

# [Chapter 6] Information attributes

Diane Pennington and Ian Ruthven  
University of Strathclyde

In this chapter, we focus on the concept of information attributes. Information attributes are the properties of information and information objects that can be used to describe and differentiate information. Being able to differentiate information objects means we can select the most appropriate objects for our tasks and we can design information systems to organise and store information in useful ways. We start by providing a discussion on why being able to describe properties of information objects is important, and then we provide a description of essential concepts in this area to provide a theoretical background to information attributes. Following this, we highlight a selection of key information attributes to give a flavour of the kind of studies that have been conducted in this area before concluding with current trends in this research area.

## 1. Introduction

Information attributes are everyday concepts and part of our everyday language when discussing information. When dealing with patient information Ann, our cardiologist, will care about issues such as the *novelty* of any information representing change in her patient's condition, its *accuracy* and the *quality* of the information she is receiving. Novelty can be assessed by the time of the information, recorded in a standard format that all medical staff use and understand, accuracy may be judged by knowledge of how the information was created and quality of information may be judged by who provided the information or how useful the information is in deciding how to treat her patients. Ann will also have to deal with various *genres* of information such as patient records, heart rhythm traces, temperature charts, etc. which can be used for different purposes and to make different decisions. Many information attributes require high levels of domain knowledge to be used correctly; part of Ann's professional training dealt with how to read these documents and how to use them to treat her patients.

Our lawyer Johan will also interact with many types of information and care about issues of *bias* in the information, the degree to which his information can be *verified* with respect to other sources of information and whether he has *confidence* in his information source. In presenting evidence to the court he will need to consider whether his evidence is *clear*, if it provides *tangible* facts and whether it is *consistent* with other information being presented. Johan also needs to care about whether his own legal knowledge is *up to date*, whether his interpretation of the law is *consistent* with other lawyers and whether he is translating specialised information in such a way that his clients and the jury will have the *ability to understand* the information he is providing.

Liila, our journalist, has to deal with many sources of information, and she will also care about novelty of information and its accuracy but she also has concerns over whether the information is *recent*, whether her sources are *authoritative*, whether the information is sufficiently *specific* to her story and the *affective* nature of the information on her potential readers. She may also have to care about the *cost* of information, whether information is *available* in time for her deadline, whether the information she is getting is *dynamic* and therefore does she have to keep updating her story and maybe issues such as whether the information only covers a certain *geographic area*. She will also care about whether the information she needs is *findable* with the tools she uses in her daily work. Ann, Johan and Liila all care about the same thing – is their information *good* information? Their differing work roles mean that what they see as 'good' information will change depending on their uses of information. The italicised words in the preceding examples are indications of kind of information attributes we can use when assessing the information we have available. In the rest of this chapter, we expand on this core idea that we can assess information by attributes of the

information and information objects to understand how others might make these decisions, and to improve the design of information systems that provide better support for making these decisions. We use the term objects to refer to the narrow concept of storable and findable items compared, a subset of the wider concept information artefacts from the previous chapter.

## 2. Theoretical underpinnings

Maron noted in an early Information Science contribution, 'Information is not a stuff' (Maron 1965). However, it is often treated as such with many papers written in information seeking, information management and information use that treat information as a vague generic thing that we create, search for, use, or disseminate, without ever specifying what type of information is being created, searched, disseminated or used or what form that information takes. This is unsatisfying on many levels; not least because it creates an imprecision in our understanding of what is happening with information in workplace settings and what types of information are being used for what purposes.

This has led to many serious, but perhaps inconclusive, attempts to define what is meant by this thing we call 'information' (e.g. Weaver 1949, Fritz Machlup 1983, Shannon 2001, Cornelius 2002, Capurro and Hjørland 2003). These definitional approaches frequently remain abstract and lack operational power to move us forward in understanding what it is we do with information in real environments. No single definition has really stuck and the debate rolls on over what we mean by information. Possibly, because 'information' itself is so complex and difficult to define, it is also challenging to identify a set list of attributes or characteristics that can be used to describe information. Below are tangible as well as more abstract notions of the qualities used in Information Science research to describe the attributes collectively comprising the concept of information: objects, properties, and relevance.

