Ranganathan in the Perspective of Advanced Information Retrieval

PETER INGWERSEN AND IRENE WORMELL

Examines Ranganathan's approach to knowledge organization and its relevance to "intellectual accessibility" in libraries. Discusses the current and future developments of his methodology and theories in knowledge-based systems. Topics covered include: semi-automatic classification and structure of thesauri; user-intermediary interactions in information retrieval (IR); semantic value-theory and uncertainty principles in IR; and case grammar. Elaborates on the usefulness and applications of analytico-synthetic methodology in designing classification schemes, thesaurus construction, query negotiation, structuring of search expressions, etc. Close parallel is recognizable between Ranganathan's principles and case-frame theory, for example in techniques of IR, contextual IR theory building and software reuse systems. Concludes that though Ranganathan's criterion of helpful sequence and other aspects of his methodology may have limitations when applied to areas other than subject classification, his principles for knowledge organization and faceted methodology are likely to find applications in cognitive modelling, neural network techniques and pattern recognition.

1. Introduction

S. R. Ranganathan, who passed away in September 1972, had time enough to sense the early achievements of automated IR processes, which were based on rather simple computer operations with one-to-one correspondence between stored bibliographic data items and queries formulated as single words and lacking the possibility to express the multidimensional growth of various subjects.

Exactly because he was a great philosopher and having already extended the horizons in Library and Information Science more than once, his ideas of

The authors work for the Royal School of Librarianship, Birketinget 6, DK-2300 Copenhagen S, Denmark.

Libri 1992: vol. 42, no. 3, pp. 184–201

categorisation naturally reach beyond an argumentation alone for the actual practice in classification at that time. They develop into our time and future of computer technology, embedded in hypertext, knowledge-based systems, and neural nets – but in addition, evidently encompassing ancient storage media and information technologies like clay-tablets, parchment or paper as well as handcopying and printing.

To transcend the media, avoiding their various limitations and pitfalls, moving into an abstract landscape of knowledge evolving through time, requires a particular mind. Ranganathan possessed that kind of mind – had this ability.

Albeit, some of the senior colleagues and most of his contemporaries, did not completely understand or agree with the logic and soundness of Ranganathan's thoughts and theories.

However, observing the emerging use of modern information technology in processing bibliographic data, he understood the potentialities of *deep classification* in relation to the powerful capabilities of computers in *creating multiple access points* and *dynamic search procedures* for *identifying microscopic units of thought* embodied in different pieces of information, potential to users. He stated that "Electronic engineering has begun to throw a challenge to the theory of classification"[1].

The emergence of the new technology gave intellectual stimulation to him for sharpening the theory of classification in depth on the idea level as well as on the notational level. Commenting on the Doc-Finder, one of those days' electronic aids for document retrieval, he argued:

"... it is not going to enable the classificationist or the classifier to rest on his oars. On the other hand, it will make him work even more than ever before. This should be realized by the library profession and particularly by its specialised branch, called the professional documentalists"[1].

It is interesting to note, however, that the library world has shown only limited interest for the further developments of his dynamic theory of classification.

First this paper examines Ranganathan's methology and its relevance to *intellectual accessibility* in libraries. This is followed by a discussion of present and future developments of Ranganathan's methodology and categorial ideas in knowledge-based retrieval. Four areas of interest will be touched upon:

- semi-automatic classification and thesaurus structures;
- request model building in IR interaction;
- semantic value theory and uncertainty principles in IR; and
- case-frame methodology.

This selection of areas does not pretend to be exhaustive in terms of Ranganathan's contributions neither to information science and IR itself nor to other disciplines concerned with communication of knowledge and information transfer. The intention is to demonstrate the validity and adequacy of his thoughts in relation to advanced IR solutions, based on theories of linguistics and communication.

2. A new dimension

His arguments for the re-examination of traditional practice of library classification and for the introduction of a New Foundation are based on the *recognition of the universe of current knowledge as a dynamic continuum:* "It is evergrowing; new branches may stem from any of its infinity of points at any time; they are unknowable at present. They can not therefore be enumerated here and now; they cannot be anticipated; their filiation can be determined only after they appear"[2]. He wanted to open up a new dimension in library classification so that it could stand a higher pressure from the field of knowledge than those classification systems virtually adopting enumerative schemes, having a rigid finality and still resting on arbitrary numbering at bottom.

Derek de Solla Price, another distinguished personality of our field, has expressed his critic against the static character of traditional classification systems in the following short but concise formulation: *"Last year's index is no good for this year's science"*[3].

The ignorance of Ranganathan's suggestions in the library world and the hammering critique against the "practicability" of his theories, are due to the fact that the majority of classificationists do not feel the need for basing their work on a dynamic theory. Libraries confine their services to the feeding of readers only with the macro subjects embodied in books and have shown but scarce interest, also to reveal the micro-level units of thoughts in their collections.

