
This is the unspecified version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: http://openaccess.city.ac.uk/3117/

Link to published version:

Copyright and reuse: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.
The nature of information science; changing models

Lyn Robinson  
Department of Information Science  
City University London, United Kingdom  
lyn@soi.city.ac.uk

Murat Karamuftuoglu  
Department of Computer Engineering and Department of Communication and Design  
Bilkent University, Ankara, Turkey  
hmk@bilkent.edu.tr

Abstract  
This paper, based on a presentation at the 7th conference on Conceptions of Library and Information Science, considers the nature of information science as a discipline and profession. It is argued that information science may be understood as a field of study, with human recorded information as its concern, focusing on the components of the information chain, studied through the perspective of domain analysis, in specific or general contexts. A particular aspect of interest is those aspects of information organisation, and of human information-related behaviour, which are invariant to changes in technology. Information science can also be seen as a science of evaluation of information, understood as semantic content with respect to qualitative growth of knowledge, and change in knowledge structures in domains.
Introduction; the debatable nature of information science

Debates about the nature of information science, the scope of the discipline, and its relations to other academic and professional areas, are as old as the discipline itself. These are not merely navel-gazing, or arguments about terminology. They relate to the validity and viability of the discipline, and have significance for the extent to which its unique contributions are recognised.

Information science first became known as a discipline during the 1950s. The first usage of the term in a paper by Jason Farradane (1955), in which he stated that contemporary British academic and professional qualifications were “... a pattern for establishing qualifications in documentation, or 'information science'”, following from earlier uses by Farradane of the term ‘information scientist’, to mean initially a specialist in the handling of scientific and technical information (Shapiro 1995, Robinson 2009). The discipline grew out of the longer-standing 'documentation' movement, under numerous social, economic and technical influences; see Robinson (2009) for a summary of the literature describing these origins.

It is clear that, from the origins of the terms, there has been little agreement about the nature of information science, and indeed information scientists (Shapiro 1995, Bawden 2008, Robinson 2009). Was the concern with the 'information of science', i.e. the practicalities of the handling of scientific and technical information, or with the 'science of information', i.e. the academic study of information phenomena? This question has never really been settled; in essence, it is the question of whether information science is a discipline, or a practical art.

Heilprin wrote in 1989 that "although many laws, hypotheses, and speculations about information have been proposed, adequate scientific and epistemic foundations for a general science of information have not yet appeared". Nearly twenty years later Zins (2007) concluded that "Apparently, there is not a uniform conception of information science. The field seems to follow different approaches and traditions: for example, objective approaches versus cognitive approaches, and the library tradition versus the documentation tradition versus the computation tradition. The concept has different meanings, which imply different knowledge domains. Different knowledge domains imply different fields. Nevertheless, all of them are represented by the same name, information science. No wonder that scholars, practitioners and students are confused."

There has, of course, been much debate about what kind of discipline information science is; for overviews of the issues see Hawkins (2001), Webber (2003) and Robinson (2009). It has variously been claimed as a social science, a meta-science, an inter-science, a post-modern science, an interface science, a superior science, a rhetorical science, a nomad science, a liberal art, a knowledge science, and a multidisciplinary field of study (Robinson 2009).

A further unsolved issue is the relationship between information and other academic and professional disciplines. One area of debate has been the relation with 'adjacent' disciplines such as librarianship, archiving, information systems and computer science; views here have ranged from such disciplines being the same thing, entirely distinct, distinct but inter-dependent, distinct but naturally linked, and part of a composite discipline. This debate also manifests in the question of whether there is any meaningful link between the concept of information in different domains; those who see the possibility of such a link include Bates (2006) and Bawden (2007), while those who reject it include Hjørland (2007).

We must therefore conclude that there is still no agreement about some of the basic aspects of the information science discipline. This matters since this lack of agreement as to what the discipline is about leads inevitable to a difficulty in explaining what its value and benefits may be. As Dillon (2007) reminds us, although the questions central to library / information science are of great interest to society, the answers are not usually sought from the LIS community.

