 2016, p.p. 109–117.

[10] E. Weinberger, "A Theory of Pragmatic Information and Its Application to the Quasi-species Model of Biological Evolution," Max Planck Institute for Biophysical Chemistry. Available: https://pdfs.semanticscholar.org/3ade/f57e8f5c34566fc7f1e56079b7114631238f.pdf.

[11] M. Bieliková, et al., " Personalized Conveying of Information and Knowledge. Studies in Informatics and Information Technology," In Research Project Workshop Smolenice. Bratislava: SUT Press, 2012, pp. 53-86.

[12] S. Svetsky et al, "Pan-European Language Network 21+3 for designing the infrastructure of Human Processes - driven System of Machine Translation for specialized high-quality Translation," Proposal No.: 643961 submitted to H2020-ICT-2014-1 Call.

[13] S. Svetsky, O. Moravcik, "The Implementation of Digital Technology for Automation of Teaching Processes". Presented at the FTC 2016 - Future Technologies Conference(2016, San Francisco) Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7821632.

[14] D.G. Moursund, "A Faculty Member's Guide to Computers in Higher Education," (April 2007). Available: http://uoregon.edu/~moursund/Books/Faculty/Faculty.html.

[15] N. J. Nilson, "Human-Level Artificial Intelligence? Be Serious!," in AI Magazine, 2005. Available: http://ai.stanford.edu/~nilsson/OnlinePubs-Nils/General%20Essays/AIMag26-04-HLAI.pdf.