However, the pressure on today's library services calls for sharpening of old foci and formation of new ones. Classificatory thoughts must now be tuned to these problems and more powerful devices and techniques discovered. To have the capability of isolating and securing the individualisation of

thought-units, throwing them into a helpful sequence, is requiring far more subtle principles of knowledge representation than the conventional classification practices are resting on.

Ranganathan's struggle to arrive at fundamental principles for splitting specific subjects into various resultant constituents – facets – and then to synthesize them into subjects sought for with the help of notation, was quite in line of research at his time. His search for a "universal" model to organize and classify information that appears in recorded form, has in our discipline represented a passing from macro areas to micro units in discovering the contents and the corpus of subject domains. His theory and the application of *scientific principles in analysing and synthesising* a specific subject and to individualise it out of millions, can be seen as a phenomenon similar to those research interests that have guided other scientific areas too. For example in Physics, where the atom was split and the Bohr-Rutherford Model of the constituents of the atom served as theoretical basis for the explanation of its properties; and in Biology, where RNA and DNA were observed in cells, and research got a step closer to the understanding of the micro world of genetics.

The fact, that the classification schemes and ordering systems in libraries always reflect the manner in which people think in a certain period, has got a fascinating illustration in the famous book by the Italian semiologist Umberto Eco: "Il nome della rosa" from 1980. The atmosphere in the library of a medieval monastery, the rituals, secrecy, and the traditions of the place, together with the physical shape of the library as a labyrinth, combined with a merely undecipherable notation, are all strong expressions of the Medieval Age's attitude against accessing new knowledge.

Ranganathan represents a milestone in the development of intellectual processes involved in the organisation and classification of recorded knowledge. His approach of analysis and synthesis in classification will always be valid, because his application of this scientific principle is not likely to change or appear obsolete at any point of time.

It is certain that Ranganathan knew the limitation of his contribution to the future, both in classification in general and, in particular, in relation to the Colon Classification. Thus it should be noticed, that details in his classification system may meet obsolescence but not its underlying principles!

3. Physical, bibliographic and intellectual accessibility

Subject bibliographies and catalogues are, traditionally, our intellectual channels to the recorded knowledge of mankind. Therefore, it is interesting

to take a short historical overview of the rôle of bibliography and classification schemes as instruments of communicating knowledge.

Among cataloguers there have always been two main viewpoints dominating the theoretical discussions in bibliographic research: The *macrocosmic view*, relating to the groups of researchers who were urging a broader perspective to the philosophy of bibliography than was traditionally conceived among librarians. They understand bibliography as an instrument of the communication process throughout society and they state that the main purpose of bibliographic work is to maximise the social utility of document records for the benefit of mankind. The two eager spokesmen of this view were Jesse Shera and Margaret Egan. Already in 1952 in the "Foundation of a Theory of Bibliography"[4] they take the premise that information science is concerned with the transfer of desired information from human generator to human user. They understood the rôle of bibliography as an instrument of communication and dissemination of knowledge.

The *microcosmic view* is represented by people for whom bibliographic work meant "an uncritical listing of all books and materials within the coverage of the given bibliography"[5] and consequently, "it has nothing whatever to do with the subject or literary content of the book"[6].

A new consciousness about the breadth of the problem of accessibility is reflected by the fact that nowadays several prominent scholars attest that classification and indexing are regarded as identical with the organization of knowledge and, consequently, they can be considered a subfield of the science of science[7].

If we take the ultimate goal for library activities to collect and organize knowledge rather than documents, Karl Popper's "world of objective knowledge"[8] offers a suitable theoretical foundation for discussions of the theory and practice of indexing and classification.

In *Popper's ontological scheme* World 1 represents the physical objects; World 2 is the mental world covering subjective human knowledge or mental states; World 3 stands for the objective knowledge, which is the product of the human mind and is recorded in all the artefacts man has stored and scattered around on earth. It is the world of ideas, art, science, language, ethics, institutions – the whole cultural heritage, in short – encoded and preserved in such World 1 objects as brains, books, buildings, machines, films, computers, pictures, records, etc.

Although the three worlds are interacting with each other in various forms, World 3 represents an autonomous and independent world where knowledge is no longer subjective, belonging to the knowing subject who created it, but is existing objectively in recorded form and potentially accessible for subjective use to all who want to share it.

With reference to Popper's three worlds, and to the awkward and delicate problem of gaining "objective knowledge" from written works, a book represents not just a physical entity (in World 1), but it is a composite, poly-dimensional intellectual product (in World 3) which can hardly be put into a limited physical mono-dimensional system occupying only one linear position.

As Manfred Kochen states, "the modern book has a dual nature"[9].

It is, on the one hand, a physical object such as a scroll, a bound copy, or a microfilm strip. It is, on the other hand, an abstract, mathematical object which is invariant under transcription from one form of physical recording to any other. In this sense, a book is analogous to the mathematical object we call a line, most often hairpin bended and crossing but with a zero width, and of which pencil marks on paper suggest, through a reader's eyes, what the communicator intends.

