Taxonomy of Novel Abstracts Based on Empirical Findings

Jarmo Saarti

Kuopio University Library, Finland

Jarmo Saarti works as a project manager at the Kuopio University Library in Finland. He has studied content representation of fiction and received his PhD in Information Studies at the Oulu University in 2000. For further information, please contact him at jarmo.saarti@uku.fi or visit his website http://www.uku.fi/~saarti/indexeng.html

Saarti, J. (2000). **Taxonomy of Novel Abstracts Based on Empirical Findings.** *Knowledge Organization, 27(4).* 213-220. 13 refs.

ABSTRACT: This paper examines the content description of fictional works, specifically novels. The data for the empirical part of the study was gathered in Finnish public libraries. The aim of the study is to find out how library clients and library professionals of public libraries describe novels by abstracting them – what are differences in their characterisations and what are the similarities between their abstracts. Also, a taxonomy of novel abstracts is given. The abstracts are classified into four categories in the taxonomy: plot or thematic abstracts, cultural-historical abstracts, subjective or personal abstracts and critical abstracts. In the final portion of the paper, a model for a search and retrieval system for fiction is presented.

Introduction[1]

Fiction retrieval is a special issue that has recently aroused much interest in information science. This interest is due to several factors, the most important being the need for fiction retrieval and developments in computer systems that have made it possible to create retrieval systems for fiction. Retrieval is now possible through the use of a multifaceted approach for analysing and describing fictional texts. On the other hand, it has been noted that when libraries computerise their collection, a large amount of material remains unindexed from the years of paper catalogues – and most of that unindexed material has been and still is fictional texts.

Furthermore, the gradual shift to the digital distribution of materials requires new tools for analysing the content of fictional materials as well as for indexing[2]. Indeed, it is difficult to retrieve texts and other materials that have not been analysed and classified and/or indexed in full text databases. Thus, the distribution of this kind of material becomes very difficult, if not totally impossible. One further aspect, which emphasises the need for content analysis of fictional materials, is that free text searches cannot be used efficiently when searching for fictional material. This problem becomes apparent if we compare, for example, a search for fiction with the search and retrieval of natural science material, where texts are usually very topical and unambiguous.

There are very few empirical studies into the content representation of fiction. This paper is based on a test situation, where the aim was to determine what kind of abstracts the library clients and library professionals wrote for novels. Abstracting is a good, and perhaps an even better way than indexing, of describing the contents of a novel. By abstracting, one can describe also complex thematic and chronological entities (cf. Pejtersen's idea of concrete indexing strings (1996, 25) and the idea of synopsis).

This paper summarises the parts of Jarmo Saarti's dissertation (1999b) that deal with content description of fiction done by abstracts. In this paper, the empha-

[1] A grant by Ella and Georg Ehrnrooth's foundation has made this paper possible. I am indebted to Dr. Ewen Macdonald for his help with my English.

[2] There are already quite sophisticated fiction retrieval systems in use in the on-line bookstores.

https://doi.org/10.5771/0943-7444-2000-4-213 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

sis is on the taxonomic aspects of content description, as well as on the information systems of fiction and how these taxonomies are based on the information process of fiction. The other part of the dissertation work concerns the consistency of the fiction content description, predominantly the indexing of fiction (see Saarti 1999b, 2000).

Information Process of Fiction

The primary participants in the information process of fiction are: the work of art, its creator (i.e. the writer), the reader, and the social-historical environment where the publishing and reception takes place (see fig. 1.). Due to the nature of fictional works, the reception of the work of art is not fulfilled unless all of the participants mentioned above are involved in the process. The role of the writer is to write works of art – novels, short stories, poems, plays etc. – which are to be published. The role of the work of art is to be the medium through which the artist can communicate with his/her audience. Furthermore, the work of art has its own, autonomous life: after the book is published, the writer can only have a role as one of the readers, for example, as an interpreter of the work.