One approach to overcome this problem is to attempt an understanding of information science in terms of two well-established concepts within the field: the communication chain and domain analysis.
Domain analysis and communication chain

This approach, by which the discipline of information science is located in the examination of the information chain through the methods of domain analysis, was put forward by Robinson (2009).

Many accounts and explanations of information science from the 1960s onwards have focused on the idea of the ‘information chain’ or ‘information life cycle’; the sequence of processes by which recorded information, in the form of documents, is communicated from author to user (Robinson 2009, Webber 2003, Summers, Oppenheim, Meadows, and Kinnell 1999). ‘Documents' may be understood broadly, to include entities other than conventional written documents (Buckland 1997, Frohmann 2009, Turner and Allen 2010). As Meadows (1991) points out, there are a number of variations of this chain, according to the type of information and information-bearing entities involved; the nature of the chain has changed considerably over time, largely under the influence of new technologies (Duff 1997). But typically the chain has been described as having components such as: creation – dissemination – organisation – indexing – storage – use. Zins’ (2007) recent Delphi study shows that, to a large extent, perceptions of information science still revolve around these concepts.

The communication chain by itself may reasonably be seen as too restricted a focus for the discipline. Our viewpoint therefore complements it by a framework for studying it, and for improving its effectiveness in practice: domain analysis.

Domain analysis, in the sense in which the term is used here, was introduced by Birger Hjørland, who regards it as encapsulating the unique competences of the information specialist (Hjørland 2002, Hjørland and Albrechtsen 1995).

Hjørland sees domain analysis, as practised by the information scientist, as comprising eleven distinct approaches, any of which may be used to help to understand the information of a domain. These approaches are:

- production of literature guides and subject gateways
- production of special classifications and thesauri
- research on indexing and retrieval in specialist subjects
- empirical user studies
- bibliometric studies
- historical studies
- studies of documents and ‘genres’
- epistemological and critical studies
- studies of terminology and special languages, discourse studies
- studies of structures and organisations in the communication of information
- studies in cognition, computing and artificial intelligence

The domain analysis concept has been extended by several writers, building on Hjørland’s ideas to extend and clarify the concept of ‘domain’, to introduce new aspects, as to extend the
range of areas of applicability of the concept; see, for example, Tennis (2003), Hjørland and Hartel (2003), Feinberg (2007), Sundin (2003), Hartel (2003) and Karamuftuoglu (2006).

This leads us to a simple conceptual model for the information science discipline: the six-component information chain as the focus of interest, examined by the eleven approaches of domain analysis. Some of the approaches will ‘fit’ clearly with certain components – the production of special classifications with indexing, for example – but, in principle, any component/s of the chain may be studied by any approach/es. This leads to a to a three level model, able to describe any topics within information science (Robinson 2009). It involves defining a ‘context’ in terms of scale and media involved, thus:

- Component (of chain)
- Approach (of domain analysis)
- Context (scale / media)

A study of, for example, use of social network resources by historians might be described as:

- component: use
- approach: empirical user studies
- context: group (history discipline) / social networking

This model also provides a way of showing the way in which related disciplines are linked: through the appropriate domain analysis approach, from one or more components of the chain in the appropriate context.

Computer science, for example, is seen to be linked primarily through the ‘indexing and retrieving’ approach, via the overlap area of information retrieval, which may be argued to belong to both disciplines. Its artificial intelligence aspects (not always regarded as part of computer science proper) are linked through the ‘professional cognition and artificial intelligence’ approach. Robinson (2009) gives other examples, showing the advantage of this perspective in explaining the validity, and nature, of relations and overlaps between information science and other disciplines.

This model gives a clear picture of the nature of information science, and a way of understanding its relations to other disciplines. However it does not explicitly identify those topics, which are entirely unique to information science.