The burden is on the reader to ascertain a subjective value of each curve and crossing point. Classification means to provide adequate access to potential pieces of information which may become of use, that is, give meaning and mental activity. In such a system "whales" might be grouped with "fish", although this relation is "false" from a generic stand, but "true" from a locational view.

The traditional library may thus be concerned primarily with the collection, preservation, physical organization and use of books as physical objects and they share with museums and monuments the duty of our cultural heritage. But the rapid changes in literature and in the demand for library services, accompanied by the equally rapid changes in technology, constitute an outside pressure on libraries. A new institutional form, which we might call the modern library, is emerging and is primarily concerned with the content, abstracted from any physical form. This institution is of central interest to further developments of classification theories and to the improvement of bibliographic work in today's libraries.

To master the complex problem of access to recorded knowledge in the form of potential information requires of the classificationists a deep understanding of the nature of knowledge itself and an appreciation of its rôle in society. The latter implies a common and dynamic code applied between knowledge providers, mediators and users.

Ranganathan's methodology can be seen as a solid investment in such a code.

To assist in the understanding of the need for several types of representations and types of access in the concept of "modern library", Wormell has suggested the introduction of the tripartite conception of accessibility, such as *physical, bibliographic and intellectual*[10]. Her assumption is that

access to library records is not synonymous with access to the contents of document items represented by the bibliographic records; one can gain bibliographic access to records without necessarily gaining intellectual access to the information contained in the document items. According to the conception, the aim of the traditional classification schemes in libraries is to provide physical access to their collection, that is, to place books in a sequence which enables the user to find, in a certain linear place, books treating the same subject. Therefore, traditional classification schemes are best regarded as having a function designed to physically bring together material likely to be required at the same time. The limitations of existing classification schemes as retrieval devices, arise from the fact that these systems are based upon the book as a physical entity.

This gives rise to the question: is it on the whole possible to produce classification schemes for the discovery and presentation of potential information on the contents of documents, i.e. to provide intellectual access?

The answer is affirmative if we allow generated knowledge entities to be structured in such a way that they can interconnect and intermingle across their physical carriers in time and space. This implies to accept several classifications for one physical entity. Two basic problems exist. First, the amount and nature of the intellectual efforts required to produce such structures. Secondly, the degree to which the concept of "meaning" should be involved in the classification process.

In the following some of the recent developments will be highlighted, and especially the techniques, experiences, and ideas arising from them which are likely to effect a fundamental change in the notion of subject representation and bibliographic control. All of them have a shared basis in Ranganathan's theories on the use of faceted frames for the analysis and synthesis of data representing the real world elements and relationships between them.

4. Conceptual models related to faceted methodology

Two mainstream approaches can be observed associated with faceted categorisation of recorded knowledge:

1. the library and information science (LIS) approach;
2. the linguistic approach.

The first approach is on the one hand concerned with attempts to automate classification of document contents, and on the other hand suggesting the application of faceted techniques to request modelling in knowledge-based retrieval, section 4.1 and 4.2 below.

The second relationship to faceted analysis and synthesis is found in text analysis based on linguistic rules. This approach has very recently been applied to the development of new IR techniques, section 4.3.

4.1 Knowledge classification and thesaural structures

The LIS research into knowledge classification reaches far back in history, as pointed out by Vickery[11] among others. The faceted dimensionality of document contents enters with Ranganathan in the forties and fifties with the PMEST structure[2] and has indirect influence on the idea behind Doyle's "semantic road-maps" in 1961[12] and direct effects on the theories applied by the Classification Research Group, UK, leading to domain specific schemes e.g. in natural science[11] and more universal in the Broad System of Ordering[13]. The notion of *structure* is essential, since it implies both a specific *notation* consisting of meaningful facets, also viewed as "rôles" or "cases", and *concepts* related to the various facets/rôles/cases. To a reader, the faceted structure forms a *context,* providing a better understanding of subject matter than enumerative systems do. Because we are dealing with classification, a *specific sequence* of facets must be kept, e.g. decreasing complexity or decreasing concreteness, in order to form the structured string of concepts making out the final classification of each document entity.

However, in our opinion the most interesting feature in faceted theory is not concerned with the *sequence* of facets, but the facets themselves. With a faceted structure at hand, it is possible to move across the various document entities either by means of a concept chained to a specific facet/rôle/case, by combination of several concepts and their specific facets – implying a particular conceptual relation – or by means of concepts alone. The structured string of concepts constitutes a context which may be regarded as representing the various aspects of the contents of a document entity. By making each concept in the string searchable, we are in close range of actually *indexing* the document in a structured way.