### *Objects*

Even if we can't agree on the meaning of 'information', what Information Science has successfully done is provide ways to classify information sources and *objects* to create a vocabulary around objects, their organisation and use. We can perform this classification in two ways: by characterizing properties of the information objects such as documents, video, and email, or by properties of the information we obtain from these objects.

Information objects have many useful properties that can help us access, use and discuss information. Factual, if sometimes controversial, properties such as source, age, language, and type, can provide useful metadata for querying. The objects' contents can provide the raw material for indexing schemes to allow end-user searching. Usage statistics can provide the basis for recommender systems and for quality metrics such as impact factors. The different natures of objects can lead to systematised forms to which we ascribe names such as 'memo', 'report', 'brochures', 'press releases', etc. and which helps us understand what objects we are interacting with.

We can also create algorithms to help us assign attributes to information objects, for example based on the level of textual complexity or readability level, and there is a rich stream of such research seeking to analyse information automatically to provide new features for searching and browsing.

### *Properties and relevance*

A separate stream of research is in describing information by what we see as *properties* of the information we obtain from interacting with these objects, e.g. the value of the information to us - a subjective concept compared to the more objective concept of cost. Many of these properties are relations between us and the information objects we are dealing with. This idea of looking at what characteristics we use to decide whether information is relevant to our needs was investigated robustly by Barry and Schamber in two studies, both based on workplace information seeking (Barry,

1994, Barry and Schamber 1998, Schamber, 1991). Their investigations showed that there is a large set of relevance criteria, or values that people use to decide on relevance, and that these criteria are a mixture of object properties (such as source), information properties (such as validity of the information), and personal properties (such as ability of the user to understand the information). Further studies have demonstrated that different relevance criteria may be used in different workplace settings but that they form a stable core of decisions we make about information and information objects.

The concept of relevance itself is an important one as workplace tasks are built around being able to access relevant, rather than just any, information. Relevance is another term that has at the same time too many and too few definitions; too many that we can confidently say what it means and too few that operationalise it in a way that we can confidently measure. Nevertheless, there are some useful definitions. In 1975, Saracevic claimed that Information Science 'emerged as the third subject, along with logic and philosophy, to deal with relevance ... Relevance is considered as a measure of the effectiveness of a context between a source and a destination in a communication process' (Saracevic 1975).

Later, in 1996, Saracevic presented five contrasting definitions of relevance (Saracevic 1996):

- *System relevance*, or the relationship between the representation of an information object and the representation of a user's search request within a given information system. In some ways this is the lowest meaning of relevance as it is based on algorithmic matching or similarity but it is a very common way of interacting with information. This level says nothing about an object beyond its content.
- *Topical or subject relevance* which is the relationship between the topic of the information object and the topic of the user's search request. This goes beyond information that is contained within the object to what the object is 'about' at a more thematic level. This 'aboutness' may be inferred from external sources such as user-generated content but still stays at the content level. In both system and topical relevance, the indexable properties of the objects are important.
- *Cognitive relevance* or the relationship between the state of knowledge and needs of the user and the texts (rather than the representation of the texts). Here what is important is not just content but also issues such as novelty, quality, and the ability to understand the material with which we are interacting.
- *Situational relevance* or the relationship between the situation and the information objects are key here. What is relevant is what helps us in progressing through a situation and this may be about better decision-making, problem-solving, or establishing precedent. Situational factors can therefore include issues about cost of information or findability of information that vary across situations.
- *Motivational relevance* or the relationship between the goals of the user and the information objects. This goes beyond moving ahead in a situation to thinking about issues such as satisfaction with how a situation has been tackled. At this highest level of relevance, we see issues that relate to preferences and motivations that will relate to aspects such as the complexity of information, or its readability or affective issues such as how interesting texts are to read.