The role of the reader is that of the interpreter of a work of art. The interpretation, as well as the creation of a work of art, takes place within a social-historical context that defines the language used and a work's

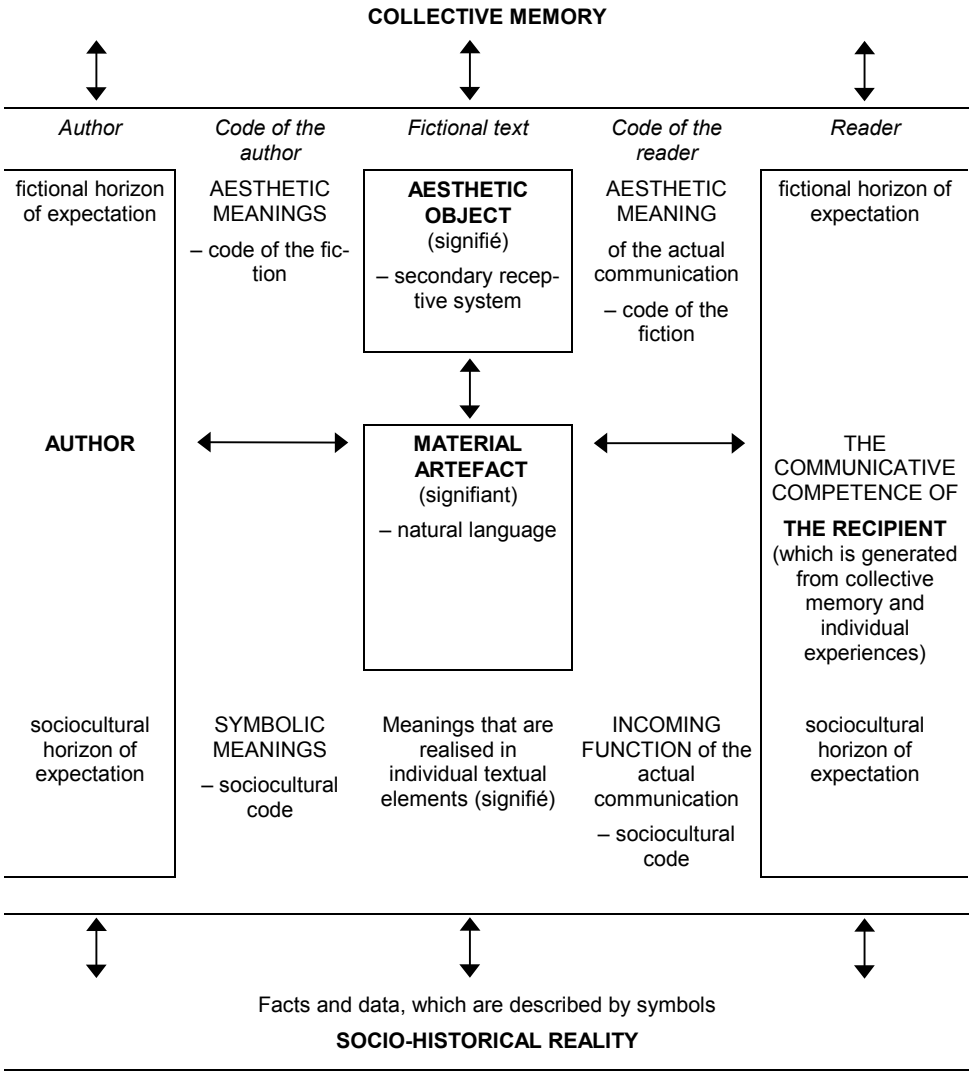

Fig. 1. Communication process of fiction. (Adapted from Segers 1985, 72 and Martens 1975, 36.)

https://doi.org/10.5771/0943-7444-2000-4-213 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

means of artistic expression. Without this common language, no communication between readers and writers could occur.

Fictional communication is typically twofold. It consists of factual meanings, such as references to actual happenings, historical and geographical facts (see e.g. Ranta 1991, 20-23). On the other hand, it has an aesthetic function and is thus based on the individual interpretation and reception of the work of art. This duality impacts the content description of fiction: on the one hand objective grounds can be identified, but on the other hand, some aspects are more likely to be subjective and thus personal, diverse and varied.