In attempting to automate this type of classification certain problems arise. Although it is possible to analyse a document in a way similar to linguistic text analysis, whereby the end product of the process consists of several faceted strings of concepts, a complex set of rules and a well structured vocabulary is needed to synthesise, for example in the form of generalisation or compression. Without profound success Sparck-Jones[14] tested several methods in the beginning of the seventies. Very recently, classification expert systems based on facet theory are under development, but have not overcome the problem of synthesis. This issue is attempted to be solved in a semi-automatic way, simply by letting the classification expert system suggest

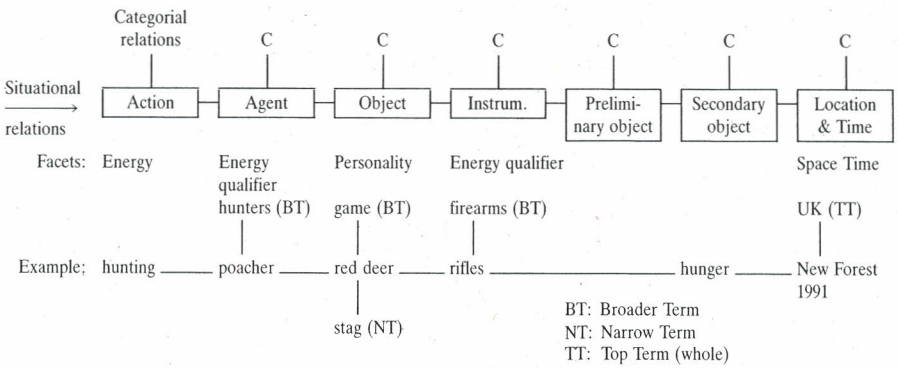

Fig. 1. Conceptual model of Categorial and Situational relationships between terms, and the similarity to some of Ranganathan's facets, based on a model by Ingwersen and Strunck[18].

potential faceted concepts in the context to be validated, modified or rejected by a human classifier[15].

One of the reasons for the success of (human) faceted classification in the natural science and technology disciplines is the *situational nature* of these fields. A document on, say, "hunting rifles" has a high probability of also containing potential information on aspects like "wild life" and/or "hunting" or "rifle manufacturing". As originally suggested by Ingwersen[16] and further developed based on psychological studies[17], one may classify knowledge of the world in the human memory into "categorial" and "situational" relationships. The categorial relations encompass, for instance, generic, property, part-whole and synonymous relationships. Situational classification deals basically with activity or processes (events), mainly guided by concepts representing "action", i.e. verbs. The correspondence between human memory classification in general and the resulting generated knowledge in documents, very often in the form of events and processes, emphasizes the adequacy of applying faceted techniques as a means of representation.

The two types of categorisation seem interconnected in the human mind in Long Term Memory, forming part of both semantic and episodic memory[17]. If we regard a *thesaurus* as a representation in recorded form of relationships between concepts in a domain, i.e. its semantics, we may easily observe the categorial relations (broader, narrower, etc.). Further, a close examination of the so-called "related terms" will easily reveal that these terms basically consist of concepts in situational relationships to the leading term. Also observable are the various facets/rôles/cases introduced by Ranganathan, see Figure 1.

However, one must have in mind that a thesaurus does not refer directly to

any document contents. It is either a tool for assigning keywords during indexing, i.e. to *discriminate* between various documents in contrast to classification, or it serves as a feedback device during searching in the form of a semantic representation of term relations in a domain.

The structure, Figure 1, is based on the surface grammar structure in Sanskrit[19] and also influenced by Ranganathan, combined with traditional categorial classes. For each facet/rôle/case in the situational classification, e.g. Action, Agent, Object, etc., categorial relations may exist, e.g. "game-red deer-stag". The "Preliminary object" signifies the source in the action, e.g. "bricks" in case of "house construction", where "house" is Object and "construction" Action. "Secondary object" implies reason/goal for the activity, e.g. "hunger" or "entertainment". In Sanskrit, each of the situational rôles would carry a significant suffix. In the figure, the PMEST facets are associated to the analogue rôles or cases with the Agent and Instrument rôles as Action qualifiers in a faceted scheme. The reason for placing Action as root in the model originates from linguistic traditions, foremost expressed by Fillmore[23].

Hence, we may look at the model, Figure 1, as a thesaural structure of term relationships – a semantic map in three dimensions. Two dimensions are shown explicitly on the figure. A third dimension is constituted by topical connections to other situational strings of concepts containing some but not all of the same terms and perhaps, but not necessarily, in identical rôles. For instance, all the terms in another string might be identical except "poacher", substituted by "hunter" (agent) and "hunger", substituted by "sport" or "entertainment" (secondary object). The two resulting strings of structured concepts would consequently signify two slightly different topics.

However, one might also regard the model solely as a faceted structure meant to classify document contents. Two documents, exemplified by the two above mentioned connected strings of concepts, would then be rather closely related, i.e. "near" in document space: one document on "poaching", the other on "sport hunting". A third document, say on "rifle manufacturing in 1991", would be more loosely connected since "rifles" would change into the Object rôle (Personality facet) and most of the other rôles' concepts in the context would be completely different. The only link might actually be the concept "rifles" itself, in addition to "1991" in the Time facet.

From the examples shown the problems of generalisation in automated faceted classification become evident. Without a thesaurus of categorial relationships there is limited possibility for relating "poaching" to "hunting" (or "poacher" to "hunter"), the former as a special instance of the latter.

4.2 Facets as instruments in request modelling

By relaxing the principle of "helpful sequence" and instead concentrating only on the facets or rôles themselves, Ingwersen has suggested to apply the facets to structure modelling of requests for information in IR. Either by an information specialist or librarian during search interviewing, to which the model Figure 1 originally was intended[18], or by a knowledge-based intermediary mechanism[17].

It must be stressed that we are *not* talking about *classifying* user requests. In contrast, the suggestion is to make deliberate use of the faceted (or case) structure when posing questions to the user about his problem area and/or substance of the information need. In particular, when a knowledgeable user at the initiation of the interaction with the system formulates his request in the form of a *label*[20, 17], that is, one or two terms or concepts only, the system (or librarian) may attempt to extract other relevant concepts from that user by having the faceted structure in mind.

Basically three methods exist for this extraction. The two first methods belong to the so-called "intelligent" approach to knowledge-based retrieval. The third one adheres to the class of supportive retrieval systems[17].

1. The intermediary contains a faceted or similar structured knowledge-base or thesaurus which is displayed as feedback, centered around the concepts given by the user. The user may pick or modify concepts. Instead of displaying conceptual structures, the intermediary may pose questions as to the relevant faceted relations, logically leading *from* a recognized concept. This latter method has actually been applied in the Plexus Project which is based on a domain dependent BSO-structured conceptual map[21];

2. The intermediary does not possess such a thesaurus or not even a morpho-syntactic dictionary, or the terms are not recognised. It may hence pose questions about the topic, structured according to the particular domain (s) covered by the system. For instance, the user initially chooses the "hunting" domain; by domain and task analysis made prior to its design the intermediary mechanism knows that some dedicated structure consisting of "action", "object", "instrument/tool", "agent", and "location" conceivably may exist. However, the intermediary is ignorant of the actual concepts carried by these facets. Depending on the physical lay-out, the interface may provide a facet or case frame explained properly, to be filled by the user with concepts constituting the request. In Boolean logic this method allows the interface to OR "agent" and "action" concepts and ANDing other rôle concepts. Another way is to allow the user first to state his request in natural language. By surface text analysis, i.e. simple use of prepositions, the mechanism itself attempts to place the concepts in appropriate facets or cases. For

example, the user enters "poaching in Wales". The mechanism may ascertain "Wales" as "location", but does not know whether "poaching" belongs to the "action" or, say, "instrument" rôles. This might be verified and further questions asked to extract other faceted concept. Obviously, the faceted questioning is solely applied in order to make the interaction move on. The facets themselves are not used for retrieval purpose, since (or if) the related information source does not possess such structures. It must be noted that Tome Searcher makes use of natural language input in the described way. The difference lies in the way Tome asks questions in case of non-recognition of the concepts entered. It asks categorial questions, such as instance of or property related questions – not situational or faceted ones.

3. The intermediary mechanism is designed on top of a facet structured information source – the database(s) underlying it. Mark Pejtersen's Bookhouse system[22] makes use of facets which are generated from a vast cognitive task analysis of the actual domain: fiction literature. Furthermore, the Bookhouse facets are influenced by Ranganathan's theories, but the *facets are self-contained in a multidimensional way*. In the Bookhouse the contents of each novel is represented by "personality" and "matter" (what the story is about), "energy" (the plot), "space" (where does the story take place), etc. in addition to other domain dependent facets or rôles, such as "secondary object" (author's intent). Because of the intensive domain and task modelling which took place prior to design, the system is "user driven". By means of icons and symbolic metaphors the user himself finds his way into a "strategy selection room" where one of the strategies to choose exactly makes use of these multi-dimensional representations. Then, the user finds himself in a new "space" or room with images representing the dimensions. So, for example, "time" is naturally a clock, "location" is a globe, etc. In this system it is the users who pick, browse and taste the contents of facets, they themselves normally would use to describe fiction. The Bookhouse is not "intelligent" – but intelligently designed as a supportive IR system.

4.3 Linguistic text analysis

One may also view the model, Figure 1, as an instrument for text analysis. It is remarkable to observe the similarity between Ranganathan's or Vickery's faceted schemes, and Fillmore's linguistically based case grammar[23] or Lindsay and Norman's LSD scheme[24]. Obviously, some similarity ought to exist since facet theory and classification in the LIS approach also is concerned with generated knowledge contents in documents, mainly consisting of text. By dealing with cases/rôles, more or less identical to facets, linguistic text analysis might indeed be applied in the analysis phase. However, a difference

exists between a linguistic approach to sentences as entity and the LIS classification approach to much larger text entities: the demand for synthesis and generalisation is already outlined above.