Therefore, different levels relate to different attributes of information objects: lower levels such as the system level involve simple, although easily manipulated, properties of objects such as content, whereas higher levels involve attributes that require more complex system support. This support may come in the form of more sophisticated indexing features, such as the algorithms used in music

retrieval to classify music by mood or genre, or in the form of improved interactive support. This support can involve making available more information to the user, which, in turn, gives more power to make better decisions. We already see different information attributes being surfaced in different applications: price, origin, seller feedback for online shopping sites, for example, or in the case of video sites, making available information on the number of views, source, popularity, etc. which can be used to help assign values to the various relevance criteria which a user may be working with.

Relevance is not the only factor influencing users' selection of sources. For example, an additional one is the objective *cost* of information. Hardy (1982) analysed the Cost/Benefit model of selecting information sources, which 'proposes that information seekers select information sources on the basis of expected benefits and expected costs of using an information source' (p. 289). In his study of forestry workers, costs were much more important than benefits in source selection.

There are many information attributes. In what follows we present an analysis of some of them to illustrate the range of research in this area. We look at classical attributes such as complexity, attributes, such as genre, that are being constantly redefined and new attributes, such as findability, that are arising from the need to have new ways of thinking about information in new information environments.

### *Genre*

A document's genre, or what type of document it is, can influence whether a workplace user believes a retrieved document is relevant and where to look for information in the first place. For example, if software developers want to know how other developers wrote code for a particular situation, discussion forums will be more helpful than vendors' manuals. In other words, the task the worker needs to complete influences what type of document is needed (Freund, Toms et al. 2005).

Genre can be described as a set of texts that share a purpose as identified by the community which produces or uses them (Swales 1990). Yates and Orlikowski (1992) stated, "genres are social institutions that are produced, reproduced, or modified when human agents draw on genre rules to engage in organizational communication". That is, groups of people working together create document *types* to help them communicate for different purposes and these types tell us something about what each group of people see as important to completing their tasks, or as Watt (2009) has observed: "convergence on a set of standardized document structures is both natural and helpful". We can see genres developing over time in information repositories as authors gradually develop information structures to help readers find information easily within articles (Clark, Ruthven et al. 2009). There is also evidence from eye-tracking studies that genre familiarity reduces the cognitive effort in analysing texts (Clark, Ruthven et al. 2014).

Freund, Clarke et al. (2006) studied software engineers to develop a taxonomy of 16 genres for documents including manuals, presentations, product documents, tech notes/tips, tutorials and labs, white papers, best practices, design patterns, discussions/forums, cookbooks and guides, engagement summaries, problem reports, and technical articles. Other disciplines have their own standard and emerging genres that are important to understand and be able to use in order to become part of that discipline or profession.

Some work-related genres seem universal: lectures, user manuals, company reports, journal articles, etc. will be instantly recognisable across cultures and organisations. Others may be entering history as they disappear from use. The memorandum pinned to the company noticeboard was at one time a critical means of disseminating instruction and information. Now, email and texting have almost eliminated this type of communication. Equally, new developments create new genres such as

webinars, vlogs, and tweets that have to be blended into institutions' information activities and require learning by individual workers and entire institutions on how to best use them.

In many areas of work life, there are particular types of genre that only those in that area truly understand how to create and read. The 'convergence' of a community into Watt's 'standardized document structure' makes it easier for those working in a work community to communicate better and work more efficiently but makes it harder for those outside to understand and access information. Certain areas of work, such as the legal field with its case notes, affidavits, summons, subpoenas, plaint notes, contracts, etc., are so rich in genre that we often need professional intermediaries, such as journalists, to explain what it all means.

Genres can provide the basis for information organisation in many workplaces. Sometimes it is as simple as having all staff records in one filing cabinet and all invoices in another one. For larger institutions there will be often more sophisticated support, including specialist IT systems, for storing and searching within and across genres, and we may have specialist sub-units like Human Resources teams who take charge of developing particular genres and their deployment within institutions.

Genres make it easier for us to know where to start looking for information: if Ann our cardiologist wants to learn about new treatments for heart disease she knows to start with recent medical journals rather than post-mortem reports. The 'standardized document structures' highlighted by Watt make it easier to create IT systems to support our access to information by allowing us to search with this structure or to use the structure to perform textual analyses like summarisation.