Research Questions, Data and Methods

The basic theoretical approach used in this study was qualitative: specifically, a grounded theory approach was used (see. e.g. Strauss & Corbin 1990). Since the aim of this study was to construct a taxonomy of the aspects used in the content description of fiction, both qualitative (mainly content analysis) and quantitative methods (mainly statistics) were used in the analysis of the empirical material gathered. This triangulative approach was chosen to confirm the findings of the study (see e.g. Strauss & Corbin 1990, 18-19, Breitmayer & Ayres & Knafl 1993).

Five novels from different genre categories were chosen for the experiment., which made it possible to analyse the differences between the content descriptions of different genre types, as well as the differences between serious (high-brow) and recreational literature. These were (their genres in parentheses):

- Dostoyevsky, F. M.: *Notes from the Underground* (serious, foreign).
- Kauranen, Anja: *Kultasuu* (*Golden Lips*) (serious, domestic).
- Mustonen, Enni: *Nostovara* (romance, the title is a pun and impossible to translate, domestic).
- Paasilinna, Arto: *The Year of the Hare* (humour, domestic).
- Simenon, Georges: *Maigret at the Gai-Moulin* (detective novel, foreign).

Altogether, 30 people were chosen as abstractors for the novels mentioned above. Abstractors were selected from five different Finnish public libraries: three patrons as well as three librarians or other library staff members from each library. Library staff had previous knowledge or working experience of the indexing of fiction. The patrons chosen were avid readers of fiction. The aim here was to determine how the role of the library professional affected the content descriptions and how literate individuals describe novels.

In order to ensure the comparison between different test subjects, they were all required to read the same novels. They read all the above books within two months (the books were all about 200 pages long). They were then invited, as a group, for a discussion and asked to complete the questionnaire. During the group session, they all abstracted[3] the above novels. The writer of this study was with them *in situ*. The whole process of abstracting lasted about two hours in each library.

First, the abstracts were analysed statistically, and the basic statistics were calculated for the abstracts. After that, a content analysis of the abstracts was completed with the aid of an observation matrix (see Miles & Huberman 1994, 207-224). The aspects on which the content analysis was based were derived from the model of fictional communication (see previous section) and from previous studies on the content description of fiction (see e.g. Pejtersen & Austin 1983 and 1984, Beghtol 1994 and Saarti 1999a, 1999b).

The abstracts were divided into the following categories for the primary content analysis:

- Descriptions of the novel's content elements (thematic elements, elements of the plot, personae, settings, time etc.).
- Descriptions of the novel's structure.
- Descriptions of the reading experience.
- Descriptions of the novel's place in the history of the fiction.
- Descriptions of the author.
- Descriptions of how the work was evaluated.

In order to generate analytical categories, the primary categories above were divided, according to the main participants in the communication process of fiction, into four categories (see fig 2). These include all the analytical elements in each category (see also Table 1.).

The categories were specified as follows. Descriptions of the novel's content elements included all the notes made by abstractors that referred directly to the content of the novel. These included the main aspects defined in previous studies, such as setting, time, personae, actions and thematic elements. Descriptions in this category can be seen as direct references to the

[3] The same novels were also indexed after abstracting them by all the participants, see. Saarti 1999b, 2000.

https://doi.org/10.5771/0943-7444-2000-4-213 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Descriptions of the novel's content elements and structure: – thematic elements, – elements of the plot, – personae, – settings, – time, – structure of the work.	Descriptions of the author: – personal history, – history of publications.
Descriptions of the reading experience: – interpretation, – evaluation of the work.	Descriptions of the cultural-historical continuum: – cultural-historical elements that have influenced the work, – intertextual connotations, – status of the work in the hierarchy of fiction.

Fig. 2. The aspects used in content analysis.