**A unique academic territory?**

We may argue that the uniqueness of information science lies in the focus on the combination of the information chain and domain analysis. Though other disciplines and professions may be involved in components of the chain (publishers in dissemination, computer scientists in retrieval, etc.) and in aspects of domain analysis (philosophers in epistemology, historians in historical studies, and so on), only information science is interested in the totality of intersection of the totality of the two concepts, and in all the various uses of information (Kari 2010). The information scientist therefore has a uniquely generalist approach to all aspects of the communication of information.

We may find a more precisely stated ‘uniqueness’ in the ideas propounded over many years by Brian Vickery (see, for example, Vickery's last (2009) article, and an informal summary in Bawden (2010)). Regardless of advances in technology, Vickery insists there are some fundamentals of human information-related behaviour and of the organisation of information, which do not change. This is not to say, of course, that information behaviour and information organisation do not change in new technological environments; rather that, at a deeper level, consistent explanatory factors may be found. It is the business of the information scientist to investigate these, and to show their relevance in whatever information environment they may be instantiated. It is the business of information science to investigate ‘technology invariant’ patterns of human information behaviour and issues of information organization, and to apply the findings to the design of systems and services. The area of unique interest to information science is therefore to be found within this part of the intersection of domain analysis and the communication chain.
This gives us an understanding of information science as a very 'real' discipline, with its own academic and professional scope. But to find a further conceptual basis for the discipline, we need to consider the only quantitative theory of information, that of Shannon and Weaver, and extensions from it. We shall find, paradoxically, that this leads us to an appreciation of information as concerned with qualitative changes in knowledge.

Quantitative Models of Information

We will argue that Information Science (IS) is set apart from other disciplines by its unique object of study, namely, "the problem of evaluating information understood as semantic content with respect to qualitative development of knowledge in a given domain". While semantic conceptions of information developed in the wake of Shannon's syntactic theory of information, and more generally computational approaches, study quantitative change, IS studies qualitative change, as every non-mechanical relevance judgment requires a qualitative leap. We will discuss each of the salient points in the above sentence, namely, "quantitative change", "qualitative change", "non-mechanical relevant judgement" and "qualitative leap" in detail below. However, before that we will review briefly Shannon's syntactic and Barwise's semantic theories of information.

Shannon's Mathematical Theory of Communication (MTC) is concerned with the transmission of information from a source to a receiver over a physical communication channel (Shannon and Weaver, 1949). The average amount of information, $H$, associated with a source, $S$, from which symbols are selected to compose a message, is given by:

$$H(S) = - \sum_{i} \text{Pr}(S_i) \log_2 \text{Pr}(S_i)$$

(1)

where, $\text{Pr}(S_i)$ is the probability of selection of a particular symbol, and $N$ is the number of unique symbols in $S$. For instance, for a source that has 8 distinct symbols with equal probabilities of selection, $N = 8$ and $\text{Pr}(S_i) = 1/8$. Information generated when a particular symbol is selected from a set of possible symbols is called self-information or surprisal, which measures the uncertainty associated with the selection of the symbol, and is given by:

$$I(S_i) = - \log_2 \text{Pr}(S_i)$$

(2)

For $N$ equiprobable symbols, equation (1) reduces to (2). For example, when $N = 8$, both $H(S)$ and $I(S)$ are equal to 3 bits.

MTC is a syntactic theory of information, as it is not concerned with the meaning of the symbols/messages transmitted but their quantity. In a system of two symbols ($N=2$), say head and tail of a coin, 1 bit of information is transmitted regardless of whether the head or tail of the coin symbolizes nuclear war or who is going to do the dishes.

MTC is rightly criticised for not being relevant to IS, as IS's main concern is the interpretation of documents, i.e., what documents are about or mean. Situation Theory (ST) developed by Barwise & Seligman (1997), Barwise & Perry (1983), Devlin (1991), and others attempt to provide a semantic theory of information, based loosely on the MTC paradigm.