### *Complexity*

Some information we find easy to follow; some we struggle to understand. If we cannot understand information or rather cannot understand it quickly or easily enough for our purposes, then we may reject it altogether. For the purposes of this chapter, we will refer to the *complexity* of information as denoting how easy or difficult it is to follow. This doesn't help explain what complexity is and, like in most fields, it can mean different things but it does denote the general sense of 'hardness', 'challenge' or 'difficulty' of some texts over others.

Complexity is usually a relation between the person reading a document and the document they are reading. We may find one document to be complex but you do not as you have more experience in that area, or more familiarity with the vocabulary used in the text than we. Experience and practice does count and Liila's long experience of reading court reports will give her an advantage over a more junior reporter in knowing which parts of the document to read first, what to look for in the document and what various concepts mean. Complexity, therefore, is relative to the person doing the reading.

However, complexity has an objective side as well. There are many tests of so-called 'readability' that attempt to measure how difficult texts may be to read based on textual features such as sentence length, word length and word frequencies. One of the more famous ones, the Flesch reading-ease test, gives higher scores to material that is seen as being easier to read. A score from 90-100 denotes text that is very easy to read and seen as easily understood by an average 11-year-old whereas a score of between 0-30 indicates text that is difficult to read and most suitable for those with university-level education. This chapter has a Flesch reading-ease score of 32.9!

A related test, the Flesch-Kincaid Grade Level test, maps scores from similar textual features onto education levels indicating what level of education is necessary to read the text. Although criticised by some for being overly simplistic, these readability scores have been influential in encouraging analyses of whether information provided by certain organisations is too complex. In some cases, this

has led to laws to ensure that information by some companies, e.g. by insurance companies, is at the right level of complexity to be understood by most of its customers.

Low complexity is not a desired result in all cases. Van Der Sluis et al. for example have shown that newspaper articles that are too simple are just as likely to be rated as uninteresting as articles that are too difficult to read, suggesting complexity is balanced against other factors when deciding what information we prefer (Van der Sluis, Broek et al. 2014). There is indeed a tension around communication in workplace environments regarding the complexity of information; we know experts can deal with complex information more effectively than novices and information that is insufficiently succinct is annoying. However if information is too complex then we reduce the number of people who can deal with it.

### *Novelty*

Barry and Schamber (1998) distinguished between three different types of novelty: content, source and document novelty. Content novelty is the extent to which the information presented is novel to the user, source novelty the extent to which a source of the document (i.e., author, journal) is novel to the user and document novelty or the extent to which the document itself is novel to the user. These are all related to the user of the information and their current state of knowledge; we may also add in recency or the extent to which information is recent, current, or up-to-date – a more objective notion of novel.

The ability to distinguish what is new is critical to many workplace environments: Ann needs to keep up to date with the latest medical knowledge, Johan needs to know what laws are current to be able to adequately defend his clients and Liila's whole occupation is based around what is 'new(s)'. For many workplaces, keeping up to date with the latest information and having a good range of high-quality information sources is essential to competitive success and attention has to be given to how to ensure workers are kept up to date. For all organisations there is also a requirement to keep up to date with compliance issues such as health and safety, tax laws and other business purposes.

How we keep up to date shows a variety of practices. In academic work, journals and conferences are standard ways to recognise the latest findings in a research field. But we have to know which conferences and journals to read or we may miss important information. We also need to know how to select which are the best ones to read. Those who are better connected will also know which sources of research information are doing the best research and may have indirect connections to these researchers to get highlights before they become official literature. In other areas, such as journalism, having good sources of information that can guarantee the latest information is also a critical advantage.

In other fields, particularly entrepreneurial areas, *being* novel is crucial. Investing in a new product could be a complete waste of time and resources if a rival has already patented the idea. In areas such as intellectual property search (or patent searching), search success is the confidence that one has not found a previous patent (Trippe and Ruthven 2011), i.e. success can be defined as the failure to find relevant information. However, knowing that we have failed to find information because it does not exist – no one else has patented our idea – rather than we have just searched ineffectively can never be properly resolved. Several studies, including influentially (Blair and Maron 1985), have indicated that we are not very good at telling whether we have all the information that we need when searching.