Category	Definition	Example
Descriptions of the novel's content elements	Aspects describing the contents of a novel: thematic elements, elements of the plot, personae, settings, time etc.	" Anja Korpela works in a little branch office of a bank, living in the same building with her two kids." (The book describes) "... loneliness and anguish of a man."
Descriptions of the novel's structure	Aspects describing the structure of the narration in a novel.	" The beginning of the book..." " Detailed narration and description of circumstances..."
Descriptions of the reading experience	Aspects describing the personal reading experience of a reader.	" ...boring reading for me." " ... very catchy, although it didn't have any special effect on me."
Descriptions of the novel's place in the history of the fiction	Aspects describing the status of a novel in the history of literature.	" (It reminded me of) Hercule Poirot's and Miss Marple's 'adventures'." "...similar to the Kerouac's work On the Road..."
Descriptions of the author	Aspects describing the author and his/her other works.	"... typical of the Anja Kauranen of the 80's." " I mix up in my mind the book and the movie and possibly all the other Paasilinna books."
Descriptions of the evaluation of the work	Aspects describing the objective evaluation of a novel.	" Nearly worthy of five stars and it was best of these five books I had to read." " Novel is a typical example of a classic that will last through the years."

Table 1. The categories, their descriptions and examples used in the analysis

plot and contents of the novel without deeper personal interpretation.

Descriptions of the novel's structure included direct references to the structural aspects of the narration. Descriptions of the reading experience included references to one's own *personal* experience of the reading process. During the actual content analysis, these two were the most difficult aspects. Indeed, the former caused difficulty because one can see the chronological listing of the content elements of a novel as a description of its structure. This type of description was evident here also; the majority of abstracts of this type listed these elements in the same order as they occurred in the novel. In this study, this problem was resolved by including only direct verbal references to the structure[4].

[4] In the actual indexing of fictional works this can be solved with the aid of logical, meaningful indexing strings or, as Pejtersen (1996) has presented, with the aid of concrete indexing strings.

https://doi.org/10.5771/0943-7444-2000-4-213 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

The latter, reading experience, was difficult in this analysis because it was sometimes difficult to discern the difference between reading experience and evaluation of the book. This was solved by including only those elements that could be considered objective – those elements based on objective facts such as the hierarchy of literature – in the evaluation of the book category. In these abstracts, this kind of evaluation was done by comparing the novel with other works, especially with classics. The reading experience category included those comments that the test subject named as his/her own, such as interpreting and analysing one's own life and tastes against the book.

subject, the length of abstracts varied from 38 to 186 words. The standard deviation among the abstracts was 27 words – this being similar for library professionals (25 words) and patrons (29 words).

The content analysis of the abstracts was completed with the aid of an observation matrix (see previous section). The contents of the abstracts could be analysed exhaustively using the categories mentioned above, (apart from a few comments about the actual reading process made by some test subjects). Thus, it seems that the aspects utilised here are sufficiently wide-ranging and exhaustive in the content analysis of novels.

% N=3206	Novel's content elements	Novel's structure	Reading experience	Novel's place in the history of the fiction	Descriptions of the author	Evaluation of the work	All
Library professionals	39,2	6,2	2,2	0,4	0,9	1,9	50,8
Patrons	35,7	5,7	3,3	0,5	0,7	3,3	49,2
All	74,9	11,9	5,5	0,9	1,6	5,2	100

Table 2. Frequency of the categories used in content analysis

References to the author and to the cultural-historical continuum were, on the other hand, easily distinguished, as when test subjects referred to the writer, literary school, other writers or historical period in the abstracts.

Some examples of the previously mentioned categories are listed in table 1. The most typical fragments are chosen as examples.

Abstracts describing the contents of novels

The corpus consisted of 150 abstracts. The test subjects were asked to abstract each novel in about 50-80 words. They followed these instructions rather well. The average length of the abstracts was 68 words – library professionals averaged 69 words and library patrons 67 words. Thus, there is no significant difference in the average length between professionals and patrons. However, the length of abstracts did vary greatly, both between different test subjects and also between abstracts written by a single abstractor. The longest abstract written by a library professional was 146 words, and the longest abstract by one of the patrons was 186 words. Similarly, the shortest abstracts were, respectively, 28 and 23 words long. For one

Table 2. shows the number of units of analysis in each category. One must be cautious in drawing conclusions, because the units were not similar. One category consisted of single words and another of sentences or even paragraphs (see also table 1.). With that proviso, one can obtain a clear picture of the elements' frequencies used in the different categories.

