

Elsevier Editorial System(tm) for International Journal of Medical Informatics

Manuscript Draft

Manuscript Number: IJMI-D-05-00101R1

Title: The usability axiom of medical information systems

Article Type: Special Issue Article

Section/Category:

Keywords: Artificial Intelligence; Decision Making, Computer-Assisted; Information Systems; Medical Informatics; Algorithmic Information Theory; Case-based reasoning

Corresponding Author: Mr Stefan Valerian Pantazi, MD

Corresponding Author's Institution: University of Victoria

First Author: Stefan Valerian Pantazi, MD

Order of Authors: Stefan Valerian Pantazi, MD; Andre Kushniruk, PhD; Jochen R Moehr, MD, PhD

Manuscript Region of Origin:

Abstract:

Introduction

In this article we begin by connecting the concept of simplicity of user interfaces of information systems with that of usability, and the concept of complexity of the problem-solving in information systems with the concept of usefulness. We continue by stating "the usability axiom" of medical information technology: information systems must be, at the same time, usable and useful. We then try to show why, given existing technology, the axiom is a paradox and we continue with analysing and reformulating it several times, from more fundamental information processing perspectives.

Discussion

We underline the importance of the concept of representation and demonstrate the need for context-dependent representations. By means of thought experiments and examples, we advocate the need for context-dependent information processing and argue for the relevance of Algorithmic Information Theory and Case Based Reasoning in this context. Further we introduce the notion of concept spaces and offer a pragmatic perspective on context-dependent representations. We conclude that the efficient management of concept spaces may help with the solution to the medical information technology paradox. Finally, we propose a view of informatics centred on the concepts of context-dependent information processing and management of concept spaces that aligns well with existing knowledge centric definitions of informatics in general and medical informatics in particular. In effect, our view extends M. Musen's proposal and proposes a definition of Medical Informatics as context-dependent medical information processing.

Summary

The axiom that medical information systems must be, at the same time, useful and usable, is a paradox and its investigation by means of examples and thought experiments leads to the recognition of the crucial importance of context-dependent information processing. On the premise that context-dependent information processing equates to knowledge processing, this view defines Medical Informatics as a context-dependent medical information processing which aligns well with existing knowledge centric definitions of our field.

The usability axiom of medical information systems

Stefan V. Pantazi

Andre Kushniruk

Jochen R. Moehr

School of Health Information Science, University of Victoria, BC, Canada

Keywords

Artificial Intelligence, Computer-Assisted Decision Making, Information Systems,
Medical Informatics, Algorithmic Information Theory, Case-based reasoning

Corresponding author

Stefan V. Pantazi, MD

School of Health Information Science

University of Victoria

HSD Building, Room A202

Victoria, BC, Canada, V8W 3P5

Email: spantazi@uvic.ca

Phone (250) 472-4300

Fax (250) 472-4751

Abstract

Introduction

In this article we begin by connecting the concept of *simplicity* of user interfaces of information systems with that of *usability*, and the concept of *complexity* of the problem-solving in information systems with the concept of *usefulness*. We continue by stating “the usability axiom” of medical information technology: information systems must be, at the same time, *usable* and *useful*. We then try to show why, given existing technology, the axiom is a paradox and we continue with analysing and reformulating it several times, from more fundamental information processing perspectives.

Discussion

We underline the importance of the concept of representation and demonstrate the need for context-dependent representations. By means of thought experiments and examples, we advocate the need for context-dependent information processing and argue for the relevance of Algorithmic Information Theory and Case Based Reasoning in this context. Further we introduce the notion of concept spaces and offer a pragmatic perspective on context-dependent representations. We conclude that the efficient management of concept spaces may help with the solution to the medical information technology paradox. Finally, we propose a view of informatics centred on the concepts of context-dependent information processing and management of concept spaces that aligns well with existing knowledge centric definitions of informatics in general and medical informatics in particular. In effect, our view

extends M. Musen's proposal and proposes a definition of Medical Informatics as *context-dependent medical information processing*.

Summary

The axiom that medical information systems must be, at the same time, useful and usable, is a paradox and its investigation by means of examples and thought experiments leads to the recognition of the crucial importance of context-dependent information processing. On the premise that context-dependent information processing equates to knowledge processing, this view defines Medical Informatics as a *context-dependent medical information processing* which aligns well with existing knowledge centric definitions of our field.

Brief

1. The axiom that medical information systems must be, at the same time, useful and usable, is a self-contradiction (i.e., paradox) because it implies that systems are somehow able to fill in, automatically, the information gap between *usable, simple, abstract descriptions* and the *useful, relatively complete* representations required to solve complex problems.
2. The axiom is centred on the problem of representation which fundamentally, could be defined as a function that maps the reality of an object to be represented onto an arbitrary representation medium.
3. Context-dependent representation functions are a higher order of information processing that could be better described as knowledge processing. Additional evidence of the importance of context-dependent information processing comes from Algorithmic Information Theory (AIT) and Case Based Reasoning (CBR) research.
4. Context-dependent information processing is a computationally intensive problem that could be addressed through an efficient management of concept spaces that are extremely high dimensional, extremely sparsely populated, dynamic and which exhibit a similarity-based organization.
5. One of the most efficient ways of managing concept spaces with the aforementioned properties is by linking entities explicitly in richly connected structures (i.e., graphs).

6. Medical Informatics could be defined as *context-dependent medical information processing* and this aligns well with existing knowledge centric definitions of our field.

Introduction

Information systems are commonly thought of being composed of two important parts: their *user interface* and their *problem-solving* engine. From a usability engineering perspective, if one connects the concept of *usability* with the user interface of information systems and the concept of *usefulness* with the complexity of problem-solving, the following “usability axiom” holds: information systems must be, at the same time, *usable* and *useful*. However, because *usable* user interfaces need to be *simple* and because *useful* information systems able to solve complex problems require *complex* problem-solving engines, the usability axiom is also a paradox: information systems must be, at the same time, both *simple* and *complex*. The paradox vanishes only if we could divide information systems completely and address separately their user interfaces and their problem-solving engines. Only in this case would we be able to build *simple*, low complexity user interfaces to systems that solve *complex* problems. Unfortunately, given existing information technology, the concepts of usability and usefulness are dependent of each other, as if their sum remains constant. Highly *usable* systems are often less *useful* because they typically solve trivial problems (e.g., generic, repetitive tasks). Conversely, potentially very *useful* systems that could solve complex, specific problems (e.g., planning a trip, or devising a therapy plan for a certain patient) usually end up exhibiting *usability* problems. This is so because, given a limited problem-solving engine, the complexity of the task to be solved spills over into the complexity of the user interfaces which subsequently may become an error causing factor [1].