### *Metrics*

With so much information available, we need ways to help decide where to focus our attention. One approach is to employ metrics or quantitative measures as surrogates for quality or popularity and for other purposes. A variety of metrics, exist to track and describe information. Metrics allow owners

and users of information to answer a variety of questions about how the information is used and who is using it. There are many terms for this area: 'bibliometrics', 'scientometrics', 'informetrics', 'webometrics', etc. but what all share is a concern over what data we use to create metrics, what they represent and their reliability for decision-making.

For academics, bibliometrics is a field of research within Information Science that "boils down to the quantitative analysis of published scholarly literature, notably journal articles and the network of their bibliographic connections" (De Bellis, 2014, p. 23). Bibliometric information allows researchers to track the reach of their research, such as how many times their papers have been cited, the disciplines of the authors that have cited them, and so on.

Traditional metrics, such as impact factors, are being challenged by new types of media with their own metrics, such as a scholar's number of Twitter followers, number of likes on a Facebook post linking to a newly published journal article, etc. (Cronin and Sugimoto 2014). Altmetrics is an approach to tracking newer and previously unnoticed or non-existent forms of research impact, such as how many times an article is mentioned on Twitter, or how many blog posts discuss it (Priem, 2014).

Both within and outside academia, social media encompasses a massive amount of metrics. Some are available to the user, and some are not. For example, on Facebook, users can see how many times their posts are liked, shared, commented on, and viewed. Facebook's algorithms track what posts users view and whom they interact with most often; users do not see these metrics themselves, but the metrics do influence what users see in their Facebook feed.

Organisations can gather and analyse website metrics through simple programmes that exist on their servers, or with external services such as Google Analytics. These programmes provide a plethora of details, such as how their users found their pages, the process each user followed while progressing through the site, how long they spent on each page, and much more. These metrics are useful to understand what parts of a website its visitors value as well as where they might get frustrated and therefore leave the site.

How might metrics influence the use of workplace information? Consider a corporate intranet containing, in part, a library of documents that staff need to do their work. If staff can see how often everyone has accessed each document, a high accession rate might influence their decision to choose a particular document, because the ones used the most are likely the most useful and current.

### *Authority*

The authority of an author or other information creator is an important attribute of information. It is essential for a critical reader to consider if and why a given creator is the right person to provide given information. In a time when "alternative facts" and "fake news" have become common terms, it is difficult to know whom to trust. The authority of a creator is likely related to the reliability and quality of information provided. These are important attributes for people as they decide whether to use information based on its trustworthiness and/or whether it is relevant to their needs. In a seminal study (Rieh 2002), users' judgments of online quality and authority 'were identified in terms of characteristics of information objects, characteristics of sources, knowledge, situation, ranking in search output, and general assumption' (p. 145).

Cognitive authority is one measure of authority, which is defined as the extent to which people think they can trust information based on who wrote it or where it is published (Wilson 1983). For example, if physicians are seeking updated information about a condition, they will likely think new research published in a top medical journal is a more authoritative choice for making decisions about treatments for the condition than patients' blogs about their personal experiences with the disease.

Before the Internet, it was perhaps easier to determine cognitive authority, because apparently reliable sources were in printed form, such as a book, a journal, an encyclopaedia, or any other resource found in a library. The librarians chose the most reliable sources based on the author's reputation, the credibility of the publisher, and so on. It is no longer possible to make these relatively straightforward decisions.

In today's online environment where anyone can create and post information, it has become more difficult to ascertain cognitive authority, and new models of authority need to be negotiated (Neal, 2010). Neal and McKenzie (2011) performed discourse analysis on blog posts written by women with the female chronic illness known as endometriosis, to examine how women make decisions about the authority of information about the disease and its possible treatments. From the research, they found a different type of authority, which they coined "affective authority", as "the extent to which users think the information is subjectively appropriate, empathetic, emotionally supportive, and/or aesthetically pleasing" (p. 131). Examples included information about the lived experience of particular treatments from other women with endometriosis as shared on blogs and discussion forums, the sensitive approach of a religious leader, or an understanding friend. While these figures do not have authority in Wilson's cognitive sense, they did provide substantial affective authority to the women in the study.