The use of the novel's content elements in abstracts was the most popular method. In addition, the library professionals used these content elements slightly more than the patrons. On the other hand, patrons used more evaluative elements in their abstracts. Aspects concerning the novel's place in the history of fiction and descriptions of the author were used least. This can be seen in Table 3., which describes the percentage of test subjects that used different aspects in their abstracts. All the abstractors mentioned the novel's content elements and structure in at least one of their abstracts. The other three categories were not used by all the test subjects, although the patrons used them somewhat more often than library professionals.

The different aspects used in the abstracts can be divided into four different categories (fig 3.). In these subdivisions, the aspects are divided according to their

https://doi.org/10.5771/0943-7444-2000-4-213 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

%	Novel's content elements	Novel's structure	Reading experience	Novel's place in the history of the fiction	Descriptions of the author	Evaluation of the work
Library professionals (N=15)	100	100	80,0	26,6	46,6	86,6
Patrons (N=15)	100	100	93,4	53,4	60,0	93,4
All (N=30)	100	100	86,7	40,0	53,3	90,0

Table 3. The usage of elements in different content analysis categories.

	From singularity to generality.	
Objective analysis of the novel.	Descriptions of the novel's content elements and structure.	Descriptions of the novel's place in the history of the fiction and references to the author.
Subjective analysis of the novel	Descriptions of the reading experience and subjective evaluation.	The objective evaluation of the work

Fig. 3. Taxonomy of the abstracts

singularity/generality and objectivity/subjectivity. Thus, one can see the same transitional process described above in these abstracts (see fig. 1.). This process proceeds from interpreting one's own reception of a novel's contents to a more universal reception of fiction as a holistic process.

The abstracts can be typified according to the previous four part table (fig. 3). In this study it was usual that each abstract had elements from more than one aspect. But one could also find almost pure examples of each idealised type of abstracts, which are:

1. *Plot- or thematic abstract that describes the structure and content elements of a novel.*
2. *Cultural-historical abstract that describes the status of a novel in author's works or in literature.*
3. *Subjective or personal abstract that describes person's own reading experience.*
4. *Critical abstract that criticises the work objectively.*

The following are examples[5] from abstracts that best exemplify these idealised types – the first is about Mustonen's novel, the next two from Dostoyevsky's and the last from Simenon's novel:

Library clerk, female, 41 years – plot- or thematic abstract

The book is about the everyday life in a little village in the countryside. The main character is a single parent who works in a bank. She has been divorced quite recently and she lives with her children in the same building as the bank. There is excitement in the book, when the bank gets robbed. The effect of the economic recession becomes evident when the bank is merged with a bigger bank, where the main character gets a new job. A small IT-firm moves into the former locations of the bank. Its owner is not often there, which arouses suspension. The most important things to the main character are children, family and work. The ex-husband tries to hang around, but he gets no response. There is also a romance, because in the end the bank clerk and the owner of the IT-firm fall in love with each other.

Librarian, male, 58 years – cultural-historical abstract

Psychological novel, a Russian classic, that represents the inner realism and precedes Freud's psychoanalysis as well as existentialism. In it, the inward-looking, anxious clerk comes face to face with the petty realities of everyday life. One can see in the novel the writer's turnaround from the

[5] Also note how the type of the abstract affects the actual text, vocabulary used and rhetoric of the text.

https://doi.org/10.5771/0943-7444-2000-4-213 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

socially radical, individual bourgeois towards the Christian mysticism. Considerations about morality and justification of the deeds and intentions – idealism – become more important than the actual experience of the social turning point. Painfully accurate description of the soul, modern angst, social taboos (e.g. prostitution). The clerk, the main character, is totally the opposite of the Gogol's satirically described minor clerks, as different as night from day!

Patron, female, 57 years – subjective or personal abstract

A bit frightening, like tearing apart layers from one's self. Familiar things could be found, not always that flattering. I wonder if there would have been less grimness, if the story had been written in a romantic tower: with the blue skies around, and free birds, no locked doors. I wonder how much society around rewrites us. Quite a lot. How much one must know in order to dare to dig deep into human soul and analyse its depths.