---

1 We define information (after Floridi, 2005) in the context of this paper as “semantic content”, and more precisely, as “well-formed meaningful data”. Data, in turn, is defined as “lack of uniformity within some context”. Knowledge is sometimes defined in formal/mathematical approaches as “justified, true semantic information”. Although, this definition is useful in some contexts, it is not comprehensive enough. In a second sense, knowledge can be defined as “systematic understanding and integrated body of skills needed for performing successfully in a given domain”. The second sense has close ties to the concept of theory in science, and more useful for most IS purposes.
ST provides an ontology (objects, situations, channels, etc.) and a set of logical principles (inference rules) that operate on the objects and situations via channels. Channels are informational-links that model the semantic, conventional, casual, etc. relationships between objects. Rijsbergen & Lalmas (1996, pp. 391-392) give an example of a channel that models the synonymy relationship in a thesaurus. For instance, in the context of information retrieval, if a document contains the term “belief”, it can be deduced, using a thesaurus as a channel, that it also contains (implicitly) the term “dogma”, assuming these two terms are related in the thesaurus used as a channel.

Collectively, ontology and the set of inference rules determine the scope of deductions that can be made, and thus, the type of questions asked and answered about the state of affairs in a given situation. Changing the channel in ST amounts to changing the types of inferences made about entities/objects. The channel, thus, determines what can be known about a situation. For instance, consider the use of WordNet as a channel. Information science is related to computer science through the meronym relation in WordNet, that is, computer science is a part of information science. However, according to WordNet, documentation is not related to IS. In fact, it is not even recorded as a discipline in WordNet. However, for instance, Hjørland & Capurro (2003) take a view that documentation is an important part of, if not synonymous with, information science. Other authors take different views on the same issue.

This brief discussion illustrates that what a channel models in ST depends on the particular theoretical/epistemological position taken in constructing the ontology, which marks the limits of the usefulness of ST for IS. To put in other words, ST allows deductions once a model of the world is given in terms of objects and channels that represent the relationships between them. The main problem in ST is precisely the construction of the qualitative model of the world that provides the basis for drawing inferences. Once the model is given, quantitative inferences are relatively straightforward to compute. A more thorough discussion of ST and semantic information in the context of IS can be found in Karamuftuoglu (2009).

Quantitative and Qualitative Change

What is a qualitative change is a difficult question to answer rigorously. Intuitively, the term “qualitative” invokes the image of the creation of something new out of old where the steps involved in the transformation of the old into new are not obvious. The archetypal example is the transformation of larva into butterfly in the pupa or cocoon. This formulation is akin to the idea of “inventive step” or “non-obviousness” invoked in patent law in many countries.

A similar idea is found in computation theory (see e.g., Boolos et al., 2007). Informally speaking, an effective method is a method in which each step in it may be described as an explicit, definite, mechanical instruction, that always leads, when rigorously followed, to the right answer in a finite number of steps, ignoring physical limitations on time, speed, and storage. The essential feature of an effective method, like that of the inventive step in patent law, is that it does not require any ingenuity from the person or machine executing it. An effectively computable function is similarly defined, as a function for which there is a finite procedure, an algorithm, instructing explicitly how to compute it.

We, thus, define quantitative change as a process that leads from one state (old) to another (new) following an effective method. Inferences allowed in ST, and generally all deductive argumentation, are essentially effectively calculable. Deduction is a type of argumentation from general to particular. When the premises of a deductive argument are true, conclusions reached by it are guaranteed to be true. A complementary mode of reasoning found in

---

2 A situation could be seen as a collection of objects (entities) that corresponds to the limited parts of reality, for our purposes here.

3 http://wordnet.princeton.edu/
traditional logic is *induction*. Induction is a method of reasoning from particular to general, which produces only probable conclusions that need to be verified by future observations. The forms of these two modes of reasoning are given below.

**Deduction**

**Rule** – All the beans from this bag are white. [given]
**Case** – These beans are from this bag. [given]
**Result** – These beans are white. [concluded]

**Induction**

**Case** – These beans are from this bag. [given]
**Result** – These beans are white. [given]
**Rule** – All the beans from this bag are white. [concluded]

Symmetrically, we define *qualitative change* as a process where the transition from one state of the system to another, or one or more of the steps in the process, are not effectively calculable. Since, some of the steps in the process need to be “invented”, or require a “leap of faith”, so to speak, such processes are considered to be *discontinuous*, or involve *qualitative jumps*.

Our thesis is that in IS there are certain processes that involve qualitative changes, and judgements that require qualitative decisions. Specifically, as we will show next, relevance judgements on documents, and more generally, subject analysis and classification, require qualitative decisions and jumps.

Consider two alternative rules that specify the criteria of relevance of a document to a given query:

**Rule 1:** Documents that support measures taken to prevent the spread of the HIV virus are relevant.
**Rule 2:** Documents that support measures taken to prevent the spread of AIDS in terms of environmental and social factors are relevant.

There are two possible deductions regarding the relevance of a given document $X$, corresponding to the two rules specified above:

**Deduction A:**
**Rule** – Documents that support measures taken to prevent the spread of the HIV virus are relevant
**Case** – *Document X* does not support measures taken to prevent the spread of the HIV virus
**Result** – *Document X* is non-relevant

**Deduction B:**
**Rule** – Documents that support measures taken to prevent the spread of AIDS in terms of social and environmental factors are relevant
**Case** – *Document X* supports measures taken to prevent the spread of AIDS in terms of social/environmental factors
**Result** – *Document X* is relevant

As we have seen earlier, deductions are effectively calculable. Once, the rules are known, the relevance judgment is reduced to a mechanical inference. The question is, then, how one arrives at the rules of relevance. Arguably, the first rule is rather obvious to arrive at. Overwhelmingly, it is accepted that AIDS is caused by HIV. Therefore, a query such as “What is being done to help prevent the spread of AIDS in a certain country $Y$?” would invoke in most human judges or classifiers/analysts Rule A above, without the need for an inventive step or qualitative leap. The second rule, Rule B, arguably, is not so straightforward to conjecture, as the significance of social/environmental factors in AIDS is not universally recognised.
The second inference would, therefore, require a process akin to creation of a hypothesis in science. C.S. Peirce, prominent philosopher and semiotician, calls the process of creation of a hypothesis from incomplete evidence as abduction. Abduction is different from both deduction and induction in that neither the rule nor the case is given. The rule is hypothesised, and based on this hypothesis a case is concluded. Abduction is a creative process of hypothesis forming, in which, based on the relevant evidence, the hypothesis that best explains a given phenomenon is formulated. In Peirce's words: "Abduction is the process of forming explanatory hypothesis. It is the only logical operation which introduces any new idea" (Peirce, 1958, v 5, para. 171-172). This process is illustrated below.

Abduction (Hypothesis)

1. The surprising fact, F, is observed;
2. But if H were true, F would be a matter of course.
3. Hence, there is reason to suspect that H is true (Peirce, 1958, v 5, para.189).

When the abductive mode of reasoning is applied to document evaluation, the process is more complex than the above case where the surprising observation corresponds to a fact. In the case of document evaluation/subject analysis, it is not known in advance whether there exists a user community for whom the document is actually relevant. This is postulated. Based on this postulate and the content analysis that reveals that the document contains information about environmental and social issues in Y, it is hypothesized that there exist alternative theories of AIDS causation, which explain the epidemic in terms of conventional social and environmental factors.

In the aforementioned case of the relevance judgement on Document X, this takes the following form:

Abduction (to arrive the alternative Deduction B):

1. The surprising result, F (Document X is relevant to the topic "The spread of AIDS epidemic in Y"), is postulated.
2. But if H (there exists an alternative theory of AIDS causation, which explains the AIDS epidemic in terms of social and environmental factors) were true, F would be a matter of course (given that the document reports relevant information on the economic and social development of Y)
3. Hence, there is reason to suspect that H is true

The conclusions reached by abduction, unlike deduction, are always uncertain. The assessor, thus, needs to research to find whether the hypothesis is true, i.e., there really exist social/environmental explanations of AIDS. Failing to find one would make the document non-relevant. Conversely, discovery of a relevant theory would make the document relevant to a user group, who prioritizes such factors in explaining health problems over virus-based explanations.

One final question remains to be answered: how does in practice an assessor make a qualitative jump from the postulation of the potential relevance of a document to the rule that specifies the relevance criterion? The assessor's knowledge of the major philosophies and schools of thought would be an advantage in formulating a hypothesis in reasoning processes similar to abduction. It is unreasonable to expect from even professional classifiers/assessors to have a detailed understanding of all significant issues and theories in a domain such as medicine. It is arguable that even subject experts do not necessarily know everything relevant to a particular information need in the domain of their expertise. However, basic philosophy knowledge (knowledge in the second sense mentioned earlier), especially, epistemology, and methodology, as well as history of science, would arguably make the process of qualitative jump easier in such IS tasks as document classification and subject analysis.

We are now able to relate this idea back to the earlier material of this paper, by noting that the domain-analytic approach suggests that specific theories in a given field of study are built on more general metatheoretical frameworks (paradigms) or worldviews, which in turn are built on specific philosophical assumptions. Specific practices, tools and artefacts rest at the top of
the pyramid, partially or wholly based on specific domain-dependent theories. Figure 1 illustrates this idea.

A metatheory, or as sometimes called a paradigm, is essentially a set of principles that prescribes what is acceptable and unacceptble as theory in a scientific discipline. In epidemiology, for instance, one can distinguish between the dominant biomedical paradigm, and alternative emerging or past paradigms. The dominant biomedical paradigm focuses on the biology of disease. The dominant paradigm in epidemiology has been criticised for ignoring the contextual factors, such as level of social and economic development of a given society. The alternative paradigms in epidemiology prioritise general socio-economic and environmental factors in combating diseases and improving public health (see Karamuftuoglu, 2007, for a more detailed discussion of the above issues in the context of IS).

Kuhn's work on the history of science (Kuhn, 1962) shows that there is normally a single central paradigm, a single way of doing science, which he called "normal science", in established fields such as physics and astronomy. Before the establishment of a paradigm in a field of study, there is a period that Kuhn called "prescience", which is characterised by the existence of two or more alternative frameworks that compete to become the dominant paradigm. Similarly, at the moments of crises in "normal science", that is, when the central paradigm could no longer accommodate accumulating contradictory and conflicting results and observations, a number of alternative explanations of the anomalies compete. This is a period of "revolutionary science" according to Kuhn. When a new theory successfully resolves the anomalies, a "paradigm shift" happens, i.e., the old paradigm is replaced by the new. Science, thus, progresses discontinuously rather than in an orderly and continuous way.

Hence, it is arguable from a Kuhnian perspective that relevance assessment and classification of documents should be carried out in terms of the objective movement of competing theories and metatheories/paradigms in a domain. In other words, qualitative judgement of documents requires an understanding of the qualitative growth of knowledge, and change in knowledge structures in domains. This is another defining feature of information science.

![Figure 1: Domain-analytic view of a field of study](image-url)
Conclusions

As we have seen, there is no agreement on what the information science discipline is, or what its purpose and methods should be. The identity and relevance of information science is increasingly being challenged by developments in such diverse disciplines as philosophy of information, computer science, informatics, artificial intelligence and cognitive science. It is our contention that information science is set apart from all of the aforementioned disciplines and others, and that it is therefore very much 'real'.

We may best understand information science as a field of study, with human recorded information as its concern, focusing on the components of the information chain, studied through the perspective of domain analysis, in specific or general contexts. Its particular focus of interest is those aspects of information organisation, and of human information-related behaviour, which are invariant to changes in technology. It also has a role as a science of evaluation of information understood as semantic content with respect to qualitative growth of knowledge, and change in knowledge structures in domains.
References


Frohmann, B. (2009), Revisiting "what is a document?", *Journal of Documentation*, 65(2), 291-303


Hjørland, B. (2002A), Domain analysis in information science: eleven approaches - traditional as well as innovative, *Journal of Documentation*, 58(4), 422-462


Kari, J. (2010), Diversity in the conceptions of information use, *Information Research, in press*