In a case study of environmental activists, participants' personal perceptions of media credibility and cognitive authority were essential to judging and selecting information sources; environmental agencies were seen as most credible, while newspaper were seen as less credible due to bias (Savolainen 2007). Wathen and Burkell (2001) performed a literature review regarding what factors influence judgements of the credibility of an online source. The answers were not straightforward; the factors included an entanglement of the source, the message, the institutional quality such as the owner of the website, and the receiver's perspective, such as information need, existing knowledge, and social situation.

In any workplace setting, the authority of information consulted to perform a work task should be considered carefully. For example, for people in a new employment setting, it might be preferable to ask a superior for information until they know which peers are appropriately knowledgeable and approachable. Information stored on shared network drives or corporate intranets, although "official" and institutionally supported, may be outdated, provide conflicting details, or presented to staff in a disorganised manner. It is then left to the member of staff to determine what version of a document is the most authoritative version (based on its currency, for example) before proceeding to use the information.

### *Findability*

Another essential attribute of information is findability. If people cannot find information, they cannot use it, rendering it essentially worthless. (Morville 2005) provided three definitions of findability:

- a. The quality of being locatable or navigable.
- b. The degree to which a particular object is easy to discover or locate.
- c. The degree to which a system or environment supports navigation and retrieval. (p. 4)

Imagine our lawyer Johan preparing for a trial. He needs to look up legal precedent that relates to the case. He knows the precedent exists, and it is probably in his firm's small legal library, but he does not remember the exact date of the ruling, and he cannot determine where to browse within the many volumes that could have the ruling in them. The information he needs is (probably) there, but it is not findable. He proceeds his preparation without the details he wanted.



Increasingly, however, the necessary information is stored online. People must rely on search engines such as Google to find what they need. Over time, the ability to find 'known items', such as when we know the title or part of the text, has improved substantially. For example, if Johan's precedent were available online, he could likely type in the name of the case he wanted into a search engine and find it almost instantly.

Let us imagine that Johan knew he needed a previous case to discuss in court, and it existed online, but he did not know anything about it, such as its name or any text in it. What does he search for in order to find it? A classic problem inherent in search engines is that people must tell the search system what it is they do not know (Norman, 1988). In other words, it is difficult to find something when you do not know what it is.

The quality and amount of the organisation and description of information within a search system greatly influences findability. For example, the way a website presents the structure of its information can influence whether someone finds needed information. If Ann, the cardiologist, needed to consult an online resource to ensure her prescribed dosage of a medication is correct for a patient, how should the site be organised? She might want to access the medication information by medication name, medication type, diagnosis, symptoms, potential side effects, or a range of other facets that could sort and describe the medications. But if she goes to the resource, and the only way she can access information is through a very long page of medications, perhaps only sorted alphabetically, the resource is not organised well.

Closely related to the organisation of information is its description. In traditional Information Science terms, bibliographic description of an item refers to a set of descriptors about the items, such as its title, author, date of publication, number of pages, and so on; think about what descriptions library catalogues provide. Again, this is only helpful if you know some of the descriptive information for searching online resources. Many search engines solely or primarily search the descriptors in determining whether an item in the system matches the user's search. Additionally, the description tends to drive the organisation of the system and its findability. Consider a department store's online shop. A person wants to find a women's red t-shirt in size 10 for under £20. The descriptions provided for all the shirts, as well as how the shirts matching her search return to her, determine findability. If the shirt she wants exists on the site, but she cannot find it, the shop will needlessly lose her business.

Information attributes can be used to create better interactive displays of information objects. Once we have attributes defined, we can create attractive and useful browsing systems to allow the user to narrow down the information they require by size, shape, colour, type, cost, origin, popularity, or whatever attributes are meaningful and obtainable to that domain and set of information tasks. These displays and features improve findability because they allow users to determine what features are relevant to their needs.