Patron, female, 53 years – critical abstract

Clear, compact novel, detective novel, that describes economically the appearance of the characters. The psychological tension is reflected finely and it maintains the reader's interest in the plot of the book. The end is surprising, although the title foreshadows the role of Maigret in the story. Dialogue is lively, one can even hear the voices in one's own ears. Guaranteed, lively thriller with a plot full of surprises.

Conclusions

One conclusion that can be drawn from this study is that the role of library professionals influences their abstracts. Library professionals most often composed plot/thematic abstracts. On the other hand, patrons most often produced the subjective/personal abstracts. Patrons used more of the different aspects/categories presented above. Librarians, for the most part, described the novels, while patrons, in addition to that, also described their personal reading process. It was also evident that the literary competence of the reader influenced the abstracts. The most literary oriented test subjects made the most references to cultural-historical and critical aspects.

Different readers can be seen to use radically different elements while describing a novel's contents and the aesthetic experience of reading it. Also, all of the important participants of the information process of fiction are found in the abstracts, including work, author, reader and the cultural-historical context. Indeed, the totality of fiction's aboutness must include the entire process of fiction and its reception, not only the texts and their contents. On the other hand, these latter two aspects can be seen as the most crucial factors in the content representation of fiction.[6] Fur-

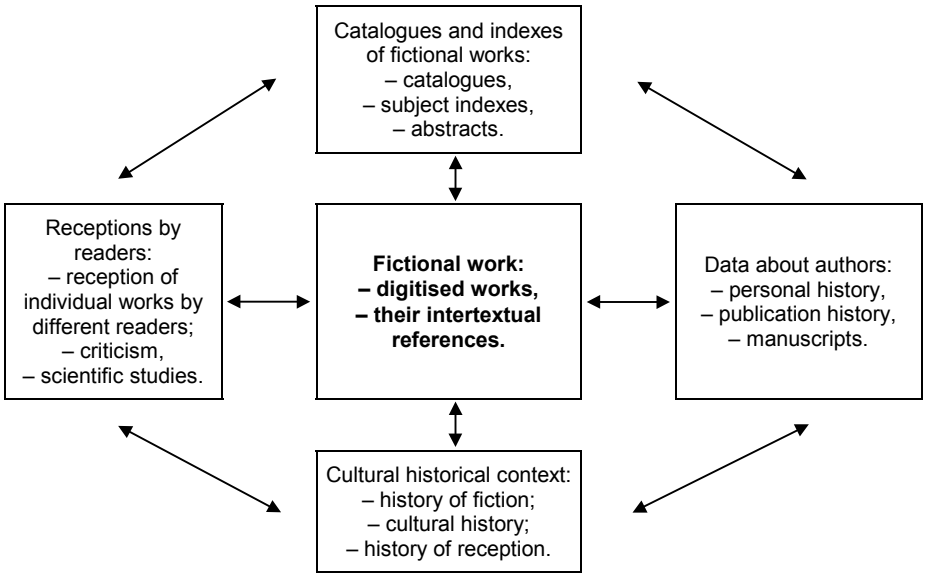

Fig. 4. A broad model for a search and retrieval system for fiction

[6] This was even more evident in the analysis of the indexing (see Saarti 2000). In this one could make standardised, typical indexing strings for each novel based on the statistical analysis of the indexing done.

https://doi.org/10.5771/0943-7444-2000-4-213 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

thermore, there is already some evidence (for example, from the internet bookstores) that subjective aspects are also important in fiction information systems. This, of course, is one topic that needs further study.

One clear result of studies that explore the indexing and abstracting of fictional works is the impact that the interpretation of a work of art has on the content description of that work. The impact of interpretation could be clearly noted in this study when analysing the abstracts. The structure and the content – those aspects in the novel worthy of description – were diffuse and manifold. The interpretative aspect of content description is a subject for further studies, not only in fictional works but also in scientific material.