Instead, linguistic text analysis, for instance on syntactic level forming case structures in a context, ought to be applicable for the purpose of indexing. This has been attempted several times, e.g. in the Precis system and by Spang-Hanssen[25]. This seems attractive, also because user requests in natural language can be analysed, and text structures in the system be matched with user requests. However, again a basic difference exists between text linguistics and information retrieval. In text linguistics one would prefer to *understand the meaning* of a given text, e.g. for translation purposes. In IR the aim is to provide relevant information to a user. The question is: do we have to understand the meaning of the user's request and the meanings embedded in documents?

Interestingly enough, the answer is negative. In IR information goes beyond meaning. van Rijsbergen has very recently demonstrated the issue[26], p. 34–35:

> Usually a query is an expression of a lack of information, a retrieved document is intended to fill that information gap. In the theory of IR that I am proposing, sentences do not have to occur in the documents to be used to describe a document. A limited case would be to identify a document with all the sentences that are true of it. This would mean that at any one stage we only have a finite set of sentences describing the document, *but a potential for finding further sentences.* A sentence would then be *about* the document rather than in it. In this way, one would emphasize the *informativeness* of a document. Information is ultimatively dependent on the interpretation the user puts on a meaning, a logic is a tool which a user can use to get at the information.
>
> In IR we do not seek an answer to the *meaning* of language, instead we seek a model that will enable the user to find information, that is, discover something she did not already know. Simply retrieving meanings is not enough; these meanings must carry *information,* hence this form of retrieval is inherently uncertain.

Our interpretation of van Rijsbergen's model makes it possible to see a document as a *kernel of sentences,* all true of it, with a space around it consisting of *explicit and implicit "semantic values"* representing the potential information in the document[17]. The "explicit" values are those which directly refer to meanings of sentences true of the document – as in the example "time flies like an arrow". By means of syntactic (or faceted) analysis we would encounter 4–5 explicit semantic values. The "implicit" semantic values are those that are *not present* in the text, but could be, from a logical (e.g. faceted) point of view. For each "explicit" set of values there will consequently exist sets of "implicit" values. It seems reasonable to assume that van Rijsbergen basically is interested in the "implicit semantic values".

This idea of "implicit" semantic values can be seen to underly the assumption proposed earlier in this paper about *structured questioning* by intermediaries and the use of *situational classification* in human memory. The example used was within the hunting domain where the text (in the user request) represented by the sentence "poaching in Wales", in addition holds the implicit potentiality of being about "game in Wales" – and consequently also potentially being about "red deer" and "stags in Wales" – *without actually containing* these terms, rôles or sentences.

Obviously, the opposite case also exists, namely that a sentence in a document contains "stags in Wales" and therefore potentially can be about "poaching in Wales". van Rijsbergen suggests to apply the logical uncertainty principle as a means to retrieval. It becomes then a question of "nearness" in the document space whether the text on "stags in Wales" (or "red deer" or "game") is found or not – starting with "poaching in Wales". This is the *Dark Matter Problem* in information retrieval which in addition includes what the searchers might "read into" texts or pictures, that is, *information.*

The problem to face is to produce sufficient *context* to retrieve e.g. "stags in Wales". This can be done through IR interaction, for instance by structured request model building, combined with the logical uncertainty principle – to a certain degree. Then the user must take over.

In a slightly different direction Smeaton et al[27], under the umbrella of the Esprit II project SIMPR, apply *morpho-syntactic* analysis to produce the so-called TSAs (Tree Structured Analytics) from document texts or user requests. The TSAs are *deliberately* made uncertain in a semantic sense, and may thus represent possible (explicit) semantic values for each sentence analysed. The difference between van Rijsbergen's theory and Smeaton's TSA analysis is that the former makes use of syntactic analysis to produce terms or concepts related to rôles, cases of facets. "Nearness" may consequently be measured in terms of either concepts or corresponding rôles, as well as conceivably missing rôles, or all together. Hence, implicit semantic values are retrievable. Smeaton does not make use of all possibilities related to syntactic analysis, that is, the cases. He is satisfied with whether terms may be nouns, adjectives, verbs, etc. Consequently, hidden semantic values may not be easily retrieved. The difference may be illustrated by the sentence "the letter arrived yesterday" in which the grammatical subject (the noun "letter") is not equivalent to the conceptual "agent", containing the implicit value "person(s)", who wrote the letter. The TSAs will not reveal the potentiality of this implicit semantic value.