### **3. Conclusion**

Because information is hard to define, the characteristics of information also present a challenge to Information Science researchers and practitioners. Perhaps as the field continues to increase its understanding of what attributes are important to information users, it will be easier to define the vital attributes that emerge from information itself. Even if we agree on which attributes are important it can be a challenge to agree on how to measure an attribute such as complexity or authority.

It is worth noting that interactions performed on a document by its users can alter its attributes. In the Internet age, people are able to add, edit and delete information attributes in previously impossible ways that can ultimately determine whether documents are retrieved. For example, the number of likes or favourites on a document posted on a social media website can influence its

relevance ranking in the website's retrieval system (Neal 2010). This is an important factor to consider in workplace settings as people increasingly use freely available social media resources such as YouTube in order to learn how to complete tasks or acquire new skills. Additionally, since so much information shared online is now in non-textual forms, such as videos, photographs, maps, music, and so on, traditional attributes may need to be reconsidered and reconceptualised (Neal 2012). It is perhaps not possible to represent important attributes of a visual document fully in words, since important features are lost in translation between word and image (O'Connor and Wyatt 2004).

As we develop more ways to automatically process information, we see more and more indexing features being developed and included in algorithmic approaches to information retrieval and management. These do tend to be relatively low-level features. Higher-level ones, the ones used in Saracevic's higher relevance levels, remain difficult to operationalise especially those that reflect more subjective assessments of information. Whilst waiting for algorithmic solutions to all information attributes there is a rich seam of useful research in developing interactive systems that support people making useful decisions on information attributes.

Some attributes require domain knowledge to be used effectively. Ann, Johan and Liila all care about issues such as authority, quality and accuracy of the information they use and will use their knowledge about sources, professional norms and how to interpret information to make such decisions. When we lack this domain knowledge, the evidence suggests we use different information attributes – ones we can more easily work with – to substitute for these 'expert' attributes. So, for example, if we cannot judge the accuracy of information in a document then we may look for evidence that other documents also contain the same information, i.e. the property that information can be verified, or we may look for the property of tangibility (containing concrete facts) to estimate accuracy (Wen and Ruthven, 2006). That is, we can be flexible about our decision-making and which attributes to use based on the information we have available.

Information attributes are a fascinating area that informs us both about how people think and act on information as well as how we create information organisation principles to support successful information interactions. As technology develops we see new forms of information develop requiring new techniques for curating, assessing and working with these information objects resulting in a very lively research area.

#### **4. References**

- Barry, C. L. (1994). User-defined relevance criteria: an exploratory study. *Journal of the American Society for Information Science*, 45 (3), pp. 149-159
- Barry, C. L. and L. Schamber (1998). "Users' criteria for relevance evaluation: a cross-situational comparison." *Information processing & management* **34**(2-3): 219-236.
- Blair, D. C. and M. E. Maron (1985). "An evaluation of retrieval effectiveness for a full-text document-retrieval system." *Communications of the ACM* **28**(3): 289-299.
- Capurro, R. and B. Hjørland (2003). "The concept of information." *Annual review of Information Science and technology* **37**(1): 343-411.
- Clark, M., et al. (2014). "You have e-mail, what happens next? Tracking the eyes for genre." *Information Processing & Management* **50**(1): 175-198.
- Clark, M. J., et al. (2009). "The evolution of genre in Wikipedia." *Journal for language technology and computational linguistics* 25(1): 1-22.
- Cornelius, I. (2002). "Theorizing information for Information Science." *Annual review of Information Science and technology* **36**(1): 392-425.
- Cronin, B. and C. R. Sugimoto (2014). *Beyond bibliometrics: Harnessing multidimensional indicators of scholarly impact*, MIT Press.

De Bellis, N. (2014). "History and evolution of (biblio)metrics." In: Cronin, B. and Sugimoto, C. eds. *Beyond bibliometrics: Harnessing multidimensional indicators of scholarly impact*, MIT Press, 23-44.

Freund, L., et al. (2006). *Towards genre classification for IR in the workplace*. Proceedings of the 1st international conference on Information interaction in context, ACM.

Freund, L., et al. (2005). "Modeling the information behaviour of software engineers using a work-task framework." *Proceedings of the American Society for Information Science and Technology* **42**(1).

Fritz Machlup, U. M., Ed. (1983). *Study of Information: Interdisciplinary Messages*.

Hardy, A. P. (1982). "The selection of channels when seeking information: Cost/benefit vs least-effort." *Information Processing & Management* **18**(6): 289-293.

Maron, M. (1965). *Mechanized Documentation: The Logic Behind a Probabilistic. Statistical Association Methods for Mechanized Documentation: Symposium Proceedings*, US Government Printing Office.

Morville, P. (2005). *Ambient findability: What we find changes who we become*. Sebastopol, CA, O'Reilly Media.

Neal, D. M. and McKenzie, P. (2010). "Emotion-based tags in photographic documents: The interplay of text, image, and social influence/Les étiquettes basées sur des émotions dans les documents photographiques: l'interaction entre le texte, l'image et l'influence sociale." *Canadian Journal of Information and Library Science* **34**(3): 329-353.

Neal, D. R. (2012). *Indexing and retrieval of non-text information*, Walter de Gruyter.

Norman, D. (1988). *The design of everyday things*. MIT Press.

O'Connor, B. C. and R. B. Wyatt (2004). *Photo provocations: Thinking in, with, and about photographs*, Scarecrow Press.

Priem, J. (2014). "Altmetrics." *Beyond bibliometrics: Harnessing multidimensional indicators of scholarly impact*, MIT Press, 263-288.

Rieh, S. Y. (2002). "Judgment of information quality and cognitive authority in the Web." *Journal of the American society for Information Science and technology* **53**(2): 145-161.

Saracevic, T. (1975). "Relevance: A review of and a framework for the thinking on the notion in Information Science." *Journal of the American Society for Information Science* **26**(6): 321-343.

Saracevic, T. (1996). *Relevance reconsidered*. Proceedings of the second conference on conceptions of library and Information Science (COLIS 2).

Savolainen, R. (2007). "Media credibility and cognitive authority: the case of seeking orienting information." *Information Research* **12**(3). <http://www.informationr.net/ir/12-3/paper319.html>

Schamber, L. (1991). *Users' criteria for evaluation in a multimedia environment*. Proceedings of the 54<sup>th</sup> Annual Meeting of the American Society for Information Science, Vol. 28, Learned Information, Medford, NJ, pp. 126-133

Shannon, C. E. (2001). "A mathematical theory of communication." *ACM SIGMOBILE Mobile Computing and Communications Review* **5**(1): 3-55.

Sluis, F., et al. (2014). "When complexity becomes interesting." *Journal of the Association for Information Science and Technology* **65**(7): 1478-1500.

Swales, J. (1990). *Genre analysis: English in academic and research settings*, Cambridge University Press.

Trippe, A. and I. Ruthven (2011). *Evaluating real patent retrieval effectiveness*. Current Challenges in Patent Information Retrieval, Springer Berlin Heidelberg: 125-143.

Wathen, C. N. and Burkell, J. (2001). *Believe it or not: factors influencing credibility on the Web*. *Journal of the American Society for Information Science & Technology* **53**(2): 134-144.

Watt, S. (2009). *Text categorisation and genre in information retrieval*, John Wiley & Sons, Chichester, UK.

Weaver, W. (1949). "The mathematics of communication." *Scientific American* **181**(1): 11-15.

Wen L., Ruthven I. and Borlund P. (2006). "The effects on topic familiarity on online search behaviour and use of relevance criteria." In: Lalmas M., MacFarlane A., Rürger S., Tombros A., Tsirikia T., Yavilinsky

- A. (eds) *Advances in Information Retrieval. ECIR 2006. Lecture Notes in Computer Science*, vol 3936. Springer, Berlin, Heidelberg.
- Wilson, P. (1983). *Second-hand knowledge: An inquiry into cognitive authority*. Westport, CT, Greenwood Press.
- Yates, J. and W. J. Orlikowski (1992). "Genres of organizational communication: A structural approach to studying communication and media." *Academy of management review* **17**(2): 299-326.