Before attempting to analyse the paradox, one also needs a proper characterization of the prototypical problem that medical information technology is supposed to solve. One possibility is to regard knowledge intensive decision-making – which includes prediction of outcomes of complex realities – as the prototypical informatics problem. At the extreme lies the prediction of one bit of information (for example, to give/not to give a certain drug to a certain patient) given a particular, usually extremely complex, patient description. If one disregards the medical connotation, this problem also falls under the realm of artificial intelligence (AI) and as with any AI problem, in order to be solved, it must be represented in the memory of a problem-solver. In the context of an information system, this implies that a description of the problem must be acquired, typically, through a user interface. But the spatio-temporal reality of a patient is complex, multidimensional and dynamic and so must be the approaches to represent it without loss of information. Because it is based on abstractions, the process of acquisition of dynamic, complex, complete representations through human-computer interfaces is extremely difficult because of at least two interconnected problems: the knowledge acquisition bottleneck [2] and the frame problem [3-5]. Relevant to any expert system design, the former problem is more related to the amount of knowledge required to overcome brittleness (i.e., the susceptibility to fail outside a narrow domain), while the latter puts the emphasis on the need for representational approaches that can cope with dynamic, rapidly changing environments. In Medical Informatics, these problems are justifications for “the enormous upkeep effort” required to manually maintain dynamic, complex medical knowledge bases that are not only vast but may rapidly become obsolete [6].

The focus on the representation issue reveals the strong coupling between the user interfaces and problem-solving engines of information systems and allows us to restate the usability axiom of medical information technology from a more fundamental perspective: information systems must be able to create *useful, relatively complete* (i.e., complete with respect to a certain purpose) internal representations of complex spatio-temporal realities (e.g., description of a patient), with minimal information loss, from simple, incomplete, abstract descriptions acquired through *usable, simple* human-computer interfaces. The self contradiction (i.e., the paradox) arises from the assumption that information systems are somehow able to fill in, automatically, the information gap between *usable, simple*, abstract descriptions and the *useful, relatively complete* representations required to solve complex problems. Historically speaking, this paradox has existed from the beginnings of information science and artificial intelligence. The usual form to be presented was as the dichotomy between symbolic processing and machine learning paradigms. Symbolic processing is the tenet of hand coded knowledge bases and rule driven expert systems. It is no surprise that these lack usefulness and suffer from the knowledge acquisition bottleneck [2] and frame problem [3-5]. This suggests that machines that learn are the only alternative with the potential to address the paradox by attaining automatically, relatively complete, useful representations while allowing for simple, usable user interfaces of information systems. The limitations of this alternative seem to lie only in the memory and processing power required for advancing from the stage of proof-of-concept “toy applications” to that of real world problem solvers.

In Figure 1 we illustrate a hypothesized relationship between usability (i.e. simplicity and ease of use of a system or user interface) and complexity/difficulty of problems solved by an information system. Ideally, in healthcare we wish to “have our cake and eat it too” [7] - i.e., have systems that perform highly useful and complex tasks, and yet are easy and simple to use. But the expectations of systems being both highly *usable* as well as being capable of performing complex and *useful* tasks are often not met in practice. It is also unlikely that this apparent contradiction is more striking in other domains than in health care, particularly in the area of applied artificial intelligence in healthcare. For example, early attempts at creating knowledge-based expert systems spawned development of a wide range of applications and systems, including MYCIN, INTERNIST and many other such systems [8]. These applications were expected to help physicians solve highly complex diagnostic tasks and yet do so in a manner that was easy to use, highly useful and that fit within the complex work practices of healthcare [9]. However, as well-stated by Miller and Masarie in their insightful and landmark article “The Demise of the ‘Greek Oracle’ Model for Medical Diagnostic Systems” it became clear that such systems that were designed to be highly useful by solving complex medical problems, would not be accepted by health care practitioners until they were also designed to actually be highly usable and simple in their operation [10]. The field of applied artificial intelligence in medicine has been marked by issues, controversy and problems arising from the tension between usefulness and complexity on the one hand and usability and simplicity on the other. In particular, newer approaches to developing systems based on intelligent agents, adaptable user interfaces and intelligent tutoring systems [11], are all examples of AI systems developed to solve complex problems, while at the same time

attempting to appear simple and usable to end users. This has become an ongoing challenge both theoretically and pragmatically.

In this paper it is argued that the technical solution to addressing the paradox is centered on the thorny problem of representation. Complicating the user side of the equation with less *usable*, sophisticated representations is impracticable, because knowledge acquisition cannot be achieved only with handcrafted symbolic representations. Given the particular features of the prototypical informatics problem, the exact nature of the solution seems to consist of endowing information systems with the capacity to acquire, with as high a degree of autonomy as possible, *useful*, *relatively complete*, *problem-specific* representations of the complex problems that need to be solved. To a certain extent, systems that “can learn” could be thought of being able “to program themselves.” If they are considered “intelligent,” solving the paradox equates to building intelligent systems.

Discussion

What does “to represent” mean?

In the previous section we introduced the concept of representation but we did not define it. We will therefore continue with a conceptual exploration of the semantic space of the concept of representation. Hopefully this will allow us to understand it at a more fundamental level and lead to a more formal definition. We use one of the best-known available sources of general semantic knowledge, WordNet [12], and explore the semantic neighbourhood of the concept in the syntactic roles of noun (i.e., “representation”) and verb (i.e., “to represent”).

All twenty-five distinct senses (Table 1, Table 2) attest to the semantic richness of the concept and suggest a potentially high degree of ambiguity. However, despite various usage contexts, all different senses of the concept retain a certain abstract, prototypical meaning. Senses of high interest are 1, 2 and 10 for the noun “representation” (Table 1) and 1, 2, 3, 5, 6 and 15 for the verb “to represent” (Table 2). Additional semantic information is gathered from related concepts that fall in general under the patterns “X is a representation” and “to X is to represent.” The two non-exhaustive categories of lexical items are labelled “REPRESENTATION” and “REPRESENT” and listed in Table 3.

From the semantic information gathered about the concept of representation one can infer that, fundamentally, the process of representation can be defined (i.e., represented) by a function (in the mathematical sense) that maps the reality of an object to be represented onto an arbitrary representation medium. Let us call this information processing function a *representation function*. Therefore, given a certain spatio-temporal reality R (e.g., a moving object), the concept of representation is a function F of R , onto another spatio-temporal region I , which is the representation, image, model or projection, that represents, stands, is projection of, or corresponds to R . Put symbolically:

$$F(R)=I$$

For example, if R were a real, actual telephone set, then, given a certain function F , $I=F(R)$ could conceivably be any of the possible representations in Figure 2.

Essentially, F can be thought of as an algorithm which takes as inputs features from R (e.g., colors, shapes, sounds, spatio-temporal patterns) and maps them onto features (e.g., pixels of an image, words in a text, phonetic symbols of a phonetic transcription) of a medium which could be paper, clay, a blackboard, magnetic media and computer memory, in order to create an image I that is a projection of R . If F is implemented into a computing device (e.g., a digital camera, a computer), then the inputs are binary sequences generated by sensors such as imaging devices or microphones, while representation media are typically computer memory, magnetic media or paper. If F is a human information processing function, then the representation medium is our memory. One property of F is important and follows immediately. Applying F to a reality R results invariably in loss of information: for any F , there are always features in R (e.g., features at molecular or atomic levels, features in obscured sides of an object, features in a distant history of an event or that are considered non-relevant for a specific purpose, etc.) that do not make it into the representation I of R and which are therefore lost. This property is related to the limits of human perception of recording and measurement hardware and of the representation media. It could also be related to a particular, relevance-driven implementation of a representation function F which strips out from a representation I the features considered irrelevant. This is what M. Blois refers to as “necessarily incomplete descriptions of natural objects” or “abstractions” [13] (p. 32). However, unlike the case of abstractions, applying a certain F (e.g., a lossless image

compression algorithm) to an existing digital object (e.g., a digital image) could result in a complete, non-abstracted representation (e.g., a compressed data file) that loses no information.

Yet, for this discussion, the most important characteristic of F is with regard to its dependence on context. In Medical Informatics the importance of context and context-sensitive processing has already been recognized [6, 13]. Blois shows why context is needed in order to make representations understandable by providing counter-examples of “totally out of context” messages and points out that “ordinary utterances are always made in a situation, and will thus have a context” [13] (p. 22-23). He acknowledges the role of context in the resolution of ambiguity [13] (p. 36) in human communications but also underlines the implicit nature (as opposed to explicit) of context that causes its “critical role” to be underestimated [13] (p. 33). In a similar vein, Moehr shows how context-deprived representations (e.g., “peptic ulcer”) do not stand necessarily for an uniform reality, that their processing must depend upon the context where they were gathered and that the difficulty of processing of health data is caused precisely by its high context sensitivity [6] (p.251). At the same time, the context dependence of medical data seems to contradict the existing biomedical research paradigms such as populational studies (e.g., randomized control trials) which purposely aim at removing the context around medical data in order to attain population-wide applicability of their findings. [5].

This very important context perspective allows us to distinguish two extreme types of F :

- 1) The simple, non-evolving, context-independent, non-adaptive representation functions,
- 2) The complex, evolving, context-dependent, adaptive representation functions.

For example, a representation function in the first category is a “shrink” algorithm that removes every second pixel from images such as those in Figure 3. Under the same category falls a zoom-out function on a world-map that generates increasingly larger scale representations. Another example could be a simple algorithm that transforms a color image into its black and white representation (e.g., Figure 2 b and c). In general, representations that are affine transformations of an input (i.e., which preserve colinearity and ratios of distances [14]) could be thought of as belonging in this first category. Such functions can be applied multiple times, in various orders, even recursively, and because they are context-independent (e.g., an image to be processed could contain anything), their implementation is trivial and holds for any input.

An example for a representation function in the second category is one that accepts as input a reality R or existing representations such as images in Figure 2 (a,b,c) and outputs abstract, conceptual, symbolic representations such as Figure 2 (d, e). Such representation functions are highly context-dependent, and are often implemented as evolving, adaptive algorithms. These algorithms fall under the realms of artificial intelligence and machine learning, which are known to be challenging tasks for existing information technology.

To summarize, representation functions can be arbitrarily complex but are virtually non-evolving, context-independent functions in the case of existing information technology: the output of a simple representation function is little or not at all altered by the history of its outputs. On the other hand, context-dependent representation functions such as those that occur in the brain are dynamic (changing with time, adaptive) and highly context-dependent: new information is processed in the context of an individual's experience relevant to that new information. Such representations could be considered a higher order of processing better described as knowledge processing rather than just information processing. This leads to defining knowledge processing as context-dependent information processing.

Context-dependent information processing

In order to better illustrate the importance of context for information processing we propose a simple thought experiment that involves the application of the same prototypical, abstract representation function to two information artefacts that are different in nature: an image (e.g., Figure 3) and a text (e.g., a textbook). In the case of an image, the representation function of choice is extremely common in information processing and performs what we would call a *zoom-out* operation which allows the visualization of various levels of abstraction in the case of images, from small scale to larger scale features. The inverse of this function would be equivalent to a *zoom-in* operation which allows one to focus on details. When such details are available in the original object, the zoom-in function is simple, otherwise the complexity of the function increases in order to account for the additional information needed to make up those details.

In the case of a text, the function is the *summarization* operation that also results in representations with various levels of abstraction (i.e., summaries) of the original text. The inverse of a summarization function aims at obtaining less abstract, richer representations from a reduced, abstract representation and would stand for what one would call *text generation* or *synthesis*. This is a function of increased complexity that often has to make use of additional information not explicitly available in the original text.

A plausible, context-independent zoom-out function implementation is one which eliminates every second pixel from a given image, thus providing various levels of abstraction of that image (e.g., Figure 3). An alternative way to look at this problem is that, for such a zoom-out operation, the identification of the pixels to be removed from the image is entirely dependent on their absolute position in the image, i.e., on their XY coordinates, but not on their colour content. This allows for an efficient context-free processing.

By the same token, a possible context-independent text summarization could be trivially implemented in a similar, context-independent way. However, eliminating every second character from a word or every second word from a text, or every second paragraph from a page or every second page from a textbook, while extremely efficient, will not result in a good summary of that text. Therefore, context-independent representation functions do not allow the appropriate visualization of various levels of abstraction of textual artefacts. This is due to the conceptual nature of textual data as well as to the fact that textual items (e.g., characters, words, phrases,

sentences, paragraphs, etc) that should be part of a summary cannot be reliably identified through their absolute position in a given text. Such identification of key items can only be attained by more sophisticated context-dependent implementations of representation functions.

This simple thought experiment shows that simple, context-free representations are never appropriate for conceptual, textual information artefacts [15] or for any other representations that make use of abstract, conceptual representations, as is the case with the user interfaces of computer systems in general [13] (p. 36) and of medical information systems in particular [6]. The amount of dependence on context seems to be the only difference between a zoom-out on an image and a text summarization. The experiment also does not dismiss a context-dependent processing of images which could form the basis of advanced image processing, image understanding and computer vision systems. At the same time, this confirms that context-dependent processing is a more general, more powerful, information processing framework situated at a higher level of the Chomsky hierarchy [16] than context-independent (i.e., context free) processing:

- regular grammars (finite state automata)
- context-free grammars (push-down automata)
- context-sensitive grammars (linear bounded automata)
- unrestricted transformational grammars (universal Turing machine)

This also allows a new formulation of the usability axiom: medical information technology must attain human-like, adaptive, *context-dependent information*

processing functions (i.e., knowledge processing) using existing, largely *context-independent information processing* models. This new vantage point allows us to speculate on a possible solution: human-like, adaptive, context-dependent information processing could theoretically be approximated with arbitrary precision by linking sufficiently many, simple, context-independent processing models into complex architectures. If this is true, then the most important implication is that intelligent, context-dependent information processing is just a quantitative aspect related to the amount of memory needed in order to store the sufficiently many, general, context-independent functions whose operational whole could yield more advanced, adaptive, context-dependent processing.

Algorithmic Information Theory

An additional indication for the importance of context-dependent information processing comes from the relatively new field of research called Algorithmic Information Theory (AIT) [17]. AIT and the notion of Kolmogorov complexity unify the fields of computer science and information theory. It is said [18] that rather than focusing on ensembles and probability distributions as in classical statistics and information theory [19], AIT focuses on the algorithmic properties of individual objects. In order to explain this statement better, we follow with an example adapted from [17].

It is known that the classical probability framework generally aims at representing the reality of probabilistic experiments, the prototype of which is the typical coin tossing. In such experiments, individual tosses are generally assumed independent of each other. This is equivalent to assuming context-independence of events, in which

previous tosses do not influence subsequent ones. For example an experiment involving ten consecutive tosses (T – tails, H – Heads) could result in outcomes such as HHHHHTTTTT, HTHTHTHTHT, TTTTTHHHHH, or THHTHTTHT all of which are assigned same probability $p=1/2^{10}$ under the classical statistics framework. In addition, because the ratio between the counts of heads and tails is exactly $\frac{1}{2}$, probability theory also implies that the experiment indeed consisted of truly random tosses of a fair coin. However, this is exactly where our biased, experience-laden intuition departs from the objectiveness of the probability theory. Not only that our perception of the probability of such regular sequences as HHHHHTTTTT or HTHTHTHTHT does not appear to be identical to that of THHTHTTHT, but most of us are likely thinking that there is something weird about a supposedly fair coin that yields HTHTHTHTHT in 10 consecutive random tosses. The only result that seems to appear legitimate is THHTHTTHT, a particular sequence of T's and H's that appears random, or patternless. This rationale forms the basis of the AIT framework laid down by A. Kolmogorov, G. Chaitin and R. Solomonoff [17].

What seems to stay at the crux of this difference between the objectiveness of the probability theory and our intuition is the fact that the number of sequences which exhibit regular structure (and hence low algorithmic complexity) is considerably smaller compared to those that appear random. This causes the former to be extremely unlikely to occur in random experiments and causes most of us to perceive such regular structures as unlikely to be the outcome of a random experiment. This perception bias exists despite the fact that, objectively, the probabilities of regular and random sequences of same length are identical and could very well serve as a possible

explanation for such trivial matters as to why people buy lottery tickets that seldom contain sequences such as 1, 2, 3, 4, 5, 6 for instance. This perception bias also seems to be continually reinforced by our reality which lacks discoveries of natural, lifeless objects that are perfectly regular (e.g., perfect square shape) and that are the result of random physical forces such as weather. However, more importantly, this same bias seems to suggest the fundamental mechanism that allows for high level cognitive functions in humans: experiential, context-dependent information processing which focuses only on regularity.

Because probabilities of events in classical statistics and information theory aim at objectivity and universal, context-independent applicability, they purposely disregard contextual information about the unfolding of experiments. For example, statistics focus only on the tally of heads and tails in the final state of the coin tossing experiment (i.e., after the 10 tosses). By contrast, if one chooses to account for the contextual information of an experiment, one is considered to be looking at the algorithmic properties of that experiment. Therefore, algorithms could be considered context-dependent spatio-temporal representation functions which describe events that unfold in a given reality.

In medicine, context-dependent representations are of high importance for efficient clinical reasoning. The potential to immediately recognize regular spatio-temporal patterns with low algorithmic complexity and which intuition tells are extremely unlikely to occur as the sole result of random processes, seems to be the mechanism that allows clinicians to decide whether certain sequences of events form causal

relationships that represent disease or not. It is also fair to say that nearly every medical event is usually represented, if not explicitly in medical records, at least implicitly put in a spatio-temporal context by the patient and/or by the diagnostician during history taking. This equates to a more complex, algorithmic representation of medical events and aligns well with another information processing paradigm, case-based reasoning (CBR) [5]. A method of individual knowledge processing that has originated in the dynamic memory models proposed by Schank [20, 21] and which aims at solving new problems based on the solutions to similar past problems, CBR focuses specifically on context-dependent descriptions of problem solving cases. Therefore, by using algorithmic representations of spatio-temporal events CBR has the potential to attain more useful representations of our reality and to provide solutions to the information technology paradox.

Toward context-dependent representations

So far, the importance of context-dependent representations might have become clear. However, it might also have become clear that context-dependent representation functions are more complex and hence more computationally expensive than context-independent ones. The nature of this complexity and some fundamental properties of such representations need to be explored. Theoretical and empirical work so far [5, 22, 23] has led us to believe that human-like, context-dependent representations could be regarded as working in *concept spaces* that are extremely *high dimensional* (thousands of dimensions) but *extremely sparsely populated*, containing only a tiny fraction of the theoretically possible number of entities. In addition, and especially relevant to Medical Informatics, such concept spaces are very *dynamic* (i.e., changing)

and necessarily exhibit a *similarity-based organization* in order to allow human-like context-dependent information processing. To illustrate some of their properties, we propose a line of reasoning which aims at setting upper bounds on certain properties of concept spaces. We call this argument “proof by resource exhaustion” because it makes use of existing, common sense bounds on existing temporal and/or spatial resources.

Let us apply this principle to a simple thought experiment. Consider the 32 by 32, black and white bitmaps in Figure 4, Figure 5 and Figure 6. Each of these bitmaps, if uncompressed, requires at least $32 \times 32 = 1024$ bits of data in order to be stored. Although crucial to defining the structure of a bitmap, the additional few bits of metadata, i.e., the width and heights and color depth of the bitmaps, are ignored in this discussion.

Each of the bitmaps in Figure 4 possesses a certain meaning to us and could be regarded as representations in a binary 1024-dimensional concept space. The bitmaps in Figure 5 have been created using pseudorandom noise with ratios between black and white pixels which allow the superposed meaningful image to still be distinguished from the noise. In this context, the following observations can be made:

- Computers are able to represent with extreme precision (i.e. without loss of any bit of data) all bitmaps in Figure 4, Figure 5 and Figure 6.
- Humans have high difficulties in representing precisely bitmaps such as in Figure 6 but can represent, however with various degrees of precision, the bitmaps in Figure 4 and Figure 5; the bitmaps in Figure 5 are represented with

little precision (i.e., high data loss) because of the noise is difficult to be retained as is;

- The representation functions required in order to assess similarities between bitmaps in Figure 4, or between bitmaps in tables Figure 4 and Figure 5 are most likely context-dependent and, by definition, their implementation is not trivial on existing computers;
- It is a trivial task for any human to assess the similarities between any of the bitmaps in Figure 4 and Figure 5; however the bitmaps in Figure 6 appear random and very similar to humans, while extremely dissimilar to trivial, context-independent, distance-calculation computer algorithms.

To generalize on these observations, computers are able to represent precisely any of the 2^{1024} possible bitmaps of dimensions 32 by 32, while humans can only represent, with various degrees of precision, a fraction of those bitmaps, i.e., those that represent something meaningful to us. Estimating the number of bitmaps meaningful to us (in effect estimating an upper bound for it) will demonstrate an important property of the concept space: that of extreme sparseness.

Let us suppose human memory was so dynamic and adaptive that humans were able to learn (i.e., acquire and represent) a completely new, 32 by 32 black and white bitmap every millisecond of their whole life. A human life could be safely considered to span a 100 years (a century). Because a year has 52 weeks, a century will have:

$$\begin{aligned}
& 1,000 \text{ msec/sec} \times 60 \text{ sec/min} \times 60 \text{ min/hour} \times 24 \text{ hours/day} \times 7 \text{ day/week} \times \\
& 52 \text{ week/year} \times 100 \text{ year/century} = 3,144,960,000,000 \text{ msec/century} \\
& = 3.14 \times 10^{12} \text{ msec/century} \cong 2^{42} \text{ msec/century}
\end{aligned}$$

So, under these extreme assumptions, the upper bound of the number of potentially meaningful 32 by 32 black and white bitmaps is about 2^{42} . Therefore, the fraction of all 32 by 32 black and white bitmaps that could, under these extreme assumptions, ever be meaningful to humans is about $2^{42}/2^{1024} = 1/2^{1024-42} = 1/2^{982}$, that is, a number with more than 290 decimal zeros.

In order to put this number ($1/2^{982}$) in perspective let us consider the number of atoms on earth. If the earth were made only of hydrogen atoms (the lightest atom possible) our calculation will overestimate but at the same time provide a theoretical upper bound on the number of atoms on earth. The atomic mass of hydrogen is about 1, meaning that 6.02×10^{23} (Avogadro's number) atoms have a mass of about 1 gram. Therefore, a hydrogen atom weighs about $1/6.02 \times 10^{23} = 0.166 \times 10^{-23}$ g/atom. The mass of the earth is estimated to be about 6×10^{27} grams. So the number of atoms in the earth can be no more than

$$\frac{6 \times 10^{27} \text{ g/earth}}{0.166 \times 10^{-23} \text{ g/atom}} = 36.1 \times 10^{50} \text{ atoms/earth} \cong 2^{5.18} \times (2^{3.32})^{50} = 2^{5.18+3.32 \times 50} \cong 2^{171} \text{ atoms/earth}$$

Finally, it is estimated that there are about 10^{81} , that is about 2^{269} atoms in the whole universe. So if we could label each atom in our universe with a distinct 32 by 32 black

and white bitmap, then the number of universes needed to make use of all possible bitmaps would be the enormous number of $2^{1024}/2^{269} = 2^{755}$.

The argument has estimated an upper bound on the number of meaningful entities in a concept space of 32 by 32 black and white bitmaps and, at the same time, hinted at the infinitesimally minute ratio between this bound and the total number of items that could theoretically fill that space. Yet the argument, while extreme in the estimation of the learning capability of human brains (i.e., one bitmap every millisecond of a 100 year human life) is very conservative with regard to the dimensionality of the concept space. Many real world problems and applications involve much richer representations than rudimentary 32 by 32 black and white bitmaps. This furthers our point and demonstrates unequivocally the extreme sparseness of the concept spaces in which human information processing operates. In addition, the argument also sets a limit on the number of training examples in order to learn and develop advanced, context-dependent processing functions. But this is not to say that processing is limited to the training examples. In fact, because human associative processing is naturally trying to seek and converge towards the learned patterns, i.e., the ones that are repeated many times and develop a meaning to us even if superposed with noise such as in Figure 5, processing can robustly generalize upon any possible input and is able to overcome much of the noise and randomness that are ubiquitous in our reality.

In general, such arguments “by resource exhaustion” are easy to construct and are often used in popular science publications. Mainstream literature and university course textbooks lack them probably due to their trivialized nature. However, the

insights generated by them cannot be underestimated. A similar argument is presented in [22] (page 53) in order to describe what Kanerva refers to as the “foremost problem” of human memory which is capable of easily operating in extremely high dimensional feature spaces unlike any known computer technology:

“There is no way to construct a random-access memory that has, say, 2^{1000} storage locations. Even 2^{100} locations would be too many to fit into the human brain, as 2^{100} molecules of water would more than fill it (the number of neurons in the nervous system is “only” about 2^{36}). With such a vast address space, most of the addresses cannot be represented by an address decoder and a storage location. However, there is hardly the need for 2^{1000} locations, because a human lifetime is too short to store anywhere near 2^{1000} independent entities (a century has fewer than 2^{32} seconds).”

To summarize, the human brain can easily perform pattern matching upon sets of items whose feature space can easily approach and even go beyond thousands of dimensions. We are able to discover and learn repeating environmental regularities (i.e., patterns) and retain them in a highly biased (or relevance based), context-dependent manner. At the expense of precision in representations, we are able to overcome randomness, to detect information inconsistencies and to master some of our complex reality by being able to predict outcomes of complex events in real time. Computers, on the other hand, are precise and powerful data storage and communication devices, but their power is misdirected toward being able to represent *any conceivable piece of data*, regardless whether that data makes sense or not, whether it appears random or not, or whether it would ever represent any reality during the entire life of our universe estimated to be around $2^{3.58 \times 9} = 2^{32}$ years (12 billion years). Currently, mainstream computer technology is limited when it comes to discovering and learning complex environmental regularities and predicting outcomes of complex events. To our frustration, computers are largely unable to process

information contextually (e.g., distinguish between “form” and “from” in spell checking), are very sensitive to noise (e.g., misspellings), and are unable to judge the consistency of information in order to take on some of the human knowledge acquisition and processing burden.

Advancing towards context-dependent representations could be recast as a problem of efficiently managing concept spaces characterised by *high dimensionality*, *extreme sparseness*, *dynamicity* and a *similarity-based organization*. Existing computational models [22, 24] suggest that one of the most efficient ways (if not the only one) of dealing with the high dimensionality, extreme sparseness, and dynamicity of representation is by linking entities explicitly in richly connected structures (i.e., graphs such as in Figure 7). Specifically, the computational models that form the focus of our current research are based on partial order sets or *posets*. Formally, a poset is a base set together with a reflexive, antisymmetric and transitive binary relation on that base set. Posets are also generalizations of trees and can be depicted using Hasse diagrams such as in Figure 7. Informally, posets could be thought of as generalizations of ordered sets in that they allow elements between which ordering relations do not exist (e.g., “c” and “d” or “abc” and “bcd” in Figure 7a and “10” and “011” in Figure 7b). Because posets are a more general way of representing information they could form the basis of more general information processing functions, however, at the expense of computing power.

Imposing similarity-based principles on their design where similar elements are close (e.g., “cd” and “c” or “abcd” and “abc” in Figure 7a), effectively turns such structures

into associative representations (or associative memories) that could grant access to similarity-based retrieval as well as to other high level information processing functions such as unsupervised pattern discovery and acquisition [25]. Because such approaches will allow us to discover and deal with new concepts which are defined through the representations of their contexts, we believe that such approaches, better regarded as context-dependent information processing or knowledge processing, hold the key to solutions for the medical information technology paradox. Context dependent information processing applications will therefore fulfil a longstanding desideratum for processing information on conceptual principles. For example:

- Case based reasoning (CBR), an area of research whose relevance to knowledge intensive problem solving such as medical decision making is already demonstrated [5], could be improved by advances in associative representations and similarity based retrieval;
- User interfaces of complex medical information systems (e.g., diagnosis coding) could be dynamically simplified (e.g., long scroll down lists shortened) based on the context-dependent relevance of their content to the situation at hand; context could be assessed through recognition of patterns of sensor data, diagnoses, symptoms, medication etc;
- Improving access control to personal and general medical information could be achieved through processing on conceptual principles the contextual knowledge about users;
- Improving the control of delivery of medical alerts could be based on contextual cues such as the recognition of workflow patterns and of the importance and of the relevance of a trigger to a certain context;

- The robustness of alternative user input interfaces such as handwriting, gesture and speech recognition, natural language processing, image and scene recognition could be improved by context dependent processing of sensor data;
- Indexing and retrieval of information on conceptual principles could lead to advances in automated searches for literature relevant to a certain subject matter, a possibility which has the potential to revolutionize information retrieval.

Finally, we believe that difficulties to implement such approaches arise solely from their inherent complexity and from the processing power and especially the amount of random access memory they require in order to overcome the stage of “toy applications.”

Conclusions

In this article, we have defined the usability axiom and the paradox of medical information technology and reformulated them, several times, from more fundamental perspectives such as that of representation. We have defined the concept of representation and demonstrated the importance of context-dependent information processing. We have introduced and defined the notion of concept space and demonstrated the extreme nature of two of its properties: high dimensionality and sparseness. We have also mentioned two additional properties, that of dynamicity and similarity-based organization and underlined their importance to context-dependent processing in general and Medical Informatics in particular. This led to concluding that context-dependent processing through efficient management of concept spaces

could provide a solution to the information technology paradox by attaining a higher level of information processing, better described as knowledge processing. In effect, this might have helped us define knowledge processing as context-dependent information processing. Finally, existing perspectives on the fundamentals of Medical Informatics [26] which place medical knowledge processing at the core of our discipline, make it possible to define Medical Informatics itself as *context-dependent medical information processing*.

Acknowledgements

We would like to thank an anonymous reviewer for suggestions and to acknowledge the useful discussions with Dr. A. Gulliver on Algorithmic Information Theory and with Dr. M. Tara on various aspects of this article.

References

1. Koppel, R., et al., *Role of Computerized Physician Order Entry Systems in Facilitating Medication Errors*. JAMA, 2005. **293**(10): p. 1197-1203.
2. Feigenbaum, E.A., *Some challenges and grand challenges for computational intelligence*. Journal of the ACM (JACM), 2003. **50**(1): p. 32-40.
3. Pylyshyn, Z.W., *The Robot's dilemma: the frame problem in artificial intelligence*. Theoretical issues in cognitive science. 1987, Norwood, N.J.: Ablex. xi, 156.
4. Dennett, D., *Cognitive Wheels: The Frame Problem in AI*, in *Minds, Machines, and Evolution*, C. Hookway, Editor. 1984, Cambridge University Press. p. 128-151.
5. Pantazi, S.V., J.F. Arocha, and J.R. Moehr, *Case-based Medical Informatics*. BMC Journal of Medical Informatics and Decision Making, 2004. **4**(1).
6. Moehr, J.R., *Health Informatics - A Scientific Challenge?* Meth. Inform. Med., 1994. **33**: p. 250-253.
7. Solomon, W.D., et al. *Having our cake and eating it too: How the GALEN Intermediate Representation reconciles internal complexity with users requirements for appropriateness and simplicity*. in *Proc. AMIA Symp 2000*. 2000. p. 819-823.
8. Szolovits, P., *Artificial intelligence in medicine*. AAAS selected symposium; 51. 1982, Boulder, CO: Westview Press. 226.

9. Grant, A., et al., *An informatics perspective on decision support and the process of decision-making in health care*, in *Using Knowledge and Evidence in Health Care*, L. Lemieux-Charles and F. Champagne, Editors. 2004, University of Toronto Press: Toronto.
10. Miller, R.A. and F.E. Masarie, *The demise of the 'Greek Oracle' model for medical diagnostic systems*. *Methods of Information in Medicine*, 1990. **29**: p. 1-2.
11. Brusilovsky, P. and M.T. Maybury, *From adaptive hypermedia to the adaptive Web*. *Communications of the ACM*, 2002. **45**(5): p. 31-33.
12. Miller, G., *WordNet: A Lexical Database for English*. *Communications of the ACM*, 1995. **38**(11): p. 49-51.
13. Blois, M.S., *Information and medicine: the nature of medical descriptions*. 1984, Berkeley: University of California Press. xiv, 298.
14. Weisstein, E.W., *From MathWorld--A Wolfram Web Resource*. <http://mathworld.wolfram.com/>. Oct 2005.
15. Schank, R.C., *Conceptual Dependency: A Theory of Natural Language Understanding*. *Cognitive Psychology*, 1972. **3**: p. 552-631.
16. Johnson-Laird, P.N., *Human and machine thinking*. John M. MacEachran memorial lecture series; 1991. 1993, Hillsdale, N.J.: Lawrence Erlbaum Associates. xix, 189.
17. Li, M. and P.M.B. Vitányi, *An introduction to Kolmogorov complexity and its applications*. 2nd ed. 1997, New York: Springer. xx, 637.
18. Chaitin, G.J., *Gödel's Theorem and Information*. *International Journal of Theoretical Physics*, 1982. **22**: p. 941-954.
19. Shannon, C.E., *A Mathematical Theory of Communication*. *The Bell System Technical Journal*, 1948. **Vol. 27**: p. pp. 379-423.
20. Schank, R.C., *Dynamic memory: a theory of reminding and learning in computers and people*. 1982, Cambridge [Cambridgeshire] New York: Cambridge University Press. xv, 234.
21. Schank, R.C. and R.P. Abelson, *Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures*. 1977, Hillsdale, N.J.: Erlbaum. 248.
22. Kanerva, P., *Sparse distributed memory*. 1988, Cambridge, Mass.: MIT Press. xxii, 155.
23. Pantazi, S.V. and J.R. Moehr, *Automated Knowledge Acquisition by Inductive Generalization*, in *e-Health 2004*. 2004: Victoria, BC, Canada.
24. Kohonen, T., *Self-Organizing Maps*. 3rd ed. Springer series in information sciences, ed. T.S. Huang. Vol. 30. 2001, Berlin Heidelberg New York: Springer-Verlag. 501.
25. Pantazi, S.V. and J.R. Moehr. *An associative memory model for unsupervised sequence processing*. in *IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM2005)*. 2005. Victoria, BC, Canada. p. 233-236.
26. Musen, M.A., *Medical informatics: searching for underlying components*. *Methods Inf Med*, 2002. **41**(1): p. 12-19.

Figure 1. The relationship between usability and problem complexity and the role of artificial intelligence. Approaches that are highly usable (e.g., calculator) tend to solve less complex problems while approaches to solve highly complex problems are often considered “less usable” (e.g., computer programming).

Figure 2. Possible representations of a real, actual telephone set. a, b, c are bitmaps of images, d is a bitmap of written symbols and e is the sequence of ASCII hexadecimal codes corresponding to characters “P”, “H”, “O”, “N” and “E”.

Figure 3. Example output of a simple representation function

Figure 4. Bitmaps (32 by 32 pixels) that are meaningful to humans

Figure 5. Bitmaps (32 by 32 pixels) that are noisy but still meaningful to humans

Figure 6. Bitmaps (32 by 32 pixels) that are only noisy and not meaningful to humans

Figure 7. Examples of partial order sets (posets)

Table 1. Selected senses out of the distinct ten senses of the noun “representation” in WordNet, in the decreasing order of their estimated usage

Table 2. Selected senses out of the distinct fifteen senses of the verb “to represent” in WordNet, in the decreasing order of their estimated usage

Table 3. Two groups of lexical items belonging to the semantic space of the concept of representation in the syntactic role of noun (i.e., representation) and verb (i.e., to represent)

Figure 1

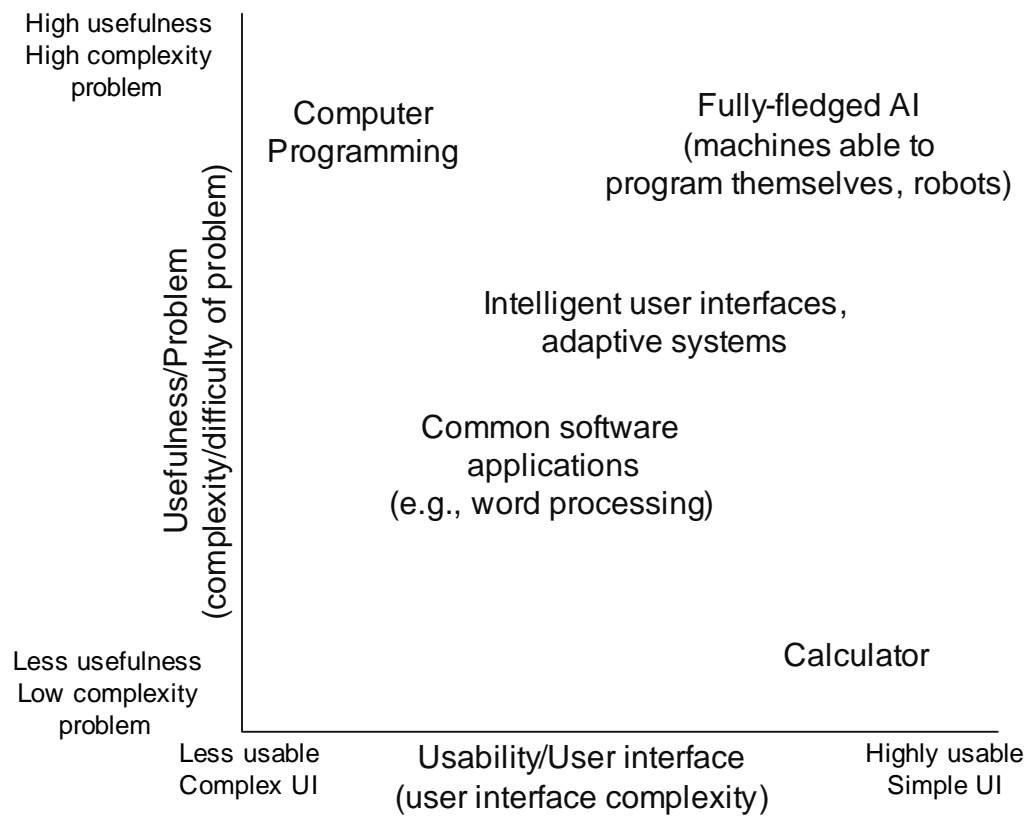


Figure 2
[Click here to download high resolution image](#)



Figure 3
[Click here to download high resolution image](#)

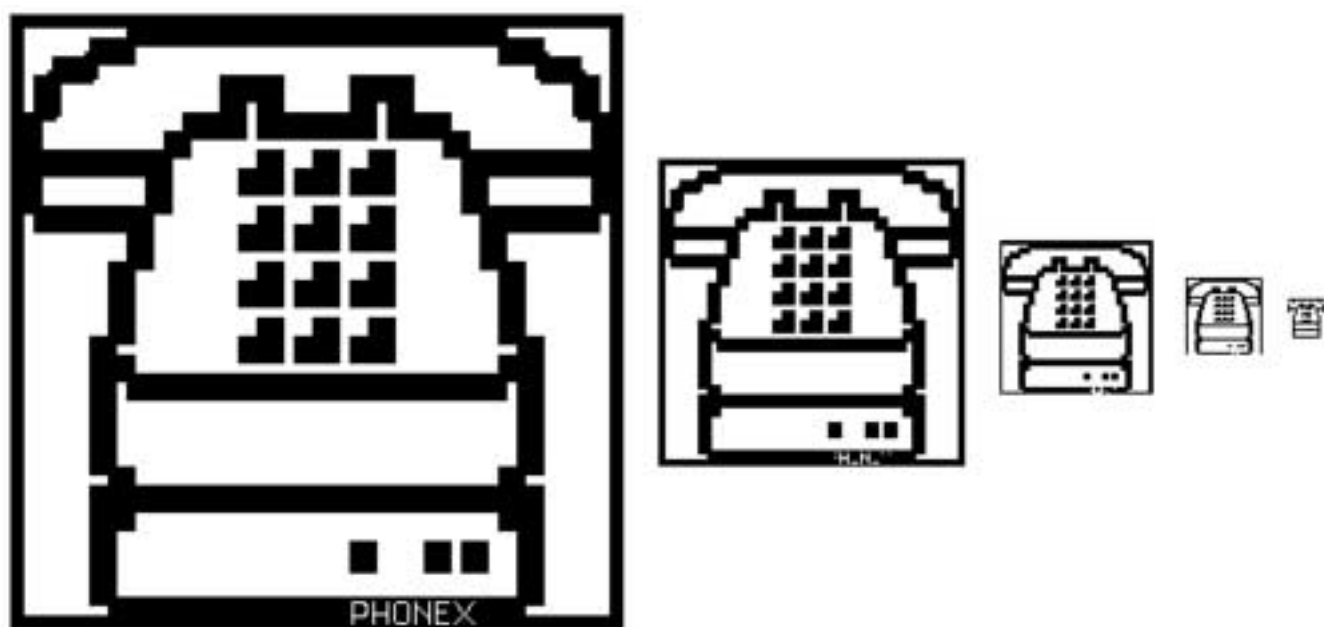


Figure 4

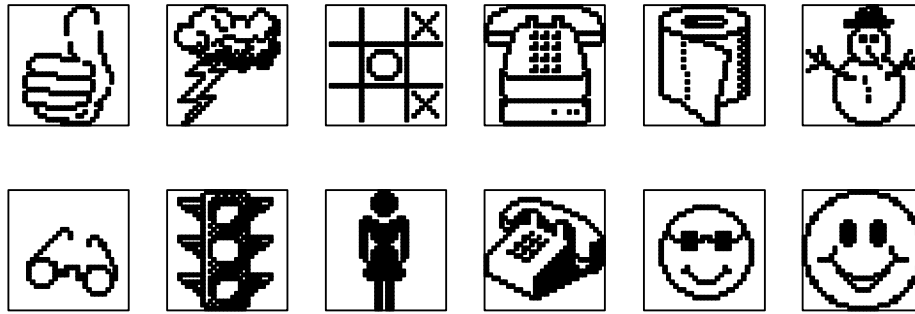


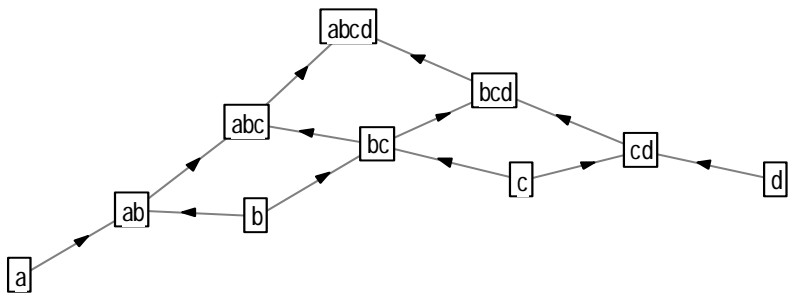
Figure 5



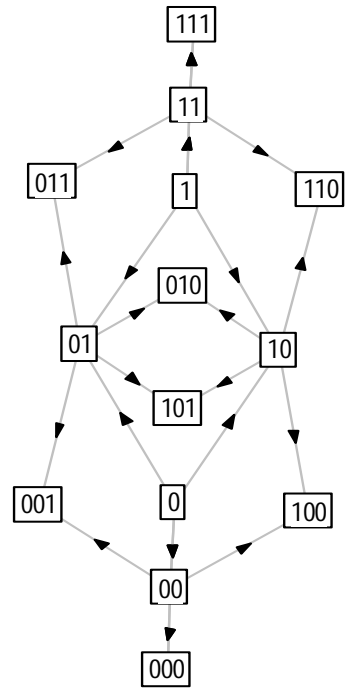
Figure 6



Figure 7



a



b

Table I

Sense #	Lemma(s)	Definitions and examples
1	representation, mental representation, internal representation	a presentation to the mind in the form of an idea or image
2	representation	a creation that is a visual or tangible rendering of someone or something
3	representation	the act of representing; standing in for someone or some group and speaking with authority in their behalf
4	representation, delegacy, agency	the state of serving as an official and authorized delegate or agent
5	representation	a body of legislators that serve on behalf of some constituency; "a Congressional vacancy occurred in the representation from California"
6	representation	a factual statement made by one party in order to induce another party to enter into a contract; "the sales contract contains several representations by the vendor"
7	theatrical performance, theatrical, representation, histrionics	a performance of play
8	representation	a statement of facts and reasons made in appealing or protesting; "certain representations were made concerning police brutality"
9	representation	the right of being representated by delegates who have a voice in some legislative body
10	representation	an activity that stands as an equivalent of something or results in an equivalent

Table II

Sense #	Lemma(s)	Definitions and examples
1	represent, stand for, correspond	take the place of or be parallel or equivalent to; "Because of the sound changes in the course of history, an 'h' in Greek stands for an 's' in Latin"
2	typify, symbolize, symbolise, stand for, represent	express indirectly by an image, form, or model; be a symbol; "What does the Statue of Liberty symbolize?"
3	represent	be representative or typical for; "This period is represented by Beethoven"
4	represent	be a delegate or spokesperson for; represent somebody's interest or be a proxy or substitute for, as of politicians and office holders representing their constituents, or of a tenant representing other tenants in a housing dispute; "I represent the silent majority"; "This actor is a spokesperson for the National Rifle Association"
5	represent	serve as a means of expressing something; "The flower represents a young girl"
6	exemplify, represent	be characteristic of; "This compositional style is exemplified by this fugue"
7	constitute, represent, make up, comprise, be	form or compose; "This money is my only income"; "The stone wall was the backdrop for the performance"; "These constitute my entire belonging"; "The children made up the chorus"; "This sum represents my entire income for a year"; "These few men comprise his entire army"
8	defend, represent	be the defense counsel for someone in a trial; "Ms. Smith will represent the defendant"
9	represent, interpret	create an image or likeness of; "The painter represented his wife as a young girl"
10	act, play, represent	play a role or part; "Gielgud played Hamlet"; "She wants to act Lady Macbeth, but she is too young for the role"; "She played the servant to her husband's master"
11	stage, present, represent	perform (a play), especially on a stage; "we are going to stage 'Othello'"
12	represent	describe or present, usually with respect to a particular quality; "He represented this book as an example of the Russian 19th century novel"
13	represent	point out or draw attention to in protest or remonstrance; "our parents represented to us the need for more caution"
14	present, represent, lay out	bring forward and present to the mind; "We presented the arguments to him"; "We cannot represent this knowledge to our formal reason"
15	map, represent	to establish a mapping (of mathematical elements or sets)

Table III

Category	Lexical items
REPRESENTATION (nouns)	representation, image, form, model, simulation, map, stereotype, schema, perception, memory, example, adumbration, copy, replica, reproduction, carbon copy, anamorphism, cutaway, display, document, drawing,
X is a representation	ecce homo, effigy, illustration, nomogram, objectification, picture, icon, projection, rubbing, cast, duplicate, triplicate, facsimile, imitation, counterfeit, forgery, knockoff, clone, miniature, toy, photocopy, print, Xerox, transcript
REPRESENT (verbs)	represent, stands for, correspond, take the place of or be parallel or equivalent to, typify, symbolize, express indirectly, be representative or typical for, serve as a means of expressing, exemplify, be characteristic of,
to X is to represent	map, imitate, simulate, recreate, replicate, reproduce, duplicate, double, clone, repeat, photocopy

Revision notes

The revision notes references to page numbers refer to the original manuscript submission and may not be accurate in the generated pdf.

1)

We performed minor changes (shuffling word order, removing a redundant sentence) in the Introduction section, on page 9, in the paragraph starting with “In Figure 1 we illustrate a hypothesized relationship between usability....”

2)

We added the following new sentence:

“At the same time, the context dependence of medical data seems to contradict the existing biomedical research paradigms such as populational studies (e.g., randomized control trials) which purposely aim at removing the context around medical data in order to attain population-wide applicability of their findings.”

in the “What does “to represent” mean?” section, on page 13, after the paragraph ending in “...processing of health data is caused precisely by its high context sensitivity [...] (p.251). “ and before the paragraph starting with “This very important context perspective allows us to distinguish two extreme type...”

3)

We addressed the reviewer’s comment with regard to the reference to Schank and Abelson’s work which indeed is very relevant to this article, from many perspectives. We include the reviewer’s comments:

As reviewer I missed a reference of the work of R. Schank and R. Abelson at page 20 at which page the authors introduce the conceptual nature of textual data. Schank and Abelson discussed the concepts of scripts as a form of generalized episodic knowledge which provides a chain of events for common activities. Many causal relations in natural language texts are not direct and instead depend on inferred intermediate events. Scripts present a formalism for representing causal interactions between events. Since the authors do introduce the work of Chomsky, the work of Schank and Abelson deserves mentioning too.

As suggested, we added in the section “Context-dependent information processing “ at page 17, the reference

Schank, R.C., Conceptual Dependency: A Theory of Natural Language Understanding. Cognitive Psychology, 1972. 3: p. 552-631,

which fits well in the context of conceptual nature of textual data.

We have also added the reference to the 1977 volume by Schank and Abelson

Schank, R.C. and R.P. Abelson, Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures. 1977, Hillsdale, N.J.: Erlbaum. 248.

not in the location suggested by the reviewer but in the context of the Case Based Reasoning under the section “Algorithmic Information Theory”, on page 21, together with an additional reference by Schank,

Schank, R.C., Dynamic memory: a theory of reminding and learning in computers and people. 1982, Cambridge [Cambridgeshire] New York: Cambridge University Press. xv, 234.

which is often cited as the a precursor of Case Based Reasoning.

We also took the opportunity to rework the section “Algorithmic Information Theory” on pages 19-21 and to clarify some delicate points that have to do with the distinction between classical information theory and Algorithmic Information Theory. In this context we emphasized on the fact that while classical statistics and information theory are universally applicable, objectively correct and necessarily context-independent, they do not completely account for what happens with human perception in a context dependent information processing paradigm. We hope that now it is clearer that the two theories complement each other rather than invalidating one another.

4)

We have tried to address the second suggestion:

A last addition to the Discussion seems requires: In the Introduction section the emphasis is on (the tense between) usability and usefulness of information systems. The authors do not explain at the end of the paper how context dependent information processing (similarity based principles, unsupervised pattern recognition and acquisition) may in the end result systems that are both useful and usable. In other words, how these methods may influence information systems' user interfaces (that are usable) is not explicitly described.

We attempted to address the comment by providing, at the end of the discussion on pages 28-29, a non-exhaustive list with some possible applications of context dependent information processing.

5)

We modified figure 3 in order to depict more accurately the idea of the output of a simple, context-free representation function and to comply with the artwork guidelines of the editorial system.