It is the authors' opinion that both methods of representation are applicable to the *plausible inference networking technique,* proposed by Turtle and Croft[28], and presently undergoing testing. This technique may best work

simultaneously on *several* types of representations for the same document and *several* user request versions for the same problem or need.

Again, this technique calls for request model building through IR interaction.

Finally, the authors wish to point to a domain in which Ranganathan's ideas, as well as related theories, seem to be of profound use: *software reuse.*

The GTE Laboratories and Contel, USA, makes deliberate use of Ranganathan's theories, comprising six user-derived facets, in order to structure the company's prototype software information system[29].

The process of facet analysis was guided by *interviewing potential users* of the classification system, i.e. software developers, who also supplied the test terms in the facets. Hence, the resulting faceted classification scheme demonstrates what Mills introduced as the "importance to the user criterion" for establishment of a helpful sequence, replacing the "concreteness" principle[30]. This utility approach resulted in a sequence with the "function" facet (=Energy) as root facet. The other user-derived facets are: Component Object; Medium; System Type; Functional Area; Setting. The latter three facets are system or application-oriented facets, whereas the former two facets cover technical aspects of the software component. In addition, the scheme is supplemented with five reusability attributes for the purpose of evaluating the effort of component integration, e.g. according to size, programming language and user experience level.

In an interesting review of IR functionality applied to software reuse Albrechtsen points to another related method used in this area, the case-based approach[31]. The British Eclipse project[32] makes use of case-frames based on the idea of conceptual dependency, originally put forward by Shank in 1972[33]. Their categories for software consist of: Conceptual Actions, i.e. fundamental functions the component covers, e.g. sorting; Nominals, i.e. objects that are manipulated or produced, e.g. lists; Modifiers, i.e. descriptors of actions or objects, e.g. quick (in sorting).

Albrechtsen points to the fact that the Eclipse Project also refers to the GTE Lab. Project, implying a conceivable analogy between the two different software reuse classification systems and Ranganathan's fundamental categories – see Figure 2.

Fig. 2. Mapping of facets in different re-use systems. From Albrechtsen.[31]

Eclipse's semantic primitives:	*GTE Lab. facets:*	*Ranganathan:*
Nominals	Objects, Medium	P, M facets
Actions*	Function*	E facet
Modifiers	System Type, Functional Area, Setting	S, T facets

*: selected root in helpful sequence

Albrechtsen is fully aware that this mapping of similarity between the schemes, Figure 2, is rather crude and that overlaps may occur between the fundamental categories in practical classification situations.

Concluding remarks

It is the authors' opinion that Ranganathan's principles for knowledge classification are valid to pursue further. In particular, the discussion demonstrates the suitability of the faceted categorisation, not only for textual documents but also with other forms of carriers of information. Faceted categorisation may provide multi-dimensional and hence structured entry points to document contents, and thus give intellectual access to generated and stored knowledge.

Although the faceted methodology seems adequate, certain drawbacks are envisaged, especially associated with Ranganathan's criterion of helpful sequence and in relation to synthesis and generalisation issues in classification. Nonethless, without doubt, his principles, and in particular his and related facet schemes, will be applied to neural network techniques in the near future. The reason being that this technology attracts pattern recognition problems, and faceted classification demonstrates exactly such problems – although of very high complexity.

More promising scenarios for application of his principles at present and in future are within the knowledge representation and information retrieval areas. The reason is the *structural approach* in his principles. This fundamental approach makes it possible to apply his and related methodologies for structured categorisation in knowledge-based IR interaction, for instance as a platform for structured questioning of users.

The paper emphasizes the close relationship between Ranganathan's principles and case grammar and case-frame theory and application, for example

in relation to retrieval technique developments, contextual IR theory building, and software reuse systems.

The authors have not questioned the universality of Ranganathan's fundamental facets, but prefer to view them as eventual starting points in cognitive modelling and work analyses of particular domains, prior to the design of systems. As such they become relative to specific domains. An interesting result of such analyses may be to compare text-derived facets and facet priorities with user-related facets and preferred sequences, in order to reach a consensus for retrieval purposes.

We are confident that this and the other issues associated with and influenced by Ranganathan's ideas discussed in this presentation, will appeal to his dynamic way of thinking.

References.

1. Srivastava, A. P.: *Ranganathan: A Pattern Maker.* New Delhi: Metropolitan Book Co., 1977.
2. Ranganathan, S. R.: *Philosophy of Classification.* Copenhagen: Munksgaard, 1951. (Library Research Monographs; University Library, Copenhagen)
3. de Solla Price, D.: Personal communication with Derek de Solla Price during his visit to Stockholm, 1982.
4. Shera, J. and Egan, M.: *Foundation of a Theory of Bibliography.* 1952.
5. Stokes, R.: *Bibliographical Control and Service.* London: Deutsch, 1965.
6. Greg, W.: *Bibliography – A Retrospect.* London: The Bibliographic Society, 1945. (The Bibliographic Society 1892–1942: Studies in Retrospect)
7. Dahlberg, I.: "New trends in classification." In: *Taylor, P. J. (ed). New Trends in Documentation and Information.* Proceedings of the 39th FID Congress, Edinburgh, 1978. London: Aslib, 1980. (FID Publication; 566) p. 214–222.
8. Popper, K.: *Objective Knowledge: An Evolutionary Approach.* Oxford: Clarendon Press, 1972.
9. Kochen, M.: *Integrative Mechanisms in Library Growth.* Westport, London: Greenwood Press, 1974.
10. Wormell, I.: *Subject Access Project – SAP: Improved Subject Retrieval for Monographic Publications.* Lund: Lund University, 1985. Doctoral thesis.
11. Vickery, B. C.: *Classification and Indexing in Science.* London: Butterworths, 1975.
12. Doyle, L. B.: "Semantic Road Maps for Literature Searchers." *J. of the ACM* 1961, 8 (1961): 553–578.
13. Coates, E. J.: *BSO Referral Test.* FID/BSO Panel. Final Report. Den Haag: FID, 1983.
14. Sparck-Jones, K.: "Automatic Classification." In: *Maltby, A. (ed). Classification in the 1970's: A Second Look.* London: Clive Bingley. p. 109–225.
15. Sharif, C.: *Developing an Expert System for Classification of Books Using Expert System Shells.* London: British Library, 1988. (BLRD; 32).
16. Ingwersen, P.: "Cognitive Analysis and the Role of the Intermediary in Information Retrieval". In: *Davies, R. (ed). Intelligent Information Systems.* Chichester, West Sussex: Horwood, 1986. p. 206–237.

17. Ingwersen, P.: *Intermediary Functions in Information Retrieval Interaction.* Copenhagen: Samfundslitteratur, 1991. Doctoral thesis.
18. Ingwersen, P. og Strunck, K.: *Informationsprocesser: Emners Beslægtethed, Brugerkommunikation, Informationssøgning.* København: Danmarks Biblioteksskole, 1980.
19. Ingwersen, P., Johansen, T. and Timmermann, P.: "A Study of the User-Librarian Negotiation Process. In: *EURIM 2 Conf. Proceedings,* Amsterdam, 1976. London: Aslib, 1977. p. 203–207.
20. Ingwersen, P.: "Search Procedures in the Library Analysed from the Cognitive Point of View. *J. Doc.* 38 (1982): 165–191.
21. Vickery, A., Brooks, H. and Vickery, B. C.: "An Expert System for Referral: the Plexus Project". In: *Davies, R. (ed). Intelligent Information Systems.* Chichester, Sussex: Horwood, 1986. p. 154–183.
22. Mark Pejtersen, A.: "A Library System for Information Retrieval based on a Cognitive Task Analysis and Supported by an Icon-based Interface". *ACM Sigir Forum,* 1989, June, p. 40–47.
23. Fillmore, C. J.: "Case for case." In: *Bach, E. and Harns, L. G. (eds).* Universals in Linguistic Theory. New York: Holt, 1968. p. 1–88.
24. Lindsay, P. H. and Norman, D. A.: *Human Information Processing.* New York: Academic Press, 1977.
25. Spang-Hanssen, H.: *Roles and Links Compared with Grammatical Relations in Natural Languages.* Lyngby, Copenhagen: Dansk Teknisk Litteraturselskab, 1976. (DTL Skriftserie; 40).
26. van Rijsbergen, K.: "The Science of Information Retrieval: Its Methodology and Logic." In: *Conference Informatienvetenschap in Nederland.* Den Haag: Rabin, 1990. p. 20–38.
27. Smeaton, A., Voutilainen, A. and Sheridan, P.: "The Application of Morpho-syntactic Language Processing to Effective Text Retrieval." In: *ESPRIT '90 Conf. Proc.* Dordrecht: Klüver, 1990. P. 619–635.
28. Turtle, H. and Croft, W. B.: "Inference Methods for Document Retrieval." In: *Vidick, J.-L. (ed). ACM/SIGIR Conf. Proc.* Bruxelles: Bruxelles University, 1990. p.1–24.
29. Prieto-Diaz, R.: "Implementing Faceted Classification for Software reuse." *IEEE* (in press).
30. Mills, J. "The Classification of a Subject Field." In: *Proc. Inter. Study Conf. on Classification for Information Retrieval.* London, 1957.
31. Albrechtsen, H.: "Information Retrieval for Software Reuse" *Inf. Proc. Management,* 1992 (in press).
32. Wood, M. and Sommerville, I.: "An Information Retrieval System for Software Components." *ACM Sigir Forum,* 22 (3–4) 1988: 11–25.
33. Shank, R. C.: "Conceptual Dependency: A Theory of Natural Language Understanding." *Cognitive Psychology*, 1972, 3. p. 552–631.