It is apparent that fiction search and retrieval systems have to be multifaceted in order to meet the varying needs of different users. In Figure 4. a model for a search and retrieval system of fiction is shown. It consists of five main blocks (databases) that represent the different participants of the fictional communication system – works of art (texts), their subject indexing and abstracts, history of the reception by the readers, history of the writers, and cultural history. Aspects belonging to each of the blocks could be seen in the abstracts analysed in this study. Based on these findings, the most important block for the library is that of the catalogues and indexes of fictional works, including plot- or thematic abstracts of the novels. The other blocks should be generated by or together with the other participants in the information process of fiction, such as readers and writers.

With the aid of this kind of system, one can holistically document the different aspects of the meaning of fiction, for example, what the fiction is about. The building of these kinds of systems has been started, but there is much work to be done.

References

Beghtol, Clare (1994). *The Classification of Fiction: the Development of a System Based on Theoretical Principles*. Scarecrow Press, Metuchen.

Breitmayer, B. J., Ayres, L. & Knafl, K. A. (1993). Triangulation in qualitative research: evaluation of completeness and confirmation purposes. *IMAGE: Journal of Nursing Scholarship* 25(3). 237-243.

Martens, Gunter (1975). Textstrukturen aus rezeptionsästhetischer Sicht : Perspektiven einer Textästhetik auf der Grundlage des Prager Strukturalismus. In: Herausgegeben und eingeleitet von Hartmut Heuerman & Peter Hühn & Brigitte Rötger. *Literarische Rezeption : Beiträge zur Theorie des Text-Leser-Verhältnisses und seiner empirischen Erforschung*. Ferdinand Schöningh, Paderborn. (ISL, 4.) S. 23-49.

Miles, Matthew B & Huberman, Michael A. (1994). *Qualitative Data Analysis: an Expanded Sourcebook*. 2nd ed. Sage Publications, Thousand Oaks.

Pejtersen, Annelise Mark & Austin, Jutta (1983). Fiction Retrieval: Experimental Design and Evaluation of a Search System Based on Users' Value Criteria: Part 1. *Journal of Documentation* 39 (4). 230-246.

Pejtersen, Annelise Mark & Austin, Jutta (1984). Fiction Retrieval: Experimental Design and Evaluation of a Search System Based on Users' Value Criteria: Part 2. *Journal of Documentation* 40 (1). 25-35.

Pejtersen, Annelise Mark & Albrechtsen, Hanne & Lundgren, Lena & Sandelin, Ringa & Valtonen, Riitta (1996). *Subject Access to Scandinavian Fiction Literature: Index Methods and OPAC Development*. Nordic Council of Ministers, København.(TemaNord ; 1996:608).

Ranta, Judith A. (1991). The New Literary Scholarship and a Basis for Increased Subject Catalog Access to Imaginative Literature. *Cataloging & Classification Quarterly* 14 (1). 3-27.

Saarti, Jarmo (1999a). Fiction Indexing and the Development of Fiction Thesauri. *Journal of Librarianship and Information Science* 31(2). 85-92.

Saarti, Jarmo (1999b). Kaunokirjallisuuden sisällönkuvailun aspektit: kirjastoammattilaisten ja kirjastonkäyttäjien tekemien romaanien tiivistelmien ja asiasanoitusten yhdenmukaisuus. (English summary: Aspects of Fictional Literature Content Description: Consistency of the Abstracts and Subject indexing of Novels by Public Library Professionals and Clients.) (Acta Universitatis Ouluesis. B, Humaniora, 33.) University of Oulu, Oulu. (Diss..) (Digital version: http://herkules.oulu.fi/isbn9514254767)

Saarti, Jarmo (2000). Fiction indexing by library professionals and users. *Scandinavian Public Library Quarterly* 33(4). 6-9.

Segers, Rien T. (1985). Kirja ja lukija: johdatusta kirjallisuudentutkimuksen uuteen suuntaukseen. (Original title: Het lezen van literatuur: een inleiding tot een nieuwe literatuurbenadering.) Translated by: Lili Ahonen. SKS, Helsinki. (Tietolipas, 97.)

Strauss, A. & Corbin, J. (1990). *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. SAGE, London.

https://doi.org/10.5771/0943-7444-2000-4-213 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY