UNIVERSITY OF JYVÄSKYLÄ

TESTING MANUAL USABILITY

- A Task-based Assessment of the Usability of Two Manual Versions

A Pro Gradu Thesis

by

Heli Parkkinen

Department of English
2002

HUMANISTINEN TIEDEKUNTA
ENGLANNIN KIELEN LAITOS

Heli Parkkinen
TESTING MANUAL USABILITY – A Task-based Assessment of the Usability of Two Manual Versions

Pro gradu –työ
Englantilainen filologia
Huhtikuu 2002 78 sivua + 6 liitettä

Tutkielman tarkoituksena on tutkia tekijöitä, jotka vaikuttavat tekstin käytettävyyteen, ja selvittää, kuinka käytettävyyttä voidaan parantaa. Tekstin käytettävyyteen vaikuttavat luettavuuteen ja ymmärrettävyyteen vaikuttavat tekijät. Tällaisia tekijöitä ovat mm. sanaston ja lauserakenteiden käyttö, koherenssi, visuaalisuus ja tekstin asettelu. Tutkielma keskittyy teknisten tekstien, tässä tapauksessa käyttöohjeiden ymmärrettävyyteen ja käytettävyyteen. Tavoitteena on selvittää tekstin eri piirteiden vaikutusta käytettävyyteen siten, että kohderyhmä ja tekstin lukemistilanne otetaan huomioon. Käytettävyyttä tutkittiin käytettävyystesteissä, joissa testihenkilöt suorittivat tehtäviä tietokoneella manuaalin ohjeiden avulla. Testissä käytettiin kahta eri versiota samasta manuaalista. Testattavana materiaalina oli käyttöohje Republican kehittämään X-Fetch Wrapper Rule Generator -tietokoneohjelmaan.

Tutkimus kuuluu lingvistiikan ja kognitiivisen psykologian tutkimusalaan, jossa tutkitaan tekstin piirteitä ja niiden vaikutusta ymmärrettävyyteen ja käytettävyyteen, sekä ihmisten tapoja prosessoida tietoa. Tutkimuksesta on hyötyä tekstin käyttäjille, eli lukijoille, sekä tekstin tuottajille. Tutkimuksen teoreettisenä viitekehyksenä on kuvattu aikaisempaan tutkimukseen perustuvia hyvän kirjoittamisen sääntöjä, joiden avulla tekstin laatua pyritään parantamaan. Tutkimusmetodeina käytettiin observointia, talk aloud –menetelmää, ja kyselylomaketta. Testitilanne nauhoitettiin videolle. Tutkimusaineisto koostui kahdesta manuaaliversiosta, videonauhoille kuvatusta materiaalista, testihenkilöiden kommenteista ja muistiinpanoistani testitilanteen aikana, sekä kyselylomakkeen palautteesta.

Tutkielmassa on kolme tutkimuskysymystä. 1. Millaisia vaikutuksia tekstin eri piirteillä on käytettävyyteen? Hypoteesini oli, että tekninen teksti, joka on kirjoitettu kontrolloidun kielen ja Information Mapping® sääntöjä noudattaen, on lukijalle ymmärrettävämpi kuin ilman sääntöjä kirjoitettu, vapaamuotoisempi teksti. 2. Mikä on optimaalinen tapa esittää informaatiota ns. noviisilukijalle, joka ei omaa juurikaan taustatietoa asiasta? Hypoteesini oli, että teksti, joka on kirjoitettu selkeästi, ja joka sisältää yksityiskohtaisia ja visuaalisia kuvauksia on noviisille hyödyllisempi ja ymmärrettävämpi kuin teksti, joka sisältää vain välttämättömän informaation lyhyesti ja ytimekkäästi kuvattuna. 3. Sisältääkö manuaali kohtia, jotka vaikuttavat käytettävyyteen negatiivisesti? Mahdolliset manuaalin suunnitteluun liittyvät ongelmat selvitettiin käytettävyystesteissä, kun manuaaliversioita käytettiin siihen tarkoitukseen, johon ne oli alunperin suunniteltu.

Käytettävyystestin tulokset osoittivat, että manuaaliversio A, joka oli kirjoitettu kontrolloidulla kielellä ja hyvän kirjoittamisen sääntöjä noudattaen, oli lukijalle käytettävämpi kuin toinen testissä käytetty manuaaliversio (versio B). Manuaaliversion A vahvuuksia olivat myös visuaalisuus ja yksityiskohtaiset kuvaukset. Kumpikaan manuaaleista ei kuitenkaan ollut optimaalisen käytettävä noviisilukijoille. Testi vahvisti näkökulmaa, että manuaalit pitäisi kirjoittaa räätälöidysti eri kohderyhmille. Lukijoille olisi hyvä tarjota eri apuvälineitä, kuten sanasto, indeksi ja online help –toiminto, joista lisätietoja tarvitsevat voisivat hakea apua. Ajoittain olisi hyvä käytettävyystestein varmistaa, että manuaali toimii teorian lisäksi myös käytännössä. Testissä kävi ilmi, että manuaaliversiot sisälsivät käyttäjän näkökulmasta paljon epäkohtia, joihin testitulosten perusteella suoraan voitiin puuttua. Tuloksista oli siis hyötyä tekstin laadun ja käytettävyyden parantamisessa.

Asiasanat: manual usability, accessibility, text quality, understandability, readability

Contents

1 INTRODUCTION

During the last decades functional texts and especially technical documentation have become an area of interest also among linguists. Research in the field of technical documentation has been expanding in the 1990's (see e.g. Pander Maat and Steehouder 1992, Steehouder et al. 1994), and interest in studies concerning the quality of technical documentation is still increasing.

According to Frase et al. (1985: 98-99) there have been two types of research on text design. The first type includes both psychological and linguistic research. These include experimental research on what makes a text easy or hard to understand. Behavioural science, that is, research on human information processing and cognitive processes, has been looking for weaknesses in human information processing, and investigating how text design could help them. The second type of research includes the work of text and graphics designers, who have been solving practical problems and creating guidelines for a better text design. My study is placed within the first mentioned research type, it is a study of textual features and their effect on the use and understanding of texts. It is useful for text producers and users (readers). The writers should know what to write and how in order to design usable texts for different audiences and purposes. Finding out the factors that affect usability improves the document quality, and makes the texts easier to use.

It is useful to study document usability, because a text with a high level of usability is fast to read and understand. Users' errors and frustration can be avoided by usable documentation, and also time is saved. In order to make usable documentation, the documentation must be planned, revised and tested carefully. Many believe that this is a waste of time and money, but according to Kaner and Pels (2000), the result is quite the opposite. By

improving the documentation companies avoid bad reputation, customers' mistrust and unnecessary calls to company's help desk. In other words, companies save money instead of spending it.

The purpose of my study is to explore the factors that affect document usability, meaning the factors that make a text readable, understandable and accessible, and therefore easy to use. The core focus in my study is the actual use of documents - I examine the use of manuals in practice, in a usability test. Two versions of a user's manual for Republica's X-Fetch Wrapper Rule Generator programme are used in the test. With the test I look for factors which would help the company to improve the design of their manuals, and examine possible differences in the usability of two manual versions.

Different documents are used for different purposes, therefore the context for using a document must be kept in mind when designing usable documents. I will focus on the use of technical texts, in this case user's manuals, because such texts are often considered difficult to understand. There are often difficult terms and complicated language structures in technical texts (see e.g. Klauke 1994). For an expert they might be self-evident, but for a reader who does not have much prior knowledge on the matter, such a text could be impossible to understand.

In this study the usability of two versions of a manual for a computer programme is tested. A technical writer (language professional) of Republica wrote the manual in co-operation with the product developers. The version A was mostly written following the rules of Information Mapping® (IMAP) and controlled language (CL). The Information Mapping® method was developed in order to help writers to write text that is easily readable and accessible. It was first introduced in 1965 by Robert E. Horn. It includes rules that are based on studies of how readers find and process information (Horn 1997: 3-A-1). Controlled language means a

controlled use of language in order to make the text easier to understand. In version A the way of presenting information is visual, and a lot of attention has been paid to the layout. The information content is the same in both versions, but in version A the user is "led by the hand", in other words, things are explained clearly and explicitly, and a lot of clues are provided for the reader on how to proceed with the manual. Version B is written in "normal" English – natural language, where the use of words or sentence structures is not limited. Also, not much attention has been paid to visual presentation of information in the manual. The usability of these versions of the same manual was tested in practice, when volunteer test persons used the computer programme with the help of the manual.

There are three research questions that my study intends to answer. Firstly, what is the effect of different textual features on manual usability? My hypothesis is that a text that is written according to the guidelines of controlled language and Information Mapping® is easier to read and understand, and therefore more usable to the end-user. Secondly, what is the optimal way of presenting information for a reader who is a novice in the area that the manual handles? My hypothesis is that a text that is written as clearly and simply as possible and which includes detailed descriptions is more usable for a novice reader than a text which includes only core information. The study provides information on what an optimal level for a novice reader is – how much information need be given? Thirdly, are there design problems in the manual? The aim is to find out whether there are, for example, errors in the content of the manual versions, or parts where information is illogically ordered. Such design problems have a negative effect on usability, and they can be found in the usability test.

The manual is usable, if the reader finds information quickly and is able to complete the task with the help of the manual. This would be an optimal situation. The errors made during the performance can provide important information on the potential design problems of the manual. There are

degrees in usability, and this study provides information on how usable the manual versions are. The results of this study are directly useful for the company, they can be used for improving the quality of the manual that was tested (manual version A was the latest version of a published customer document). This study can provide information on the human information processing and the effect of different textual features on it. Also, the results of this study can give valuable information for the use of document designers in general.

In chapter 2 the concept of usability and the factors affecting it are defined. Chapter 3 includes a description of how the usability tests were arranged and which methods were used for collecting information. Chapter 4 contains an analysis of the problems that occurred during the tests, and also suggestions of what can be done to improve the manual versions, so that the problems can be avoided. In chapter 5 the results of this study are discussed. Chapter 5 also includes suggestions for further research. Chapter 6 concludes this thesis by reviewing what kind of information the usability tests revealed.

2 THE USABILITY OF DOCUMENTS

People are often not willing to read technical documents. Carrol and Rossen (1987) have pointed out that learners at every level of experience try to avoid reading, so if there is a problem to solve or a new thing to learn, looking for information from a manual is often not the first option. Horton (1993 as quoted by Lahti 2000: 19) has described the way the users of software typically try to solve a problem: first they try and see what happens, then they ask another user, call a vendor, search the online documentation, and at last, they read the manual. It seems that a manual is used as the last resort. Some people are willing to look at the documentation before calling the helpdesk, but asking a friend or a colleague is certainly a popular option when solving problems. It is easier to formulate the problem for someone who is present, because a technical problem may be difficult to describe – and you need to know what the problem is when you call the helpdesk or start searching from the manual.

The seriousness of the problem may also have an effect on the method to solve the problem. If there is a serious problem needing a quick solution, it is more likely that the way to solve the problem is interactive, because searching for specific information from a thick paper manual is not the fastest option. Also, one reason for the unwillingness to read manuals could be the difference in learning skills and learning habits. There are different kinds of learners and thus not one type of learning method can be effective for everyone. Many people find learning in an interactive situation more effective. For learners who need interaction, electronic documents in the web (e-learning) could be an effective option. If the electronic documents or learning environments are structured, navigatable, well implemented and include a sufficient amount of metadata for the use of search engines, the electronic documents beat the paper manuals in usability. This is, for example, because the search engines can be used for finding specific information quickly. However, this study focuses on the usability of paper

manuals, because the material available for testing is paper manuals. Also, paper manuals are still the main publishing method, so the usability of paper manuals is of great importance.

Although there are claims that people do not read manuals, there is also evidence against those claims. According to the study which Kaner and Pels (2000) report, a large majority of customers claimed that they checked their manuals before calling for support. Further, Schriver (1997: 213) presents the results of a study on using documentation for household appliances. In the study 80% of the users said that they either scan the manual or use it as reference. However, if the users could not solve the problems with the help of documentation, they tended to blame themselves rather than documentation.

A company producing documentation could save a lot of money by paying attention to the quality of documentation and the smoothness of work-flow. When documentation is produced, there are many steps involved. In an example case a writer writes a draft document about how to install a device. An editor proof-reads it, and sends it back with questions about the parts that were not clear. The writer adds information to the document, and the editor checks it again, but there are still parts that are not clear. However, the time schedule is tight, and the editor has to send the text to be translated, hoping that the parts that he/she tried to make more clear are factually correct. The translator translates the document, but in the original text there are many parts that are not understandable (the translator is a language expert, not an expert in technical matters), and the translator has to ask the writer for clarification. The translator does not get an answer to all the questions before the deadline, so the document has to be published. The end-users, customers, have problems understanding the document's instructions, possibly make mistakes and even break some parts of the device, and start calling the company's helpdesk. If documentation was more usable than what people are used to, people would also be more

willing to use it. If reading a manual solved their problems easily, that would save the companies' time and money, and the end-users' nerves.

In my work as a technical writer I have gained professional experience on the usability of documents. Based on my personal experience at work and on the research on text usability, the most important factors affecting usability are the factors that make a text comprehensible, readable and accessible. These factors include, for example, vocabulary and sentence structures, text consistency and coherence, content accuracy and relevance, the level of information (detailed/core), text accessibility, layout and the context of text use. These concepts are clarified in the following sections.

2.1 Definition of usability

Duffy has studied usability issues when evaluating military technical manuals. His (1985: 116) view is, that there are four major components of usability: access, accuracy, completeness and comprehensibility. Access means the ease with which the readers can find their way to a certain page that is required in their job task. Information should be accurately presented and complete, but even if it is, it is not usable if it is not presented in a clear and understandable manner, that is, if it is not comprehensible. Comprehensibility is a usability factor that is tied to the interaction of the reader with the text. According to Duffy the comprehensibility of a text depends on the reading skill, the graphic interpretive skill, the technical knowledge of the reader, and also on situational variables.

Orna (1985: 19-20) states that a usual definition for a usable text is, that if the readers after reading the text can understand or do some things successfully (something that they had not been able to do before), the text is usable. She also points out that there are degrees in usability, a text can be very easy or very difficult or something in between for the users. Orna's (1985: 20) own definition of usability is that "a usable text is one that allows

a successful transaction to take place between user and maker". In the beginning the user's state of knowledge is unsatisfactory. The user gains access to the knowledge that the maker has structured to meet the user's needs. This way the user gets new information that he/she can apply for certain purposes.

The successful transaction that Orna refers to would be an ideal situation. In reality, however, a total success is difficult to achieve, because the views of the user and maker hardly ever match completely. Customer feedback is in an important role here, because by receiving feedback the writers get to know how well they have been able to picture the reader and the purpose of text use in their minds. There are different ways to achieve a satisfactory state of knowledge. If the learning process is smooth and the learning curve is short, the reader becomes maximally satisfied and will also have a positive attitude towards the next transaction, i.e. the next reading or learning situation.

According to Nielsen (1993: 24-25), who has studied the usability of user interfaces, usefulness means that a system can be used to achieve some desired goal. Usefulness can be divided into two categories: utility and usability. Utility means that the functions of the system can do what is needed, and usability addresses the question of how well the users can use the functions. Nielsen (1993: 115, 155) has introduced the concept of usability heuristics. Heuristic evaluation is a systematic inspection of a user interface to find its usability problems. In a heuristic evaluation evaluators examine the interface and "judge its compliance with recognised usability principles". The heuristic approach includes general rules that describe common usability properties. This kind of heuristic evaluation can also be applied to usability evaluation in documentation.

Purho (2000) has combined heuristic statements that can be used as a checklist on matters that need to be considered when designing usable

documentation. He, however, points out that heuristic statements are only generalisations and as such they do not replace careful planning and user-centred design processes. These heuristic statements can be summarised as follows. Firstly, there should be a match between documentation and the real world. The documentation should speak the user's language, in other words, the concepts used should be familiar to the user. Information should appear in a natural and logical order. Secondly, there should be a match between the documentation and the product. This means that the same terminology should be used consistently in the manual, user interface, online help etc. Thirdly, the purpose and the intended use of each document should be clear, so that the media of documentation (print, CD-ROM etc.) can be determined according to the purpose. For example, an installer working on a rooftop has more use of laminated reference cards than on multimedia CD-ROMs. The documentation should also support users with different levels of knowledge. Unnecessary information for a specific user should be hidden or it should be easy to overlook, and for expert users there should be quick reference information available. Next, information design should be effective, so that the reader could easily find and understand information. There should be support for various methods for searching information; for example, some people search through the table of contents, some use the index and some browse the document. The documentation should also contain a troubleshooting section, which helps the user in problematic situations. If the documentation set is large, there should be instructions for the user how the documentation is used.

These statements should help to design usable documentation. However, there are also opposing views on the usability of heuristic rule sets. De Jong and van der Poort (1994: 232-233) report some drawbacks. Although a lot of research has been done in the field of technical writing, according to them there are no valid and commonly accepted heuristics available yet, and in fact, the heuristics that have been developed are sometimes conflicting. Also, heuristics do not take into account the unique aspects of documents in

their functional context. With this they mean that what is good for manuals in general, is not necessarily good for a specific manual or a specific group of users. There are no heuristics available that cover for all decisions in the documentation process. Moreover, even if there is a valid heuristic available, a technical writer can apply it incorrectly. With these arguments de Jong and van der Poort argue for usability testing. According to them the heuristics and rule sets do not decrease the importance of usability testing.

My view on heuristics and methods for good writing is that they are of great help to the writer, although they cannot be generalised. Not all rules can be applied for all manuals or all groups of users, but heuristics and rule sets can be used as material for general reference. Design decisions are often context-specific, and not all decisions can be made according to general recommendations. However, the heuristics and rule sets can also be rewritten and narrowed down so that they can be used for a specific group of users. Many companies have written style guides and language guides, so often the heuristics used are company-specific. Heuristics help writers to improve the quality of documents, and if the usability of documents is also tested (and the documents are revised according to the results), the result should be usable documents with a very high quality.

Wright (1994: 11-12) defines usability as a factor in text quality. When contrasting quality with usability, thinking about the way written materials are used by the reader may be important. It shows the difference between documents if they are looked at from different viewpoints. If they are viewed as written products, text-based assessments of quality seem appropriate. If they are viewed as working environments, then the process of actual document use is important, and task-based assessment of quality is more appropriate. In other words, the usability of texts can be assessed in two ways. First, the quality of the text itself is assessed, when the text is separate from the context of use. Second, the assessment is made on the basis of how usable the text is when it is used for its planned purpose. In my

study the process of document use is of a special interest, so the task-based assessment of quality is more appropriate in my study.

My point of view is that when text usability is assessed, there are several factors that need to be taken into account. In task-based assessment of document usability the most important factors that need to be considered are the context of reading (text-reader interaction), the effect of reader's prior knowledge and the features of text. All these factors must be included in the assessment of document usability, because in my approach the documents are regarded as working environments, and their usability is assessed on the basis of how usable they are for their originally planned purpose. Therefore, for example an analysis of textual features would not be enough, because that would ignore the context of text use. The context of reading, the effect of reader's prior knowledge and features of text are dealt with in more detail in the following sections.

2.2 Text-reader interaction and the context of reading

According to Widdowson (1990: 99-114) meaning is negotiated in a text-reader interaction. The reader constructs the meaning on the basis of the text. As Widdowson (1990:99) points out, "we attach meanings to linguistic expressions...by invoking some pre-existing knowledge or other, or some co-existing feature of the situation of utterance". In other words, the writers have a picture of the reader in their minds, and they try to formulate the message so that it matches the reader's schemata, so that the readers get the intended meaning. This is not always successful. In my opinion the writer's intention has not succeeded if the reader needs to make guesses on the meaning of the text. Especially in an instructional, technical text there should be no vagueness, the text should be include just one meaning, so that there is no chance of misunderstanding. If there were a possibility of misunderstanding, this could result in expensive and even dangerous errors. Here the type of text and the context of reading are important. For example,

in fictional texts the situation is totally different, because in fictional texts there is always the possibility to interpret the meaning of a text in different ways – the meaning is a matter of opinion. Opposite to this, the meaning of an instructional text must be explicit.

Sticht emphasises that the context of meaning and the purpose for which reading occurs must be taken into account. According to him (1985: 317, originally in Sticht & al. 1977) there are two classes of reading tasks. First, there is the reading-to-do reading task – information is looked up, used, and it may then be forgotten. This happens for example at work, and it is performance-oriented. Second, there is the reading-to-learn reading task – this means previewing, reviewing, outlining and underlining in order to learn and remember. This kind of reading is done for example at high school and it is topic-oriented.

It can, however, be argued that the division between purposes for reading is not always this strict. Reading-to-learn can later on belong to reading-to-do category, if the reader first reads to learn and memorise, and then starts using the information, and reads the text as a reference text in order to be able to apply the information. Still, reading-to-do reading always has more immediate effect than reading-to-learn type of reading. Usually when a user's manual is needed, the task in question is a reading-to-do reading task. There is no need to memorise information when the user looks for information in order to solve a problem with the help of a user's manual.

2.3 Target group

When designing a text for a certain purpose, the target group, i.e. the readers, must be kept in mind. Kern (1985: 341-342) uses the concept "model of the user". When designing a job manual, the writer predicts what the user is like and how the user will use the manual. Based on this prediction the writer decides on the content, organisation and format of the

manual. The manual is usable if the writer has succeeded in estimating the actual usage of the manual. It is, however, difficult to create a model of usage in order to design a manual that is totally suitable for its purpose. It is rarely possible to define the group of users and draw a conclusion that no-one else outside that group is going to use it. Often the manual has multiple functions, and serves, for example, as a general reference text, a training text and a procedural guide for performing specific tasks.

Several studies have demonstrated that the reader's background knowledge has an effect on comprehension and learning. These studies have also shown that readers with more background knowledge tend to use more effective reading strategies and are more interested in the reading material. McNamara and Kintsch (1996: 247) have studied effects of prior knowledge on learning from high- and low-coherence texts. A coherent text is easier to read and understand, because the information is presented in a logical order, and the ideas are usually connected with logical connectors, or they are presented in a discourse structure that is familiar for the reader. A low-coherence text is naturally not so easy to read and understand as a high-coherence text. See chapter 2.4.4 for a more detailed definition of coherence. McNamara and Kintsch (1996: 247-8) examined the characteristics of the text and of the individual reader. They also investigated how these factors affect text comprehension. The results of their study indicate that high-knowledge readers can find matters relevant for their purposes quickly in the text. Because they have good background knowledge on the matter, they may understand the text even if it is not coherent. For low-knowledge readers disorganised and poorly written texts cause more trouble.

One purpose of this study is to find an optimal level of difficulty for the readers. The readers in the usability test do not have prior knowledge on the subject matter or on technical matters in general, so my assumption is that they can best cope with a text where everything is expressed clearly and in

detail. This means simple, explicit and sufficient instructions and descriptions. An expert reader may become frustrated if the instructions include matters that are self-evident for the reader, but for a low-knowledge reader instructions have to be as clear as possible. In this study the usability of two manual versions is tested when a group of novice readers conclude tasks with a computer programme. They are not familiar with the programme or the technical matters involved, and in order to solve the problems and be able to conclude the task, they need to look for information from the user's manual. The test results indicate how usable the manual versions are, and which one is more usable for a novice reader.

2.4 Features of text

2.4.1 Presentation of information and text accessibility

In addition to the context of reading and reader's background, features of text have an important effect on the usability of texts. Text accessibility plays a major role when usability is considered. If the text is not accessible from the reader's point of view, then it is also not usable. Klare's (1963) definition of accessibility is the "ease of understanding or comprehension due to the style of writing". Lassen (1997: 36) draws a line between accessibility and readability. According to him accessibility refers to a quality of a text, whereas readability tends to be understood as a quality of the text as well as of the reader. Nyyssönen's (1997: 111) definition of accessibility is that a text is accessible if it is easy to read and understand. However, accessibility from his point of view depends not only on the text but on the interaction between the text and reader. An accessible text presents information and guides the reader's interpretation by signalling the discourse structure. According to Nyyssönen (1997: 112) writers make texts more accessible to the reader by using common discourse structures. These are structures such as given-new (theme-rheme), topic-comment, general-specific. In the end, however, accessibility depends on how well the reader's

"personal schemata", knowledge of the world and of the subject matter interact with the text. In other words, if the reader's mental model of the world is contradictory with the model of the text, the reader may not accept the information, and the text is unusable and inaccessible from the reader's point of view.

Nyyssönen (1997: 116-117) emphasises that instructions are the type of text that must often be read quickly and under pressure. They are often consulted, and not read through like a book. This must be taken into account in visual design. Understanding instructions involves more than just understanding the text. Crucially, it involves the construction of a mental model of the procedure and the many steps involved in it, "such as linking the referents in the environment to *terms* in the instructions" (1997: 116, emphasis in the original). In order to make the links, "the reader must interpret linguistic expressions in the text not only as carriers of semantic meaning but also as pragmatic pointers which instruct the reader to find their referents" (1997:116).

Isomursu (1997: 85-98) reports the results of a study on the role of vocabulary in text accessibility in an actual reading situation. The material that was used was a piece of text from a computer manual. The focus was on computer terminology – the way it is used in a text and the way readers access it (for a description of how the data was collected, see Pilto and Rapakko 1995). The material that was used was not intended for a certain target group, but it could be used by a novice or an expert. It included computer terminology, but mostly the words used were neutral core words (non-technical). How difficult task the reading was depended on the readers, their background knowledge and language skills. It was clearly indicated in the study that unknown vocabulary has an effect on the quality of reading. It slowed down the process of completing the task, i.e. accessing the text. Isomursu argues that it cannot be assumed that only non-core (technical) words have a negative effect on accessibility. Core words can cause serious

trouble for the reader, because they can have variable meanings, whereas specialised terms normally have a restricted meaning. Also, core words can be used in different contexts, whereas specialised terms usually have a restricted use.

The style of writing certainly has an effect on comprehension and accessibility. In an accessible text the information is presented in a way that makes it easy for the reader to find the right type of information. Then again, what the right type of information is depends on the context of text use: the purpose for which the reader looks for information and the circumstances where the reading occurs. In my opinion accessibility is a factor affecting usability. Accessibility depends on the successfulness of the text-reader interaction, so both the features of text and the reader have an effect on accessibility, and thus on the usability of the text. Usability as a concept, in my opinion, includes a larger framework. In addition to the features of text and the reader, also the context of reading and the process of text use must be taken into account when usability is considered.

2.4.2 Controlled Language

Using natural language can result in ambiguities, especially when translation is concerned. Many organisations have published writer's guides or specifications of controlled language in order to improve the translatability and readability of technical documentation. The purpose of using controlled language is to ensure coherent communication. It includes a simplified syntax and a limited use of vocabulary. Expressions in controlled language are simplified, but more precise, and therefore usable. Recently, controlled language has been developed for the purposes of computational linguistics, for machine translation (Introduction to Controlled Language, http://www.linga.fr/LingEn/lanconen.htm, 31.10.2000.)

Controlled language was first developed to make a text more usable and translatable. In my study controlled language was used in one manual version (version A, see Appendix 5) that was tested, and one purpose in my test was to find out whether controlled language turns out to be useful in the test, i.e. whether the test persons find that the language is clear, consistent and it improves understanding. In the following paragraphs the concept of controlled language is defined in more detail.

Controlled language includes a set of rules, which guides writers. AECMA (European Association of Aerospace Industries) has published a writer's guide of Simplified English (SE) for the purposes of aeronautical industry. AECMA's Simplified English guide (1998) is an example of a controlled language guide written for the improvement of technical documentation. The rules that are described in the following paragraphs are taken from the AECMA Simplified English CD-ROM (1998). The guide includes rules, for example, about words, noun phrases, verbs, sentences, procedures and warnings. For example, according to the rules concerning the use of words, professional jargon, idioms and indirect expressions should be avoided. Only approved words, that can be found in the AECMA dictionary, should be used. Terminology should be consistent, only one name should be used for one thing. Instructions should be direct and specific. In noun phrases no more than three nouns should be written one after another, in other words, ambiguous noun clusters should be avoided.

According to the rules only certain verb forms (tenses) are allowed: infinitive, imperative, simple present, simple past and a future form. If possible, active voice must be used instead of passive. Only one topic should be expressed in one sentence (as in the rules of Information Mapping®, see section 2.4.3). Connective words and phrases should be used to improve readability and show the relations between ideas. Embeddings and non-parallel constructions should be avoided. Tables should be used to illustrate complex issues.

According to the Simplified English rules sentences should be short, only one instruction is allowed per sentence. The imperative verb form should be used in instructions. The order in which information is presented should be logical, for example, chronological or cause-consequence order. Only one topic should be handled in one paragraph. The paragraph should begin with a topic sentence, and the paragraph should be relatively short (max. 6 sentences). New and complex information should be presented slowly. Warnings should be clearly indicated (AECMA Simplified English CD-ROM, 1998.)

2.4.3 Information Mapping®

The Information Mapping® (IMAP) method is a method of presenting information, which can help improve document usability. It was first introduced in 1965 by Robert E. Horn. Horn wanted to improve the access, readability and usefulness of written communication. The method is now based on more than 30 years of research into how people read and understand information (Horn 1985:180.)

The IMAP method is based on seven principles of communication. The principles together with the guidelines for applying them ensure that the writing style is both reader-based and task-oriented, and it meets the needs of both readers and writers. The principles are: the chunking principle, the relevance principle, the labelling principle, the consistency principle, the integrated graphics principle, the accessible detail principle, and the hierarchy of chunking and labelling principle (Horn 1997: 3-A-1 – 3-A-13.) These principles are described in the following paragraphs.

According to the chunking principle information should be grouped into small, manageable units. Research has shown that people can best process and remember no more than seven plus or minus two pieces of information

at one time (see Miller 1967). If information is complex, the chunk should be smaller. Readers comprehend material in chunks more quickly and more completely. Our mind tends to group items into categories, which makes the job easier for the mind (Horn 1997: 3-A-2.) Long sentences may cause the reader to slow down and make mistakes. Therefore, the sentence length should be an average of 20 words, but never more than 30 words. There can be flexibility in this rule, but it is good to keep in mind that readers comprehend and best hold in short-term memory 7±2 units of information.

According to the relevance principle information in one chunk should relate to the main point "based on that information's *purpose* or *function* for the reader" (emphasis in original). There should be only a limited kind of information on a single topic in one chunk and irrelevant information should be left out. This assists comprehension and saves time, e.g. when a reader is scanning in order to find information (Horn 1997: 3-A-3.)

Following the labelling principle means that a label for each unit of information is provided in order to give a brief preview for the reader (in the left margin, or as subheadings). The labels, or subheadings, provide "advance organisation" and help the reader to scan, to find relevant information and to skip unnecessary information. This makes the text more usable, because people tend to understand and access information quicker if they have a preview of what is coming up (Horn 1997: 3-A-5.)

The consistency principle emphasises that writers should use similar words, labels, formats, organisations and sequencies for similar subject matters. Consistency saves the readers' time: they can find the information they are looking for quickly, and they can locate similar information and focus on content, not on form. In prose writing writers use variety to entertain readers. However, especially in technical writing variety is often confusing and time-consuming (Horn 1997: 3-A-7.)

Graphics are also an important part of the text. Research has shown that half of the adult population learns better from pictures/graphics than from words. Therefore, in the Information Mapping® method there is an integrated graphics principle, which recommends the use of diagrammes, pictures, tables, lists etc. Presenting the information visually helps the reader to locate information and to learn it more efficiently (Horn 1997: 3-A-8.)

According to the accessible detail principle one should write at a level of detail that makes the information readily accessible for the reader. The writer should "put *what* the reader needs *where* the reader needs it" (emphasis in original). There should be a more detailed illustration for inexperienced users of the product. Comprehension improves when abstract principles, concepts or procedures are supported by concrete examples. The writer should include overviews, reviews, descriptions, diagrammes and examples for all abstract presentations. They should be placed near the text they illustrate, not for example in the end of the whole document (Horn 1997: 3-A-10.)

The last principle concerns the hierarchy of chunking and labelling information. Small units of information should be organised into a hierarchy. People comprehend and remember large amounts of information better if the information is organised into groups of 5-9 items (e.g. sentences in a paragraph). If the items are labelled, learning and locating the information is faster. (Horn 1997: 3-A-12)

Following the guidelines of Information Mapping® is advantageous for both the readers and writers. Reading is faster - it has been reported that readers save 20% to 40% of their reading time, because they can find the information they need quickly, and they understand what they read more easily. Documentation written according to IMAP rules has resulted in fewer errors. Also, relearning of forgotten material has been faster, because specific information is easy to find. People trained in the method have

reported that their writing efficiency has improved. Writing is easier with an analytical, structured approach, and the result is a better organised text. Documents written according to the guidelines are more appropriate, that is, they contain the information that the reader needs, where the reader needs it (Horn 1997: 2-9.)

The point of view in my study is that the rules of controlled language and Information Mapping® improve understanding and make the texts more usable. However, the context of reading and text-reader interaction are also important factors affecting document usability. This aspect has not been taken into consideration enough in these methods and rule sets, and my perspective is that one must pay attention to all these points (the context of reading, reader's prior knowledge and textual features and rules of writing) when improving the quality and usability of documentation.

2.4.4 Coherence and comprehensibility

Coherence is also a feature of text that has an effect on the usability of documentation. According to Halliday and Hasan (1976) coherence means how smoothly ideas in a text are woven together. In a coherent discourse, the relationship between ideas must be clear so that there is a logical connection from one idea to the next. This enables the reader to perceive the message. Nyyssönen (1997: 115) states that coherence means how well the reader can understand the relations between ideas in a text. Coherence in a text is supported by cohesiveness, that is, the text provides clues for the readers so that it is easier for them to build an understanding.

Factors that affect coherence are, for example, discourse structure and logical connectors. The text is coherent if it follows a discourse structure that is familiar for the reader. Nyyssönen (1997: 112) states that given-new, topic-comment and general-specific are discourse structures that are commonly used to organise the content of texts. These structures are signalled by lexical and grammatical rules. Also, with the help of logical connectors (firstly, secondly, next, therefore etc.) the reader is told in which order matters are expressed and what the relationship between matters is, in other words, the reader is given metatext that is useful for understanding.

From the above mentioned one can conclude, that coherence means the logical order of presenting information and the provision of signals for the reader, so that the reader can better understand the text. Coherence affects the comprehensibility and thus the usability of a text, because even if a text is readable, it is not necessarily comprehensible, if it lacks coherence. By readable I mean a text that is grammatically correct and has sensible sentence structures. Such a text is readable, but if the ideas are not clearly and coherently expressed, the result from the reader's point of view is not comprehension but confusion.

3 RESEARCH DESIGN

In the present study there are three research questions that are focused on. First, the study provides information on the effect of different textual features on document usability, i.e. on readability, understandability and accessibility of the document. As has been pointed out earlier, the textual features that have an effect on document usability include the use of vocabulary and sentence structures, discourse structures, coherence, visual design and layout. My hypothesis is that if a technical text is written using the guidelines of controlled language and Information Mapping®, it is easier to read and understand, and therefore more usable to the end-user than the original text. I observed the use of manuals to find out the effects of different text versions on the way the user finds and understands information. The purpose was to find out whether it, for example, takes longer to find information from the manual version that is not written according to the rules of good writing, or whether visual design has a clear positive effect for understanding.

Second, this study explores the optimal way of presenting information for a reader who is a novice in the area that the manual handles. My hypothesis is that a text that is written clearly and which includes detailed and visual descriptions (version A) is more usable for a novice reader than a text which includes only core information (version B). The results of my study show whether either manual version A or B is written in an optimal way for novice readers. If the reader finds information quickly and without problems, then the way of presenting information and the level of information are optimal for the target group. A reader with a lot of knowledge on the subject matter may become frustrated if there is "self-evident" information in the manual, but for a novice reader this information could be necessary for understanding. Therefore, the target group must be

taken into account when writing - there are different ways of presenting information for different target groups.

Third, this study aims to find potential design problems in the manual. Possible design problems can be found in the usability tests, when a test person is using the manual for the purpose for which it has been written. If the reader cannot find necessary information from the manual, or for example, thinks that the layout of the manual is not clear, then the design of the manual must be improved.

There are already several guidelines for improving readability and accessibility (and thus also usability), and the purpose of this study is not to provide new guidelines, but to test the usability of these guidelines in a practical context. The results of my study help improve the quality of the documents that are tested, but they can also provide information on human information processing and the effect that different textual features have on it. This study can also give valuable information for the use of document designers in general. In the following section the methods used for collecting information are described.

3.1 Methods

The present study is partly an experimental study, where observation was used as the main method to collect information. I tested the usability of two manual versions for a computer programme. The programme was X-Fetch Wrapper Rule Generator that Republica has developed, and the manual versions were user's manuals for it (see Appendix 5 for manual version A and Appendix 6 for manual version B). The test was organised in a laboratory setting, and the test persons had to do a task on a computer (see 3.2 for a description of how the test was organised, and Appendix 3 for the test task). The test persons were unfamiliar with the computer programme, and novices in technical matters, so they were expected to need instructions

from the manual in order to be able to complete the task. Novices were chosen as test persons, because a group of expert readers was not available for testing. During the test I observed the test situation and made notes on the reading process.

A combination of methods was used in order to collect a variety of information on how well the user finds and understands necessary information from the manual. In addition to observation, the methods included think aloud -method and user's feedback. The research questions in this study concern the usability of instructional texts, and since there are many factors affecting usability, just one method would not provide enough information. It is necessary to collect also user's feedback because observation does not give enough information from the actual user's point of view on usability. In addition to think aloud -method, a post-test questionnaire was used to collect the readers' opinions on the task and on the manual versions (see Appendix 4 for the post-test questionnaire). In the analysis I also describe the material used in the test, and explain the differences between the manual versions.

The features of both the text and the reader have an effect on the reading process, therefore, they must be taken into account in the study. Observation provides information on the reading process and on possible problems in it, but user's feedback is also necessary to gain information on the usability from the reader's point of view.

Although the test was organised in a laboratory setting, the reading process and the purpose for reading were quite natural: the reader had to do a task on a computer with the help of a user's manual. Observing a 'real' situation, for example, in a work environment would have been impossible: it would take a long time to wait until such a situation would occur, where the use of instructional text were needed, and the researcher could hardly observe the situation as a neutral observer at someone's workplace. Also, a natural

situation would set a limit because it would not be possible to observe the use of different versions. Using a combination of methods is wise, because the researcher has to be prepared for surprises. In an experimental study the results may not be as expected and the methods may provide different type of information than expected. In the following sections the methods used in this study are described in more detail.

3.1.1 Description of texts

First the features of different versions of a user's manual are described. I will analyse the texts so that the reader gets a picture of the differences between the manual versions (see 4.1). The differences in them are described, for example, how Information Mapping® method and controlled language have been or have not been used when writing the manual. In version A a lot of attention has been paid to the way of presenting information: there are a lot of detailed descriptions and information is visually presented, and language structures and vocabulary have been carefully chosen. In version B this is not the case: the manual is not so visual, the use of language, sentence structures or sentence length is not limited and there are less detailed descriptions, the focus is on presenting core information. The comparison and description of different manual versions serves as a background for the task-based usability test and the analysis of the use of texts.

3.1.2 Observation

Schumacher and Waller (1985: 386-397) discuss process-oriented measures of text design. In a process-oriented study the user's processing of a text is observed, and the user gives feedback to help find out which parts are difficult. Schumacher and Waller (1985: 386-387) describe, for example, a

user edit-procedure, that Atlas (1981) first introduced. The basic idea of this method was used in the present study. In user edit -procedure a person is given a manual and is asked to work with a machine with the help of the manual. Pauses and errors are carefully observed, because they can provide information on the document design problems or weaknesses in documents. The user edit -procedure is normally carried out with a small number of test persons. It is oriented toward a feedback or troubleshooting approach to a single document, so the main purpose is not to determine general principles of document design. However, it may be possible to derive some principles for improving documents in general.

According to Schumacher and Waller this kind of procedure is most effective for documents, where the user's action is necessary (user's manuals), because in such documents it is easy to locate comprehension problems. There are several advantages in this kind of testing. It is a simple way to assess how a person is using a document and to find potential design problems. It results in information about specific design problems, about the content and possibly about some stylistic questions. This is not an expensive way to test documents. However, it should give valid and reliable information. Possible disadvantages are, that it may be time-consuming (depending on the situation), and also that the test persons with different kinds of backgrounds may give different types of feedback, and therefore data may be difficult to summarise. This kind of method is, however, suitable for the purposes of my study, because when comprehension problems are located, the problems can be minimised by improving the problematic areas, and thus the usability of the documents can be improved.

In this study observation was used as a method to collect information on the reading process. I observed the situation where the user looked for information from a manual for a purpose and made notes on what happened. I acted mostly as a non-participant observer, meaning, I did not actively take part in the situation or try to manipulate it. However, if the test person asked

for assistance, I helped him/her to pass the problematic situation, so that the test could go on. If my assistance was needed, it meant that the manual had not provided enough information for the user. I told the test persons that they can feel free to talk to me (think aloud-method), even though I was not actively taking part in the conversation. When thinking aloud, the test persons felt more comfortable talking to a person, therefore I was partly a participant observer. I asked the test person to mark problematic parts and after the test I asked for a description of the problem in order to find out the reason for it. I also recorded the situation on video, so that later on I could go back to the data and analyse the participants' actions.

3.1.3 User's feedback

The above mentioned observation-method was combined with think aloud -method in my test. According to Schumacher and Waller (1985: 387-389) the think aloud -method is a protocol analysis procedure, which can be used in evaluating document design. It is a technique for investigating human problem solving activities. In this method the test persons are asked to think aloud, as they proceed with reading and try to understand the document. They can provide information about many things, but especially the problems they face when reading are of interest in the present study. In this case the method is used to provide feedback on the features of particular documents. The reader can give information on difficult or unclear parts, which could be improved for better understanding. Also, the reader can give positive feedback: if the reader agrees that the rule sets and methods for good writing result in good readability and understandability, it is recommendable to use them and to pay a lot of attention to the presentation of information to improve document quality.

There are some problems that may occur when using this set of methods. The test person may provide a great deal of information, but not on the area

that the study focuses on. Think aloud -method has also been criticised, because when people are asked to think aloud, their processing of the document may not be the same as in a 'normal' situation, when they would quietly read. However, the purpose in my test was not to make the test persons talk all the time. They needed time and silence to concentrate on reading and on completing the task, but they were asked to talk aloud whenever there was a problem, something odd/interesting, unfamiliar words etc. If the person did not feel comfortable with talking, the same information could also be provided in writing. The test persons were asked to mark sections which were not clear to them, and after the test there was a questionnaire for the participants to fill in.

3.2 Description of how the test was organised

I tested the usability of two manual versions in practise, in a text-reader interaction. The manual was a user's guide for a computer programme called X-Fetch Wrapper Rule Generator. Republica's Research and Development Unit has developed the programme, and the manual versions were written by a technical writer. The test persons were asked to fetch information from the Internet and to convert it to XML by using the computer programme with the help of the manual. The users had not seen the user interface before, and they were expected to consult the manual in order to be able to perform the task. In the following subsections there is a description of how the test persons were chosen and prepared for the test, a description of the test setting and of the initial report that was written shortly after the test.

3.2.1 Test persons

In this test there were two variables which could have an effect on test results: test persons' English skills and background skills on computing. The test persons were chosen so that they all had good language skills and at

least basic abilities to use computers. There were not any major differences in the language skills, because the test persons had either passed the entrance examination to the English department in Jyväskylä, or spent a long period in an English speaking country.

However, the computer skills of the people may vary. In the questionnaire after the test the persons answered questions about their computer skills, so that if there were major differences in the persons' backgrounds, it could be taken into account in the analysis. It turned out that there were no major differences in the computer skills of the test persons. In the scale of 1-4 (1 = novice, 4 = expert) most of them assessed themselves to be on level 2, and one chose level 3. The test persons reported that they use computer regularly for e-mail, Internet, word processing, and sometimes for spreadsheet computing (Excel) and for playing computer games.

3.2.2 Familiarising the test persons with the test subject

Before the test there was a brief 'training' for the test persons, where it was explained what the test would be like. I explained some terms that I expected to be difficult (as they turned out to be). For example, I explained what HTML (Hypertext Mark-up Language) is, what it is used for and what it looks like: I showed an example of an HTML-page, as the Internet user sees it, and the same page's source HTML, with tags. Also the overall context was explained – what the programme is used for and what relation it has to other programmes. See Appendix 2 for the material used when familiarising the test persons with the test subject.

After the test all test persons agreed that the training had been useful, and that without it the test would have been hard, because they would not have understood at all what they were supposed to do (because of unclear terms). According to the test persons the amount of information given beforehand

was suitable, and it was good that too much had not been told before the test. However, the time period between the 'training' and the actual test turned out to be a bit too long (1,5 weeks). It would have been better if there had been only a few days between the training and the test, because the test persons said that they had forgotten some of the things that were explained. Another option would have been to explain some things again right before the test to refresh their memories.

3.2.3 Test setting

The manual usability tests were organised in a meeting room in Republica's premises. Six people took part in the test, and 1,5 hours had been reserved for each person. There were two manual versions to be tested, so there were three people per manual version. A video recorder was used for recording the event, and a cassette recorder was used as a backup. First I explained each person what was going to happen and what they were expected to do: to talk aloud when thinking, to mark unclear parts in the manual so that they could be discussed after the test, when filling in the post-test questionnaire. I also explained my role, that I was there as a non-participant observer, but that I would help, if there was a situation where they would have no idea what to do (and the manual would not help). I said that they could think that they are talking to me when thinking aloud, although I would not answer, and that if they directed a question to me, then I would answer and help them continue the test.

In the beginning the test persons signed a mutual non-disclosure agreement, in which they promised not to tell anyone what they saw or heard during the test. A test person sat in front of a computer, where the program was already open and ready for use. They were given a task and a manual, and directions to do the task with the help of the manual. After the task the test persons filled in a questionnaire, where they described what sort of problems they

had had, and the problems and the possible causes for them were discussed. The persons also proposed what in their opinion could be done to improve the instructions in the manual, if they thought the instructions had not been optimally usable.

3.2.4 Initial report

After the test I wrote a report about the most relevant matters that came up in the test, for example, the most problematic terms and unclear instructions were described in the report, and also solutions for the problematic parts were suggested. Thus, the manual could immediately be improved according to the report. In a company publishing the manuals it is important to get an updated version of a manual for the customers as quickly as possible, and since it was obvious that a thorough analysis takes a longer time, I wrote an initial report shortly after the test, for the immediate improvement of the manual. The results of the initial report are naturally included in the analysis of this thesis.

4 ANALYSIS

The aim of the analysis is to describe how different textual features affect the usability of texts, what kind of problems occurred when the test persons performed the tasks with the help of a manual, and how the usability of documents can be improved on the basis of this usability test. Usability of the different manual versions was observed in this test in terms of processing of information, errors, problems, goal achievement, time and user satisfaction. In section 4.1 the manual versions A and B are described in more detail. In section 4.2 the reading process is analysed, and in section 4.3 the errors and problems that occurred during the test are described. The matters dealt with in sections 4.2 and 4.3 concern both manual versions. There were content related problems (i.e. problems related to the reading process, terminology and unclear instructions) that occurred with both manuals, and therefore in sections 4.2 and 4.3 there is in most cases no specification which manual version is in question. If there is no specification, then the matter concerns both manual versions. In contrast to this, sections 4.4-4.8 deal with the differences between the manual versions. In these sections the effects of different textual features are analysed. Section 4.4 deals with the layout of the manual versions, section 4.5 with visual design and the use of graphics, and in section 4.6 the effect of language in the manual versions is considered. Section 4.7 includes an analysis of the goal achievement and the time used for performing the tasks. In section 4.8 test persons' feedback and opinions about the test in general are described.

The examples in the analysis include the test persons' comments during the test, when they were talking aloud, and also comments and answers to the questions from the post-test questionnaire. The origin (video tape or questionnaire) of an example is pointed out when the example is given. I wrote down some of the test persons' comments during the test, but to make

sure that everything is correct I rewrote my notes on the basis of the video tape, therefore, the comments originate only from the tape and from the questionnaire. Most comments were originally in Finnish and I have translated them into English. See Appendix 1 for the original comments and their translations into English.

4.1 Description of the test material

4.1.1 Manual version A

At the time when the tests were held, version A (see Appendix 5) was the latest version of the manual, which had been published as a customer document. However, by the time this study is finished many updated versions of the manual have already been published, because the application is developed continuously, and thus also the manual must often be updated and improved.

In version A the IMAP guidelines and controlled language rules were used to a great extent. Labels (side headings) are used to describe information content in the left margin. Examples of side headings are *Summary, Definition, Function, Road map, Tip!* and *Instruction.* Lines are used to divide topics, so each time a topic changes, there is a line showing the change. There are road maps telling the reader what to do in each situation and where information can be found, i.e. the reader is 'led by the hand'. For example, in chapter 1 Introduction there is a *Road map* explaining the reader what each chapter includes, so that from Introduction the readers can go straight to the information they need. Also, at the end of some chapters there is information for the reader what to do next, labelled with side heading *Next.* There is also a *Chapter summary* in the beginning of each chapter, telling the reader what the chapter includes.

There are a lot of figures in the manual version A to make it more illustrative. Pictures of the buttons of the user interface are added in the text whenever the button is mentioned. Captions are detailed and descriptive. Tables are used where possible, for example on page 4 of the manual (see Appendix 5, page 99) there are two tables. Notes and Cautions are clearly expressed and differ from the body of text. There are many means of emphasis in Notes and Cautions: indentation, bold, larger font and a figure in the left margin marking the location of Note or Caution.

Active voice is mostly used instead of passive in instructions, and the imperative form is often used ("select the area" instead of "the area is selected"). Sentences and paragraphs in the manual are relatively short, and there is only one topic per sentence/paragraph in order to avoid information overload. Also, connection words are used to improve readability, for example, *first, then, however* and *therefore*. Such connection words make the text cohesive and thus improve coherence.

The following pages (40-42) include three sample pages of manual version A (printed by permission from Republica Corp.). The same sample from manual version B is presented on page 44 in this thesis.

6 EDITOR FUNCTIONS

Chapter summary This chapter gives you the instructions how to perform data extraction from source material using Rule Generator.

Road map The order of presentation is chronological, starting from data selection (6.1 - 6.3) and proceeding to data analysis (6.4) and extraction (6.5). Previewing the XML output is described in chapter 6.6 and modifying the extraction rules in chapter 6.7.

6.1 Selecting Data Area (Add Data Area)

Chapter summary This chapter describes how to select Data Area.

Definition A Data Area contains the whole data selection area: all Entry Areas and all Fields.

Instructions

1. In Browse mode, select the HTML page from which you want to extract data.
2. Activate Edit mode (if in Browse mode) by pressing **Edit mode** toolbar button.
3. First, using left mouse button, select the area, which contains all the data to be extracted.
4. Then, click right mouse button and choose **Add Data Area**.

The Data Area has now been automatically analyzed.

 tags appear around the selected area.

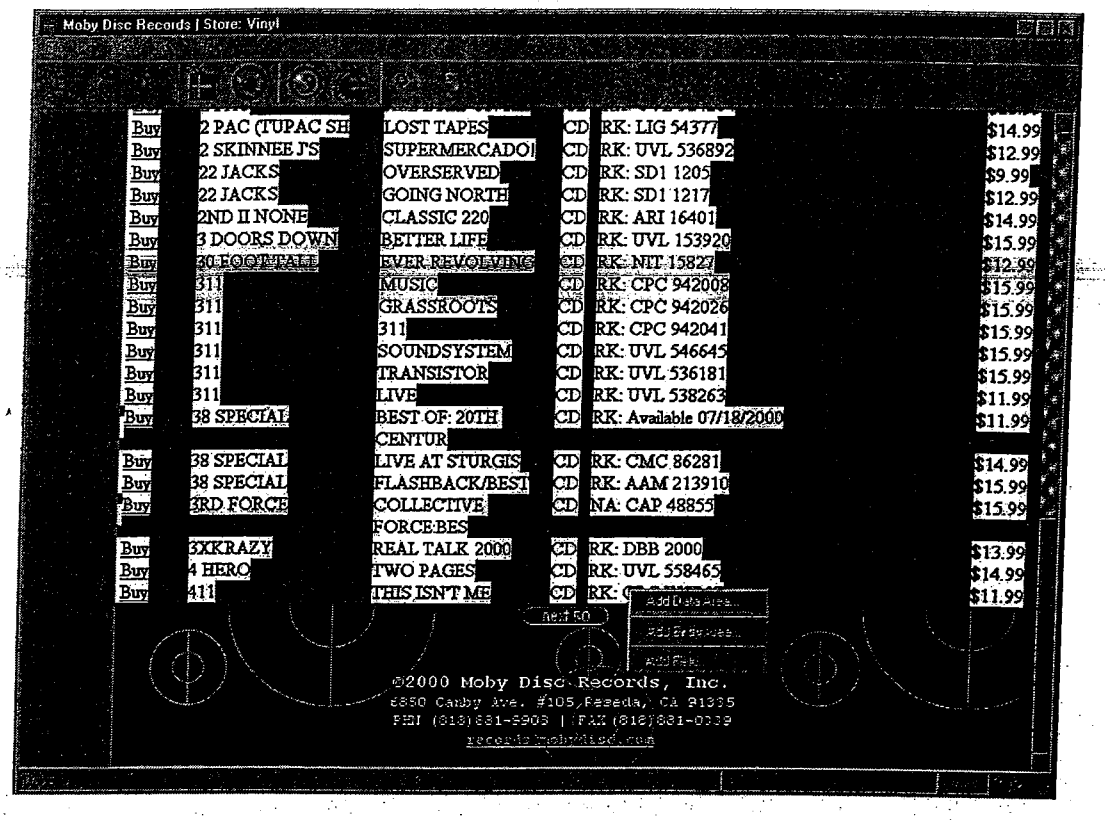


Figure 4. Adding Data Area. Select the Data Area containing all information of listed albums (shows in white). Right-click and choose Add Data Area to confirm the selection.

NOTE: No undo function is available yet. To change the Data Area definition, you have to reload the page from cache and re-create Data Area.

NOTE: Though you can select multiple Data Areas, X-Fetch Wrapper Engine does not process additional areas yet. Proceed with one data area.

Next	Proceed with selecting Entry Area.

6.2 Selecting Entry Area (Add Entry Area)

Chapter summary	This chapter describes how to select Entry Area.
Definition	An Entry Area Contains all Fields once. It is also known as All Tags Area.
Instructions	1. First, using left mouse button, select the area which contains one complete entry. In an Entry Area, all the prospective fields occur only once.

NOTE: You should not choose the first row in Data Area as Entry Area. XML output may not come out right. Rather choose the second or third row.

NOTE: **Obviously, the Entry Area must reside within the Data Area.**

2. Then, right-click and choose **Add Entry Area.**

The Entry Area has now been defined and automatically analyzed.

tags appear around the selection.

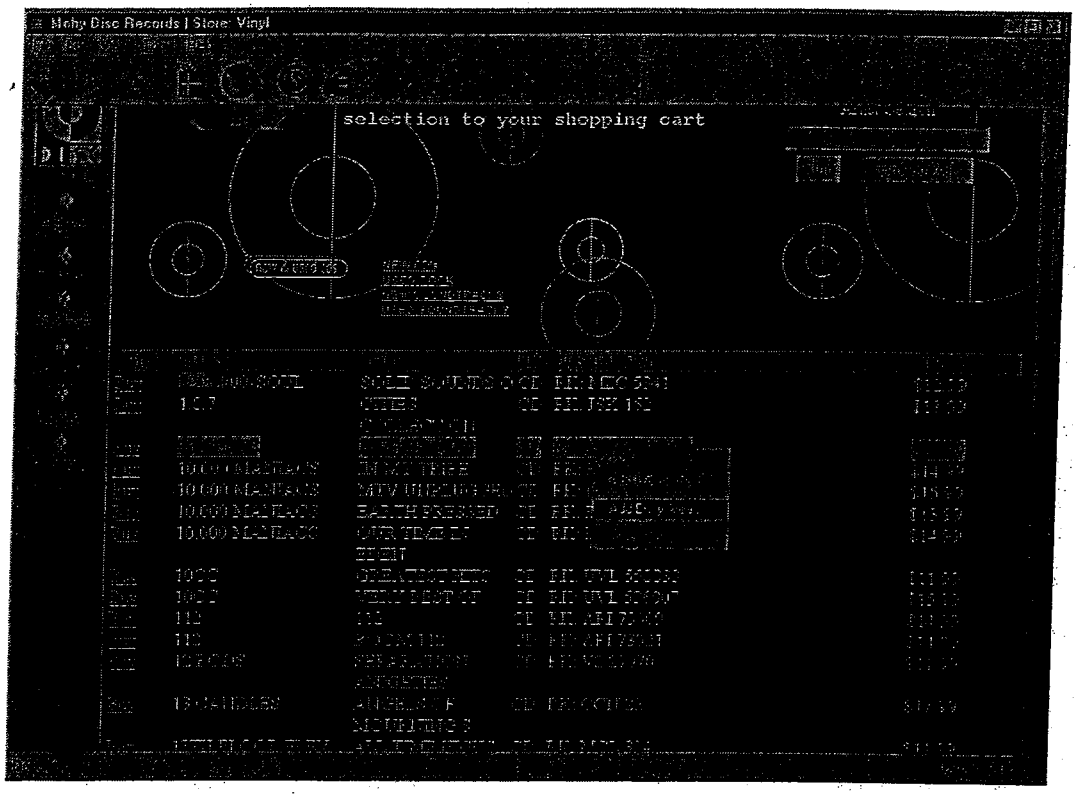


Figure 5. Adding Entry Area. Select an entire row containing the artist, album title, FMT (format), description as well as price using mouse (the area shows in white). Then right-click and choose Add Entry Area.

3. Check that the selection is how you intended. If not, redo the entire selection by pressing  **(Reload)** button.

NOTE: **No undo function is available yet. To change the Entry Area definition, you have to reload the page from cache (using Reload button) and re-create Entry Area.**

NOTE: **Though you can select multiple Entry Areas, the Engine only processes the first one. Therefore, proceed with one Entry Area only.**

Next Proceed with selecting individual Fields.

4.1.2 Manual version B

Version B (see Appendix 6) is an earlier version of the manual, in other words, version A is more advanced than version B. The contents in the manual versions A and B are the same, but the layout, language and the way of presenting information differ.

In version B the IMAP rules have not been consciously used. For example, there are no labels marking information in the left margin and there are no lines separating topics. There are no *Road maps* explaining the reader explicitly what to do next, and where information on each topic can be found. Captions are short and not as descriptive as in version A. Notes or Cautions are not clearly expressed. They do not "jump out of the text", so that the reader would not be able to miss them, but they are embedded in the text. Also, the font used in the Notes and Cautions is the same as in the body of text.

The use of language in manual version B is not limited by controlled language rules. In other words, passive is mostly used in instructions and imperative form is used only rarely ("the area is selected" instead of "select the area". The length of sentences and paragraphs is not limited, and there are sometimes several topics per one sentence or paragraph. Also, not many connections words have been used to connect topics and sentences.

The following page includes a sample page of manual version B (printed by permission from Republica Corp.). The same sample from manual version A was presented on pages 40-42 in this thesis.

6 EDITOR FUNCTIONS

6.1 Selecting Data Area (Add Data Area)

A Data Area contains the whole data selection area: all Entry Areas and all Fields.

In Browse mode, a desired HTML page from which you want to extract data is selected. Edit mode is activated (if in Browse mode) by pressing **Edit mode** toolbar button. Using left mouse button, the area which contains all the data to be extracted is selected. Right mouse button is clicked and **Add Data Area** is chosen. The Data Area has now been automatically analyzed.

 tags appear around the selected area.

Note: No undo function is available yet. To change the Data Area definition, the page from cache has to be reloaded and Data Area must be recreated.

Note: Though multiple Data Areas can be selected, X-Fetch Wrapper Engine does not process additional areas yet. Proceed with one data area.

6.2 Selecting Entry Area (Add Entry Area)

An Entry Area Contains all Fields once. It is also known as All Tags Area. The area which contains one complete entry is selected using left mouse button. In an Entry Area, all the prospective fields occur only once.

Note: The first row in Data Area should not be chosen as Entry Area. XML output may not come out right. It is recommended that the second or third row is chosen.

Note: Obviously, the Entry Area must reside within the Data Area.

Then, **Add Entry Area** is chosen by right-clicking. The Entry Area has now been defined and automatically analyzed.

 tags appear around the selection.

Check that the selection is how you intended. If not, the entire selection can be redone by pressing  (**Reload**) button.

Note: No undo function is available yet. To change the Entry Area definition, the page must be reloaded from cache (using Reload button) and Entry Area must be recreated.

Note: Though you can select multiple Entry Areas, the Engine only processes the first one. Therefore, proceed with one Entry Area only.

4.2 The reading process

Next there is a description of the problems that the test persons had when using the manual versions. There are also suggestions of how the manual could be improved in order to avoid these problems. During the test I observed how the test persons performed the task: when did they use the manual, and for what kind of information? I also made notes on how the test persons proceeded with the task, for instance whether they used the table of contents in order to find the information needed for a certain task, or whether they started reading from the beginning and go through the whole manual.

The test persons followed the tasks step-by-step. All test persons used the table of contents to start with (in the beginning of the task and when starting a new subtask). From the table of contents they looked for information needed for a certain task. If they could not find information they were looking for from the chapter referred to in the table of contents, they went back to the table of contents to see if there is something else that might be useful. Examples (1) and (2) (comments which have been taken from the video tape and translated into English) show that the test persons found the table of contents very useful:

> (1) "at least the table of contents is very clear, everything can be found easily"
> (2) "again it is easily found from the table of contents"

If the test persons could not immediately find what they were looking for, they started reading the manual, skimming and leafing through pages. Sometimes the result was that they found what they were looking for, but sometimes the result was frustration, as can be seen in examples (3), (4) and (5) (taken from the video tape):

> (3) "I am annoyed because I cannot find the instructions!"
> (4) "I cannot find it...this is hopeless"

(5) "it makes me mad when I can't find something and then I wouldn't be bothered to read everything"

Some test persons became impatient when they could not find what they were looking for. Examples (6) and (7), which have been transcribed from the video tape, show that they were eager to experiment and test the functions of different buttons:

(6) "I would rather do than read, I would like to just have a try."
(7) "what if I'll just press some button and try"

The comment (8), which was given after the test when filling in the questionnaire, also shows what happens if a person who is using the manual does not find the things looked for quickly, or does not understand them. This often results in frustration, and the manual is thrown aside as useless.

(8) "if this had been a real situation, then I would just try things out, and would have tried a lot earlier"

Comments (6), (7) and (8) all support the argument that if the instructions for the user are considered difficult to use or understand, then the instructions are not used. This is, however, not the most likely way to address a problem, if the person has not much experience on computers or on the subject matter. In this test there were also test persons who became very uncertain if they could not find explicit instructions in the manual. They easily asked for advice or verification, or talked aloud uncertain of what they can do, afraid of making a mistake that cannot be corrected. Their uncertainty is shown in examples (9) and (10) (taken from the video tape):

(9) "dare I press this...ok, I'll press it"
(10) "dare I close it?"

Here it may be that the observer's presence had an effect on the behaviour of the test persons. If the test persons had been alone in the room, they would have had to manage completely on their own, with no chance of

asking the observer for support. In this case they were partly thinking aloud and wondering by themselves whether the window can be closed, but they were also expecting a comment from the observer.

None of the test persons read through the whole manual. They all tried to perform the tasks and solve the problems with as little reading as possible. Example (11) (from the video tape, a comment when filling in the questionnaire) describes the general attitude towards using a manual:

> (11) "I think that if there is a need to use a manual, then people take the table of contents and see what they need, they don't read anything extra, maybe introduction"

As example (11) indicates, the usual way to use user's manuals is just to look for information necessary for carrying out the task at hand. The feedback from the test persons in this test confirms Nyyssönen's (1997: 116-117) view that user's manuals are often read quickly and under pressure (see page 19 in this thesis), and manuals are used as a reference guide, not as something that is read through. After the test, however, many test persons commented that in order to form an overall picture of what the manual includes, it would have been good to go through the whole manual first. Their opinions after the test are shown in comments (12), (13) and (14) (from the video tape):

> (12) "it could have helped if I had looked through the manual in the beginning, now I had no general picture but I just started doing something"
> (13) "I didn't read from the beginning, if I had done that, then maybe I would have understood better"
> (14) "I should have gone through the manual first in order to form an overall picture"

In comments (12), (13) and (14) the test persons agree that it might have helped if they had gone through the manual first. This probably would have helped, especially since the subject matter was still a bit unfamiliar for them, and they did not know what to expect from the manual. The case is different if one reads a user's manual of a somehow familiar product. For

example, most people know what to expect from a user's manual for a television or for a coffee machine. However, it can never be assumed that a novice reader knows what the manual includes. There should be clear signs in the manual what kind of information each part includes, so that one does not have to read unnecessary information.

People have different strategies when they have to solve problems. Next there is a description of one 'survival strategy' in the test. This is an example of using different means to solve a problem. The test person (user of manual version A) first read the instruction for subtask 1 *Change the program from Browse mode to Edit mode*. The test person took the manual and first checked the table of contents. She chose chapter *4.4 Browse mode toolbar buttons*, but could not find from that chapter what she was looking for. She started skimming the manual, leafing through pages and reading. Then she read through other instructions for subtasks to see if they helped, but they did not. Next she read the introduction of the manual, because she drew a conclusion that the instruction must be somewhere in the beginning (comment 15 is taken from the video tape):

> (15) "because this is the first task, the instructions for it are probably in the beginning of the manual"

The introduction, however, did not help, so she continued skimming the manual, and finally found the right instructions, and was able to complete the task. This example shows the different ways to address a problem. The reader tried to find the information needed in many ways, and also tried concluding, until she finally found the right information by chance. As in this example, most test persons used different methods in order to solve the problems, but some gave up (and asked) easier than others.

4.3 Errors and problems

I observed whether the test person made mistakes, what kind of mistakes and in which situations. I was also interested to know whether they were corrected, and how they were corrected. I wanted to see whether there were any (or a lot of) unnecessary actions. I also wanted to find out whether there was an obvious lack of understanding – did the reader always know what to do and how? If the reader asked for assistance, was it due to a problem with locating the right source for information, a problem with understanding, a problem with the formulation of the given task, or something else? In the following sections there are descriptions of the problems that the readers had with terms, instructions or with other matters. First there is a description of a problem and then there are my suggestions of how the manual could be improved in order to avoid the problems.

4.3.1 Problems with terms

Browse mode and Edit mode

In subtask 1 the test persons had to change the programme from Browse mode to Edit mode. Most test persons found it difficult to understand what the different modes, *Browse* and *Edit*, meant. When looking for information on Browse and Edit mode, they were reading or skimming chapters *4.3 Menu commands*, *4.4 Browse mode toolbar buttons* and *4.5 Edit mode toolbar buttons*, but were still confused about what these terms meant (example 16 is taken from the video tape and example 17 is from the questionnaire):

> (16) "here is something about the modes but I don't know how to switch between them"

> (17) "I found the instructions for both Browse and Edit, but not to the fact how one can change from one mode into another."

Examples (16) and (17) show that the test persons could not locate the right instruction for changing the programme from Browse mode to Edit mode. The instruction does, however, exist. It is in chapter *4.5 Edit mode toolbar buttons*, but it is not clearly enough indicated, especially since the terms confused the readers. Therefore, they were not exactly sure what they were looking for. Example (18) (from the video tape) shows the test person's confusion when the expected chapter could not be found.

> (18) "should I read this from the beginning, since I can't find...I'm trying to find some chapter where it would say changing..."

In (18), the test person had an idea of what to look for, she was looking for words that would match between the task and the manual (and also user interface). She had created a model in her mind of what to expect from the manual. There should be clear signs in the manual about the purposes the instructions can be used for, so that the manual user does not have to read extra information or start guessing where the needed information could be found.

There were also positive notions in the test. When the test persons could not find the instruction they were looking for, they tried to solve the situation somehow by reasoning. For example, one test person did not find the right instruction, and was reading aloud chapter *4.4 Browse mode toolbar buttons*:

> (19) "when you open...so when you open Rule Generator it is in Browse mode, and from Edit mode you can switch to Browse mode by pressing Browse toolbar button, so probably that works the other way around, too"

In the example (19) above, the test person concluded that because there is a Browse toolbar button, there must be an Edit toolbar button, too, and tested

the buttons in the toolbar with the mouse pointer, until the right button was found. This shows an ability to infer even when there are no clear instructions. However, the instructions should be explicit, so that there would be no need for reasoning, because when concluding things there is always the danger of drawing wrong conclusions.

On the basis of these examples I suggest that in the manual before chapter *4 Editor Appearance* there could be a separate (introductory) chapter explaining generally that the programme has two modes, a short explanation of what can be done in each mode, and a description of how the user can change the programme from one mode into another (and in what kind of situations that change is needed). In this introductory chapter there could be references to the following chapters, e.g. "for a more detailed description of modes, see chapters Browse mode toolbar buttons and Edit mode toolbar buttons".

Data Area

In step 2. the test persons had to select Data Area. The test persons had minor problems with understanding the term *Data Area*. They were able to conclude what a *Data Area* is, but that was partly because the given task was so clear (select five rows). The test persons who used manual version A commented that the figures and captions were helpful and they clarified what the *Data Area* meant. However, the definition could be made clearer also in the text. In manual version B there was no figure showing the selection of *Data Area*, just a picture of tags that should appear around the selected area.

Entry area

Subtask 3 was to select entry area. The term *Entry Area* was very confusing for the readers. First of all, the area has two names in the manual, *Entry Area* and *All Tags Area*, and some found this confusing.

According to controlled language rules one name should be used for one thing, consistently. My suggestion for the improvement of the manual is to use one term only (in chapter 6.2 in both manual versions). If it is not absolutely necessary to use two terms, term *All Tags Area* could be left out from the manual.

The manual's definition of *Entry Area* turned out not to be clear enough (in chapter 6.2). Many test persons had problems with understanding it, as can be seen in examples (20) and (21) (taken from the video tape):

> (20) "what does entry area mean...it is a bit awkward when you don't know the actual meaning of the terms"
> (21) "is it entry area that I have already selected...I'm a bit lost now"

Obviously, the term Entry Area should be clarified in the manual. In case of manual version A, the figures and captions received good feedback from the test persons. Comment (22) (from the video tape) is on figure 5 (in version A) which explicitly explains how to select the Entry Area:

> (22) "in the caption it is explained more clearly what an entry area is"

Also, the information in Note in chapter 6.2 (explaining that the first row should not be chosen) comes too late, many had already automatically selected the first row in *Data Area* as an *Entry Area*, before they continued reading and noticed that this should not be done. The test person who had selected the first row before reading the information in Note commented as follows (on the video tape):

> (23) "this could have been said earlier, before the instruction to select an entry"

My suggestion for improvement is that in chapter 6.2 there should first be a more clear definition of the term: "An *Entry Area* must be within *Data Area*. It includes a piece of *Data Area*, i.e it is, for example, one row, where all the fields occur only once. You should not choose the first row...". Only after this there would be the first instruction on how to select the area.

Field

In subtask 4 the test persons had to select and name fields. The term *Field* was, however, problematic for them, as examples (24) and (25) (from the video tape) show.

> (24) "now I need to find out what that field is...this does not explain it"
> (25) "I wonder what those prospective fields are"

Examples (24) and (25) show that there should be a clearer definition of a *Field* in the manual. It took unnecessarily long for the test persons to find out what they should select as a field in this subtask. In case of manual version A, the figures and captions helped the readers to figure out what all these terms meant, so they were considered good. In version B there was only a small picture of tags that should appear after the selection, and the users of manual version B hoped for more pictures. For more information on the effect that pictures have on usability, see chapter 4.5.

Some test persons had problems with the wording *naming the field* at first. The reason for it was the formulation of the task: select and name the fields. In the manual the test persons could find a chapter *Selecting fields*, but before reading the chapter through, they were wondering about the naming part (example 26 is from the video tape):

(26) "now I don't understand how the fields are named"

This is, however, not essential from the point of view of the manual's usability, because this minor problem was caused by the formulation of the task. Some readers were confused when they found a chapter about selecting fields but no chapter about naming fields (it is included in the chapter about selecting the fields).

Other problematic terms

One test person mentioned that in chapter *6.2. Selecting Entry Area*, in the second Note the word *reside* could be replaced with an easier synonym. The word *reside* is in this sentence: *Obviously, the Entry Area must reside within the Data Area.* According to Simplified English/controlled language rules one should keep the language as simple as possible, and here it is possible to use a simpler alternative, for example, verb *be.*

There were also other terms that were problematic in some cases, for example, *generating the rules*, *XML conversion rules* and *toolbar button.* Generating the rules and XML conversion rules were unfamiliar concepts and were difficult for most test persons. Especially generating the rules sounded more complicated than it actually was (the test persons only had to press one button to generate rules). In at least one case the test person was wondering what a toolbar button is, before correctly concluding what it meant.

Many test persons hoped for a glossary, where the problematic terms would have been explained (example 27 is from the video tape and example 28 is from the questionnaire):

> (27) "a glossary might be quite good...but then how do you notice that there is a glossary, it should be marked somehow...but I also started looking for some kind of a term list from the beginning of the table of contents"

> (28) "The terms were sometimes unclear. It would have helped if there had been a glossary in the manual, explaining what the most general terms (select, generate, extract, entry/data area etc.) mean in this context."

One test person suggested that including a separate glossary in the manual would be better than longer explanations in the text. Then people who need more information in order to understand the instructions could consult the glossary, but people who have more experience and can cope with less detailed instructions would not have to spend time on reading information that they find unnecessary.

In general the users of manual version A and B had similar problems with the difficult terms. It was, however, easier for users of manual version A to solve the problems, because the figures in the manual helped in defining what the terms meant, at least in parts where areas had to be selected.

The results of the test indicate that mostly the problems with terminology were caused by technical words, but not always. In this sense the results of my study are in agreement with Isomursu's study (see p. 19-20 in this thesis). Based on the analysis, it is clear that unfamiliar vocabulary slowed down the reading process. However, not just technical words were causing problems, also non-technical words were problematic when they appeared in an unfamiliar context. For example, word *select* was not quite clear in this context, where it meant 'painting with the mouse pointer'. *Select* has several meanings depending on the context, it could, for example, mean just clicking with a mouse. When text usability is considered, the context of reading must always be taken into account, as well as reader's background and the features of the text itself.

4.3.2 Unclear instructions

Selecting an area

In one case the test person did not know that after one area has been selected, the right mouse button must be clicked so that the mouse pointer is over the selected area. The problem occurred in subtask 2 *Select data area*, where an area from the display had to be selected for the first time. It is not explained in the manual that the mouse pointer must be over the selected area. It is regarded as self-evident, but for a novice this is not self-evident.

In this case, my suggestion for improvement is that in chapter 6.1 there could be an addition: "Then, click right mouse button (make sure the cursor is over the selected area) and choose Add Data Area".

Generating rules and extracting data

In subtasks 5 and 6, where the test persons had to generate rules and extract data, they understood the manual's instructions well and acted according to them, although many of them said afterwards that they did not know what they were doing, but just followed the instructions. However, none of them was certain when the rule generation was finished, and whether they had done it right, as the examples (29), (30) and (31) (from the video tape) show:

> (29) "something should have happened now but nothing happened...a bit weird, I should think that it would give some kind of a sign that ok"
> (30) "I wonder what's happening now, nothing...that button just dimmed"
> (31) "what happens now, it seems that nothing is happening, but I guess something is happening"

From a novice's point of view the terms sound complicated, because it is not clear what they mean. It is hard for a novice to believe that something that sounds as complicated as *generating rules for data extraction* is so easily done, just by pressing one button (example 32 is from the video tape):

> (32) "quite weird that generate the rules for data extraction...some rules are created there, and then it means just pressing one button...I was expecting that one can create some fancy definitions and rules of one's own"

My suggestion for the improvement of the manual is that in chapter *6.4 Generating the rules* there could be a more detailed description of what happens, for example: "After you have pressed F-button, wait for a moment. When the hourglass disappears and View Rule Set button becomes available, the analysis is complete". Then the user would not wonder whether something is actually happening, but he/she would wait patiently while the programme processes the data. Additionally, there could be a message on the display, for example, "Please wait, information is being processed."

In general it must be considered how the printed manual, online help and messages in graphical user interface could be improved to function in a smooth co-operation. For example, the error messages should be descriptive and optimally they should also include a solution for the problem. The user does not benefit from a message "an error has occurred", but the message should include information on what kind of an error and what can be done about it. If the online help and messages on the screen were usable, then the importance of printed documentation would decrease.

Previewing XML output

Subtask 7 was to preview XML output in Internet Explorer, which caused some problems. All test persons tried to press the e-button in the toolbar, but failed because the XML-preview window was still open. They were

wondering whether the window should be closed, but were afraid to do so, because there is no explicit instruction in the manual to close the preview window. Comments (33) and (34) (from the video tape) show the uncertainty of the test persons:

> (33) "should I close this?"
> (34) "it doesn't accept it...should I close this somehow?"

After some attempts to press the toolbar button they either closed the window and pressed the toolbar button, or noticed that there is an e-button also in the preview window, and pressed that button.

My suggestion for improvement is that there should be a clearer instruction on this in the manual: "To view the XML-output in Internet Explorer, press e-button on the lower right corner of the preview window. You can close the window by pressing "v"-button (also on the lower right corner of the preview window). To view the XML-output in Internet Explorer after you have closed the window, press "eye"-toolbar button."

The above mentioned problem occurred also in subtask 8 *View the conversion rules*. Again the test persons did not know whether they were allowed to close the previous preview window or not, but finally closed it after unsuccessful attempts to press the toolbar button (examples 35 and 36 are from the video tape):

> (35) "do I have to close this again, although I just got it opened (clicks the toolbar button without success)...apparently it must be closed then"
> (36) "do I have to close that window...it wasn't said anywhere that the window must be closed...and also in this picture the window is open?"

Obviously, an instruction about closing the window should be added in the manual.

4.3.3 Other notions

It turned out in the test that there were parts in the manual that need to be updated. Figure 11 did not match with the preview window on the user interface, so the figure should be changed, so that it would be up-to-date.

There was an unclear reference in chapter *6.6 Previewing XML output in Internet Explorer*, in Note: "...but with *this* button, you can view it...". Two different buttons are mentioned before the reference, and in one case, the reader was confused about which of the two buttons was meant by *this* (example 37 is from the video tape):

> (37) "I wonder where that *this* refers to, which button?"

As a solution for this problem, the name of the button ("eye") should be used instead of the word *this*.

It also turned out that a troubleshooting section would be useful in the manual, i.e. instructions on what to do in which situation, if a problem occurs. It is said in the manual (in both versions) that there is no undo function and to make changes or to start again one has to press *Reload* button. If there was a problem, a test person either followed this instruction or 'panicked' and asked me what to do. Example (38) (from a video tape) is a suggestion from one test person about dealing with errors and correcting them:

> (38) "there could be an instruction in each part that if this goes wrong, then what to do"

It would, however, be difficult to include information everywhere in the manual about what to do if something goes wrong, as comment (38) suggests. But, if there was a separate troubleshooting section in the manual,

the user could easily check what each error message means, and what must be done if something does not function as expected.

4.4 Layout

The layout of the manual versions divided opinions. The layout of manual version A was considered both good and bad. One test person considered the general layout confusing (example 39 is from the video tape):

> (39) "in the margins there is text and then there is normal text, it might also be clear but it confused me, and then when there are a lot of notes and many kind of fonts, I didn't really know what to concentrate on"

However, the same test person commented that when one becomes more familiar with the manual and the style of the manual is not strange anymore, then it also becomes easier to use the manual. Then again, according to another user of manual version A the layout was clear, there was a suitable amount of information in each place (small pieces), and in case of a problem the side headings helped when searching (skimming) for information (example 40 is from the questionnaire):

> (40) "When it was clear what to do, I did not pay much attention to the side headings. When there were problems, I paid more attention to them."

Information Mapping® labelling principle recommends the use of labels (subheadings, side headings) which give a preview for the reader of what each section or paragraph includes. Example (40) implies that the side headings helped in case there was a problem. Other test persons did not comment on the use of side headings.

One test person commented that there are too many lines which do not, for example, logically divide chapters, as the reader had expected. Others did not comment on the lines, so their attitude towards them seemed to be

neutral. In manual version B there were not any side headings or lines, so there were also no comments regarding this part of layout. One test person mentioned about the version B that the text is too dense, and does not tempt to read (example 41 is from the video tape):

> (41) "the introduction is too dense, forbidding, the line spacing should be at least 1,5"

To make the text look more tempting and easier to read, the line spacing should be more than one, or then there should be short paragraphs, not too much text put together. Short paragraphs are easier to read and skim. When looking for certain information one does not read through the paragraphs, but usually it is enough to read the first sentence to find out whether the paragraph includes useful information.

According to the test persons who used manual version A it was good that Notes and Cautions were clearly separated from the body of text by emphasising them, as example (42) (from the video tape) indicates:

> (42) "notes and cautions are clearly presented"

There were, however, too many Notes. The test persons found it disturbing that the text flow was interrupted so often, and often this was without a good enough reason. Consider example (43) (from the video tape):

> (43) "marking the notes and cautions was good...there shouldn't, however, be unnecessary ones, they must be useful or then one doesn't bother to read them, when one notices that they don't include anything important"

For example, on page 18 in the manual (in version A, see page 113 in Appendix 5) there is a Note four times, and only little text. A solution for this problem is that the information from some Notes is changed to be in a normal text. This should be done throughout the manual. Notes should be reserved for information that truly includes important information for the user.

In case of manual version B, the test persons complained that the information in Notes cannot be noticed, it looks too much like the body of text. In manual version B the Notes were written the same way as the body of text (same font and font size). Comments (44) and (45) (from the video tape) show that this is not an optimal way of presenting important information:

> (44) "notes could jump up from the text more, for example, a bigger font and inside a box...and on the side there could be for example an exclamation mark to draw attention"
> (45) "notes do not differ from the normal text well enough"

It helps the manual user if certain things are always presented in the same way (e.g. cautions and warnings with the same font type and size which differs from normal text). When one gets used to the style then it is easier to find information and notice the important things. However, in case of manual version A it must be considered whether the layout should be simplified a little, so that it would not be confusing even when used for the first time. Perhaps there could be less lines, not a line between every topic change. The amount of Notes must be cut down, only truly important information should be included in Notes. It also must be considered whether there are too many means of emphasis for a Note. It differs from the normal text in the following: indentation, different font size, bold, lines separating it from the body of text and a figure in the margin. Notes and Cautions do jump out of the body of text, but less emphasis would probably have enough effect.

From the above, a conclusion can be made that it helps the manual user if information is organised according to the level of importance. Important matters should be clearly visible to the user, they should be emphasised. Non-important matters, however, should be treated in an opposite way, i.e. they should not be emphasised. Emphasising non-important matters confuses the reader, e.g. in the test the readers became frustrated with the

emphasised Notes, they considered the Notes somehow deceptive, because in many cases the Notes did not include as important information as expected. The writer should carefully consider what information is relevant from the point of view of user's action, and organise the information in the manual accordingly.

4.5 Visual design

In case of manual version A, the figures and captions helped the readers figure out what difficult terms meant, so they were considered good, as the examples (46), (47) and (48) indicate (examples 46 and 48 are from the video tape, example 47 is from the questionnaire):

> (46) "in the caption it is actually more illustratively explained than in the text...I would still want a bit more clarification about what a field is"
> (47) "the figure was illustrative, for example, selecting fields, otherwise I would not have figured out what a field is"
> (48) "the figures were good, for example, the field part I could figure out on the basis of the picture"

Also in case of manual version B the test persons got support from figures. In parts with instructions of selecting areas there were no figures showing how to select the areas, there were just small pictures of tags, showing how they should appear. The test persons compared the tags with the picture (example 49 is from the video tape):

> (49) "the result should be those"

So, even little visual signs helped. However, all test persons who used manual version B hoped for more pictures (examples 50 and 51 are from the video tape, example 52 is from the questionnaire):

> (50) "in unclear parts there could have been more pictures, they helped a lot in illustration, because you could check whether the view on the display was the same...the pictures of buttons in the middle of the text were very good"
> (51) "there could be pictures in those parts where areas are marked up, so that you would be able to see a model"
> (52) "The pictures were good and made this more clear, but a visual learner always hopes for more pictures."

All test persons were pleased with the small pictures of buttons in the middle of the text. In version A there were pictures of buttons always when the button was mentioned, in version B the picture was not always added, and it turned out that it would be helpful for a novice reader if the picture was always included:

> (53) "if you want a foolproof manual, then the pictures of buttons could always be added in the text when the button is mentioned"

In conclusion, the users of both manual versions regarded the pictures in the manual helpful. Users of manual version A especially mentioned the captions, they said it helped a lot when the figure was explained in detail in the caption. In manual version B the captions were a lot shorter, and the users of manual version B hoped for more pictures.

The integrated graphics principle in Information Mapping® method recommends that graphics are used to illustrate information whenever possible. In manual version A a lot of pictures have been used, and because of this the manual version A turned out to be more usable than version B, which includes less pictures.

4.6 Language

In general the language in the manual version A was considered good and easy to understand, just unclear terms were considered difficult, as example (54) (from the questionnaire) shows:

> (54) "As regards language, the text was very understandable, but the terms were confusing and needed clarification."

Also the users of manual version B were quite satisfied with the language, as comment (55) (from the video tape) indicates:

> (55) "the language is quite easy, simple verbs, just terms are difficult"

However, users of manual version B in general needed more time to read the instructions, or even had to read them several times. For example, in subtask 2 *Select data area* a test person who used manual version B read the instructions in chapter 6.1 several times, before realising what had to be done. Example (56) (from the video tape) below shows that the instructions were not as clear as they could have been:

> (56) "a simple matter has been explained in a really complicated way"

The use of the passive voice confused the users of manual version B. Another test person read the instructions for the above mentioned subtask, but found them unclear – because of the passive voice it was unclear whether something has been done already or if something should be done, as can be concluded from example (57) (from the videotape):

> (57) "well how can it be selected already, shouldn't it say here that will be selected, if I have to select something?"

It is clear that manual version A, which was written according to controlled language and Information Mapping® rules (the rules of using imperative and active voice in instructions), was more usable. This conclusion can be drawn because the use of the passive voice in the instructions in version B proved to be confusing and time-consuming for the readers.

4.7 Time used, goal achievement

The time used for performing the task was measured, because I expected that the difference in the time spent with different versions indicates usability or non-usability. The test persons were told in the beginning that the time is measured, but that they do not have to hurry, instead, they could use all the time they felt they needed for the task. The average time for using manual version A was 27 minutes, and for using manual version B the average time was 37 minutes. In other words, it took longer to perform the tasks with manual version B, but the difference was not remarkable, therefore general conclusions on the usability of texts on the basis of the time used are difficult to make.

In general the time needed for performing the task was longer than expected. It can be concluded that neither of the manual versions was optimally usable because of the time needed for the task. However, there are some reasons that could have affected the speed of performance in the test. First, the test situation perhaps made the test persons nervous and it was harder for them to concentrate. Second, the subject matter was unfamiliar, and the test persons spent time wondering what strange terms meant. If a test person made a fatal mistake and had to start from the beginning, the procedure was a lot faster the second time. On the basis of this, it can be assumed that the performance would have improved and become a lot faster if the test persons had done the same task twice, so after all, the learning curve in the test seemed to be short.

All the test persons were able to perform the task, i.e. achieve the goal, because in cases where they could not solve problems by themselves (with the manual), they got advice from the observer. Task 1 was most difficult for the test persons, almost everyone had to ask for help when changing the programme from Browse mode to Edit mode. If a test person had to ask the observer for help, that was a clear sign that the manual had not functioned in

its purpose. In those cases it was best to help the test person to continue, otherwise the test would have ended prematurely and a lot of important information could have been left out from the study.

4.8 User satisfaction

After the test the test persons assessed the difficulty of the task on the scale of 1-4 (1=easy, 4=difficult). The average was ca. 2,8. There was no remarkable difference between the users of different manual versions on how difficult the task was in their opinion: the average of users of manual version A was ca. 2,7 and the average of users of manual version B was 3. Most test persons said that the task itself was not very difficult. The task was considered suitable, not too difficult and not too long. In the beginning the test persons were a bit nervous about the test situation, but this did not affect the performance much after the start (examples 58 and 59 are from the questionnaire):

> (58) "the task itself did not feel difficult otherwise than in the beginning, before one got started with the manual and the task"
> (59) "not complicated once one got started"

However, the test situation had some effect on the performance. One test person commented that the situation was not disturbing, and the camera did not disturb, but for example at home one would have been able to do the test in a more natural setting.

All test persons had some difficulties during the test. Reasons for difficulties were, for example, unfamiliar terms, unclear instructions in the manual and little experience on computers. Also, many test persons had a general feeling of not knowing what one is doing (example 60 is from the questionnaire and example 61 is from the video tape).

(60) "it was difficult to understand why each thing had to be done"
(61) "if you would ask now what I did, I wouldn't be able to say anything"

Both users of manual version A and B agreed that there is a need to develop the manual further. Neither of the versions was considered bad, but not optimally usable either.

5 RESULTS AND CONCLUSIONS

My first research question concerned the effect of different textual features on document usability. In my hypothesis I assumed that manual version A is more usable than version B in the test. It turned out that manual version A was in general more usable than version B. However, the difference between the usability of these two manual versions was not as big as I expected. There were problems that occurred with both manuals, and then some additional problems that occurred with manual B. In general it can be said that the users of manual version A solved the problems faster and more easily. This was because in version A the pictures helped, language was clear and therefore the instructions were easy to understand.

The second research question concerned the optimal way of presenting information for novice readers. My hypothesis was that version A is more understandable for novice users than version B. Although version A turned out to be more usable and understandable, as such it is not optimally usable for novice readers. Version A is a published customer document, and the same document is used for all types of readers and users. However, at the moment it is not assumed that total novices would start using the programme and reading the manual, so the manual's target group is not novices, but more advanced readers. The test strengthened the view that it would be best to write manuals separately for each target group, so that the contents would be customised according to the target group's skills. For novice readers there should be more detailed descriptions and more pictures for visual learning, nothing should be expected to be self-evident. Terminology should be explained in detail. For expert readers less information is enough. It is, however, difficult to say what level of detail is correct for each user group. For example, it can be said that core information is enough for expert readers, but where is the line between core information and "extra", unnecessary information drawn? It would be

interesting to organise usability tests for readers with different background skills, so that the correct level could be determined – how much information is needed for each group. The problem with customising the contents is usually the lack of time and money. The deadlines for publishing manuals are often tight, and there is enough work to do to get the contents factually correct. However, it would improve the usability of manuals if they were written with the target group in mind, and preferably the manuals should also be tested in order to check that they actually work from the point of view of the manual user.

In my third research question I was hoping that the test would reveal design problems in the manual. The test was successful – it revealed design problems and provided a lot of information on how the manual could be improved (e.g. by explaining difficult terms and by making unclear instructions explicit). The improvement of a specific manual was the most visible, concrete benefit that this study has resulted in. My suggestions for the improvement of the manual were handled in the analysis. These concrete suggestions have already led to improvements. However, the analysis includes also points to consider in the future document design, e.g. the importance of visual design, a glossary and online help. General conclusions about improving manuals are difficult to make on the basis of this study, because the test group was small and the contents of only one manual were tested for usability. However, the test was able to provide some ideas that could help in improving the manuals in general. The test provided an insight into the point of view of a manual user, and showed that, for example, by customising the contents and by organising information according to the level of importance, usability could be improved.

In this study the material used was printed documents. An interesting field for future research would be the usability of online help (electronic documents). The need for printed documents would decrease if online help were designed usable, and if the graphical user interface functioned well, in

other words, help was provided for the user immediately in case of a problem. This requires that the user's possible mistakes are taken into account well in advance. In addition to studying the usability of online help, it would also be interesting to study the usability of a glossary and index. In this test, there was no glossary, index, or online help available, so their usability could not be tested.

The analysis revealed that in general the readers read as little as possible when looking for information. If the manual does not help, people become frustrated and the manual is easily thrown aside. People are eager to experiment and do not want to spend time searching for information. Explicitness of instructions in the manual is important. If the instructions are unambiguous or vague, people become uncertain and are afraid of making mistakes. There should be no need for guessing what the instructions in the manual mean. It also turned out in the test that a troubleshooting section would be useful, it would help the reader if there was a place for checking instructions in case of a problem.

The layout of the manual versions divided opinions. In both manuals there were aspects that could be improved, but also aspects that received good feedback. The problem with version A was that although Notes were clearly presented, there were too many of them. The use of side headings received both positive and negative feedback. In general the layout of version A could be simplified a little. Short paragraphs in version A were considered good. In contrast to this, the paragraphs in version B were considered too long and the text too dense.

It turned out in the test that visual design (an extensive use of graphics) helps the readers a great deal. In case of version A, the figures and captions helped the readers figure out what difficult terms meant. Also in case of manual version B the test persons considered the figures helpful, but they all hoped for more pictures.

Language in version A was in general considered good and easy to understand, but terminology was considered difficult. The use of imperative and active voice in instructions received positive feedback. Language in manual version B was also considered quite easy, just the difficult terminology caused problems. However, in general the users of manual version B needed more time to read the instructions, or had to read them several times. It turned out that the use of passive voice made the instructions harder to understand.

As already mentioned, terminology was difficult for the readers, therefore, a glossary could have been of great help. Many test persons hoped for a glossary. It turned out that the explanations of terms in the manual were inadequate for novice readers. Unfamiliar (technical) terms slowed down the reading process, but also some non-technical terms were problematic when they were in an unfamiliar context.

In conclusion, version A turned out to be more usable, although the difference to the usability of version B was not as big as I expected. However, even version A was not optimally usable for novice readers. The usability test was able to provide information on the needs and point of view of novice readers. The test also revealed design problems in the manual, and many problems caused by difficult terminology, unclear instructions etc. could be corrected after the test. Thus, the manual could be improved according to the results of the usability test. The next step would be to arrange usability tests in order to verify that the manual has become more usable after the corrections. Usability should be controlled at regular intervals as manuals are updated.

This was a case study on the features of one manual (of its different versions), and the main purpose was to find out information concerning the usability of that manual. Even though the test group was small, the test was

able to provide information about how people process information and on the textual features that affect readability and text accessibility. This was not a typical empirical study (I was testing certain texts with a small test group), therefore, the main purpose was not to derive general conclusions on the usability of texts. However, the results of my analysis indicate that certain textual features that follow the rules of controlled language and Information Mapping® method improve the usability of technical texts, and thus they could be recommended to be used in general. The test was also able provide a lot of information on how the tested manual can be improved, and what could be done in order to customise the manual for novice readers.

BIBLIOGRAPHY

AECMA Simplified English 1998. Air Transport Association Publications, CD-ROM.

Atlas, M.A. 1981. The User Edit: Making Manuals easier to read. *IEEE Transactions on Professional Communication. PC-24*, 28-29.

Carroll, J.M. & Rossen, M. B. 1987. Paradox of the Active User, in Carroll (ed.), *Interfacing Thought: Cognitive Aspects of Human-Computer Interaction*. Cambridge, Mass: MIT press, 80-111.

Duffy, Thomas M. 1985. Readability Formulas: What's the Use?, in Duffy & Waller (eds.), 113-143.

Duffy, T. M. and R. Waller (eds.) 1985. *Designing Usable Texts*. The Educational Technology Series. Orlando: Academic Press.

Elling, Rien 1992. Identifying Communicative and Organizational Text Functions in Evaluating the Usability of Texts, in Pander Maat & Steehouder (eds.), 21-32.

Frase, Lawrence T., Nina H. MacDonald, and Stacey A. Keenan 1985. Intuitions, Algorithms, and a Science of Text Design, in Duffy & Waller (eds.), 97-112.

Halliday M. A. K. and R. Hasan 1976, *Cohesion in English*. London: Longman.

Horn, Robert E. 1985. Results with Structured Writing Using the Information Mapping® Writing Service Standards, in Duffy & Waller (eds.), 179-212.

Horn, Robert E. 1997. *Participant's Manual for Developing Procedures, Policies & Documentation*, An Information Mapping® Seminar. Massachusetts.

Horton, W. 1993. Let's Do Away With Manuals...Before They Do Away With Us. *Technical Communication* 40:1, 26-34.

Introduction to Controlled Language, an article in http://www.linga.fr/LingEn/lanconen.htm, 31.10.2000.

Isomursu, Tuija 1997. 'In a way it's a problem you don't know that data terminology' – on the relationship between vocabulary and text accessibility, in Nyyssönen & Kuure (eds.), 85-98.

De Jong, Menno and Pieter van der Poort 1994. Towards a Usability Test Procedure for Technical Documents, in Steehouder et al (eds.), 229-238.

Kaner, Cem and David Pels 2000. Improving User Documentation and Customer Care. An article available at http://www.stcsig.org/usability/newsletter/0004-docsandcustcare.html, 5.11.2001. Reprinted from *Usability Interface*, Vol 6, No. 4, April 2000.

Kern, Richard P. 1985. Modeling Users and Their Use of Technical Manuals, in Duffy & Waller (eds.), 341-375.

Klare, G. 1963. *The Measurement of Readability.* Ames: Iowa State University.

Klauke, Michael 1994. National Standards – Their Impact on Text Production and Quality, in Steehouder et al. (eds.), 161-170.

Lahti, Maria 2000. *Efficiency of Communication in Software User Documentation.* Pro Gradu Thesis, University of Tampere.

Lassen, Inger 1997. Variables of Accessibility in Technical Manuals: A Reader-based Survey of Style Preferences, in Nyyssönen & Kuure (eds.), 35-64.

McNamara, D.S. and W. Kintsch 1996. Learning from Texts: Effects of Prior Knowledge and Text Coherence. *Discourse processes*, 22:3, 247-288.

Miller, G.A. 1967. *The Psychology of Communication.* New York: Basic Books.

Nielsen, Jakob 1993. *Usability Engineering.* Boston: Academic Press.

Nyman, Sirkku 1995. Sequentiality and Iconism in Operating Instructions, in B. Wårvik, S-K. Tanskanen, and R. Hiltunen (eds.), *Organization in Discourse.* Proceedings from the Turku Conference. Anglicana Turkuensia 14, 403-408.

Nyyssönen, H. & L. Kuure (eds.) 1995. *Principles of Accessibility and Design in English Texts - Research in Progress.* Publications of the Department of English, University of Oulu. A. Text and Discourse Studies 12.

Nyyssönen, Heikki 1997. Accessibility and Text-reader Interaction, in Nyyssönen & Kuure (eds.), 111-120.

Nyyssönen, H. & L. Kuure (eds.) 1997. *Principles of Accessibility and Design in English Texts - Research in Progress 2.* Publications of the Department of English, University of Oulu. A. Text and Discourse Studies 14.

Olson, David R. 1985. On the Designing and Understanding of Written Texts, in Duffy & Waller, 3-15.

Orna, Elizabeth 1985. The Author: Help or Stumbling Block on the Road to Designing Usable Texts?, in Duffy &Waller (eds.), 19-41.

Pander Maat, H. & M. Steehouder (eds.) 1992. *Studies of Functional Text Quality.* Utrecht Studies in Language and Communication 1. Amsterdam: Rodopi

Pilto, Risto & Tuija Rapakko 1995. Testing Accessibility of Utility Texts – Work in Progress, in Nyyssönen, H. & L. Kuure (eds.), 37-58.

Purho, Vesa 2000. Heuristic Inspections for Documentation – 10 Recommended Documentation Heuristics. An article available at http://www.stcsig.org/usability/newsletter/0004-docsheuristics.html, 5.11.2001. Reprinted from *Usability Interface*, Vol 6, No. 4, April 2000.

Schriver, Karen. 1997. *Dynamics in Document Design.* New York: John Wiley & Sons.

Schubert, S., J. Spyridakis, H. Holmback and M. Coney 1995. The Comprehensibility of Simplified English in Procedures. *Journal of Technical Writing and Communication*, 347-369.

Schumacher, Gary M. and Robert Waller 1985. Testing Design Alternatives: A Comparison of Procedures, in Duffy & Waller (eds.), 377-403.

Steehouder, M, C. Jansen, P. van der Poort and R. Verheijen (eds.) 1994. *Quality of Technical Documentation.* Utrecht Studies in Language and Communication 3, Amsterdam: Rodopi.

Sticht, T. 1985. Understanding Readers and Their Uses of Texts, in Duffy & Waller (eds.), 315-340.

Sticht, T., Fox, L., Hauke, R., & Welty-Zapf, D. 1997. *The Role of Reading in the Navy.* San Diego, CA: Navy Personnel Research and Development Center.

Wright, Patricia 1994, Quality or Usability? Quality Writing Provokes Quality Reading, in Steehouder et al (eds.), 7-38.

APPENDIX 1: quotations from the video tape and from the questionnaire, and translations into English

(1) "ainakin sisällysluettelo on hyvin selkeä, hyvin löytyy kaikki"
"at least the table of contents is very clear, everything can be found easily"

(2) "taas löytyy hyvin täältä sisällysluettelosta"
"again it is easily found from the table of contents"

(3) "ärsyttää kun en löydä ohjeita!"
"I am annoyed because I cannot find the instructions!"

(4) "en mä löydä sitä...ei tästä tuu yhtään mitään"
"I cannot find it...this is hopeless"

(5) "raivostuttaa kun ei löydä jotain ja sit ei viittis lukee kaikkee"
"it makes me mad when I can't find something and then I wouldn't be bothered to read everything"

(6) "mieluusti tekis ennen kuin lukis, tekis mieli kokeilla vaan"
"I would rather do than read, I would like to just have a try"

(7) "Jos mä vaan painan jotain nappia ja kokeilen"
"what if I'll just press some button and try"

(8) "jos tää ois ollu oikee tilanne ni sitten mä vaan kokeilisin, ja oisin kokeillu jos paljon aiemmin"
"if this had been a real situation, then I would just try things out, and would have tried a lot earlier"

(9) "uskallanko painaa tätä...ok, mä painan sitä"
"dare I press this...ok, I'll press it"

(10) "uskallanks mä sulkee sen?"
"dare I close it?"

(11) "musta tuntuu että jos tarvii jotakin manuaalia käyttää niin ihmiset ottaa sisällysluettelon ja kattoo mitä tarvii, eikä lue mitään ylimääräistä, ehkä introductionin"
"I think that if there is a need to use a manual, then people take the table of contents and see what they need, they don't read anything extra, maybe introduction"

(12) "olis voinu auttaa kun ois kattonu manuaalin alussa kokonaan läpi, nyt ei ollut alussa mitään kokonaiskäsitystä vaan alotti vaan jotakin tekemään"
"it could have helped if I had looked through the manual in the beginning, now I had no general picture but I just started doing something"

(13) "en lukenut alusta asti, jos ois lukenut alusta niin ois ehkä paremmin ymmärtänyt"
"I didn't read from the beginning, if I had done that, then maybe I would have understood better"

(14) "ensin olis kannattanut käydä koko manuaali läpi kokonaiskuvan muodostamiseksi"
"I should have gone through the manual first in order to form an overall picture"

(15) "koska tää on eka tehtävä ni luulis että ohjeet löytyy manuaalin alusta"
"because this is the first task, the instructions for it are probably in the beginning of the manual"

(16) "täällä puhutaan modeista mutta en tiedä miten niitten välillä siirrytään"
"here is something about the modes but I don't know how to switch between them"

(17) "Löysin ohjeet sekä Browseen että Editiin, mutta en siihen, miten yhdestä voi siirtyä toiseen."
"I found the instructions for both Browse and Edit, but not to the fact how one can change from one mode into another."

(18) "pitääköhän mun lukee tää niinku alusta asti läpi, kun en löydä...mä yritän ettiä jotain sellasta kappaletta missä lukis changing..."
"should I read this from the beginning, since I can't find...I'm trying to find some chapter where it would say changing..."

(19) "when you open...eli kun avaat Rule Generatorin niin se on Browse modessa, ja Edit modesta pääsee Browse modeen painamalla Browse toolbar buttonia, joten varmaan tuo toimii toisinkin päin"
"when you open...so when you open Rule Generator it is in Browse mode, and from Edit mode you can switch to Browse mode by pressing Browse toolbar button, so probably that works the other way around, too"

(20) "mitä tarkottaa entry area...vähän kummallista kun ei tunne termien varsinaista sisältöä"
"what does entry area mean...it is a bit awkward when you don't know the actual meaning of the terms"

(21) "entry areako mulla on jo valittuna...mä oon nyt kyllä vähän pihalla"
"is it entry area that I have already selected...I'm a bit lost now"

(22) "kuvatekstissä tulee selkeemmin esiin mikä on entry area"
"in the caption it is explained more clearly what an entry area is"

(23) "tän ois voinu sanoo aikasemmin, ennen ku käskettiin valitsemaan entry"
"this could have been said earlier, before the instruction to select an entry"

(24) "mun täytyy nyt ettiä et mikä tuo on tuo field...ei tää selitä sitä"
"now I need to find out what that field is...this does not explain it"

(25) "mitähän nuo on nuo prospective fields"
"I wonder what those prospective fields are"

(26) "nyt en kyllä ymmärrä miten alueet nimetään"
"now I don't understand how the fields are named"

(27) "sanasto vois kyllä olla aika hyvä..mutta miten sen sit huomaa että siellä on olemassa sanasto, sen pitäis olla jotenkin merkitty..mut mäkin kyllä rupesin ettimään sisällysluettelon alusta onko siellä jotain termistöselitystä "
"a glossary might be quite good..but then how do you notice that there is a glossary, it should be marked somehow...but I also started looking for some kind of a term list from the beginning of the table of contents"

(28) "Termit olivat välillä epäselviä. Olisi auttanut jos manuaalissa olisi ollut termihakemisto, jossa olisi määritelty mitä yleisimmät termit (select, generate, extract, entry/data area jne.) tässä yhteydessä tarkoittavat."
"The terms were sometimes unclear. It would have helped if there had been a glossary in the manual, explaining what the most general terms (select, generate, extract, entry/data area etc.) mean in this context."

(29) "nyt ois pitäny tapahtua jotain mut mitään ei tapahtunu...aika jännä, luulis et se antais jotain merkkiä että ok"
"something should have happened now but nothing happened...a bit weird, I should think that it would give some kind of a sign that ok"

(30) "no mitähän siellä nyt tapahtuu, ei mitään...tuo meni vaan pimeeks tuo nappula"
"I wonder what's happening now, nothing...that button just dimmed"

(31) "mitä nyt tapahtuu, ei näytä mitään tapahtuvan, mutta ilmeisesti jotain tapahtuu"
"what happens now, it seems that nothing is happening, but I guess something is happening"

(32) "aika kummallista että generate the rules for data extraction...siinä luodaan jotain sääntöjä, ja sit se onkin vaan että painetaan yhtä nappia...mä ootin että siinä saa luoda jotakin omia hienoja määritelmiä ja sääntöjä"
"quite weird that generate the rules for data extraction...some rules are created there, and then it means just pressing one button...I was expecting that one can create some fancy definitions and rules of one's own"

(33) "pitäskö tää sulkea?"
"should I close this?"

(34) "ei se huoli...pitäskö tää sulkee jotenkin?"
"it doesn't accept it...should I close this somehow?"

(35) "pitääks mun nyt sulkee tää taas, vaikka just sain sen auki (klikkaa toolbar buttonia, ei onnistu) ilmeisesti se on sit pakko sulkee"
"do I have to close this again, although I just got it opened (clicks toolbar button without success) apparently it must be closed then"

(36) "pitääks tuo ikkuna sulkee...ei missään sanottu että ikkuna pitää olla kiinni...ja tässä kuvassakin on ikkuna auki?"
"do I have to close that window...it wasn't said anywhere that the window must be closed...and also in this picture the window is open?"

(37) "mihinkähän tuo *this* viittaa, kumpaan nappulaan?"
"I wonder where that *this* refers to, which button?"

(38) "joka kohdassa voisi olla, että jos tämä menee pieleen niin mitä pitää tehdä"
"there could be an instruction in each part that if this goes wrong, then what to do"

(39) "marginaalissa on tekstiä ja sitten normaali teksti, se saattaa olla selkeekin mutta mua se hämäs, ja kun väleissä on paljon noteja, ja monenlaista fonttia, ei oikein tiennyt mihin keskittyä"
"in the margins there is text and then there is normal text, it might also be clear but it confused me, and then when there are a lot of notes and many kind of fonts, I didn't really know what to concentrate on"

(40) "Kun oli selvää mitä tehdään en kiinnittänyt niin paljon huomiota sivussa oleviin väliotsikoihin. Kun oli ongelmia, niihin kiinnitti enemmän huomiota."
"When it was clear what to do, I did not pay much attention to the side headings. When there were problems, I paid more attention to them."

(41) "intro on liian tiheä, luotaantyöntävä, riviväli saisi olla ainakin 1,5"
"the introduction is too dense, forbidding, the line spacing should be at least 1,5"

(42) "notet ja cautionit on selkeästi esitetty"
"notes and cautions are clearly presented"

(43) "notejen ja cautionien merkitseminen hyvä...ei saisi kuitenkaan olla turhia, niistä täytyy olla oikeasti jotakin hyötyä tai sitten niitä ei jaksa lukea, kun huomaa ettei niissä ole mitään tärkeää"
"marking the notes and cautions was good...there shouldn't, however, be unnecessary ones, they must be useful or then one doesn't bother to read them, when one notices that they don't include anything important"

(44) "notet voisi tulla enemmän esiin, vaikka isompi fontti ja laatikon sisällä...ja vieressä voisi olla vaikka huutomerkki joka kiinnittää huomion"
"notes could jump up from the text more, for example, a bigger font and inside a box...and on the side there could be for example an exclamation mark to draw attention"

(45) "notet ei eroa leipätekstistä tarpeeksi hyvin"
"notes do not differ from the normal text well enough"

(46) "kuvatekstissä on oikeestaan havainnollisemmin selitetty ku tekstissä...vielä kyllä kaipais vähän niinku selvitystä siitä mikä on field"
"in the caption it is actually more illustratively explained than in the text...I would still want a bit more clarification about what a field is"

(47) "kuva havainnollisti, esim. selecting fields, muuten en olisi tajunnut mikä on field"
"the figure was illustrative, for example, selecting fields, otherwise I would not have figured out what a field is"

(48) "kuvat oli hyviä, mm. field kohdan tajusi kuvan perusteella"
"the figures were good, e.g. the field part I could figure out on the basis of the picture"

(49) "pitäs tulla tommoset"
"the result should be those"

(50) "epäselvissä kohdissa ois voinu olla enemmän kuvia, ne auttoi tosi paljon havainnollistamisessa, kun voi tarkistaa onko näytöllä oleva näkymä sama...tekstin seassa olevat buttonien kuvat oli tosi hyviä"
"in unclear parts there could have been more pictures, they helped a lot in illustration, because you could check whether the view on the display was the same...the pictures of buttons in the middle of the text were very good"

(51) "kuvia voisi olla alueiden merkkaamiskohdissa, että näkis mallin"
"there could be pictures in those parts where areas are marked up, so that you would be able to see a model"

(52) "Kuvat olivat hyviä ja selkeyttivät, mutta visuaalinen oppija kaipaa aina lisää kuvia."
"The pictures were good and made this more clear, but a visual learner always hopes for more pictures."

(53) "jos haluaa idioottivarmaa manuaalia niin nappulan kuvat voisi aina liittää tekstiin kun nappulasta puhutaan"
"if you want a foolproof manual, then the pictures of buttons could always be added in the text when the button is mentioned"

(54) "Kielellisesti erittäin ymmärrettävää tekstiä, mutta termit olivat hämääviä ja kaipasivat lisäselitystä."
"As regards language, the text was very understandable, but the terms were confusing and needed clarification."

(55) "kieli on aika helppoa, yksinkertaisia verbejä, termit vaan vaikeita"
"the language is quite easy, simple verbs, just terms are difficult"

(56) "tosi vaikeesti selitetty yksinkertanen asia"
"a simple matter has been explained in a really complicated way"

(57) "no miten se voi nyt jo olla valittuna, eikö tässä pitäis olla että will be selected, jos mun täytyy valita jotakin?"
"well how can it be selected already, shouldn't it say here that will be selected, if I have to select something?"

(58) "tehtävä sinällään ei tuntunut vaikealta muuta kuin alussa, ennen kuin pääsi manuaaliin ja tehtävään vähän kiinni"
"the task itself did not feel difficult otherwise than in the beginning, before one got started with the manual and the task"

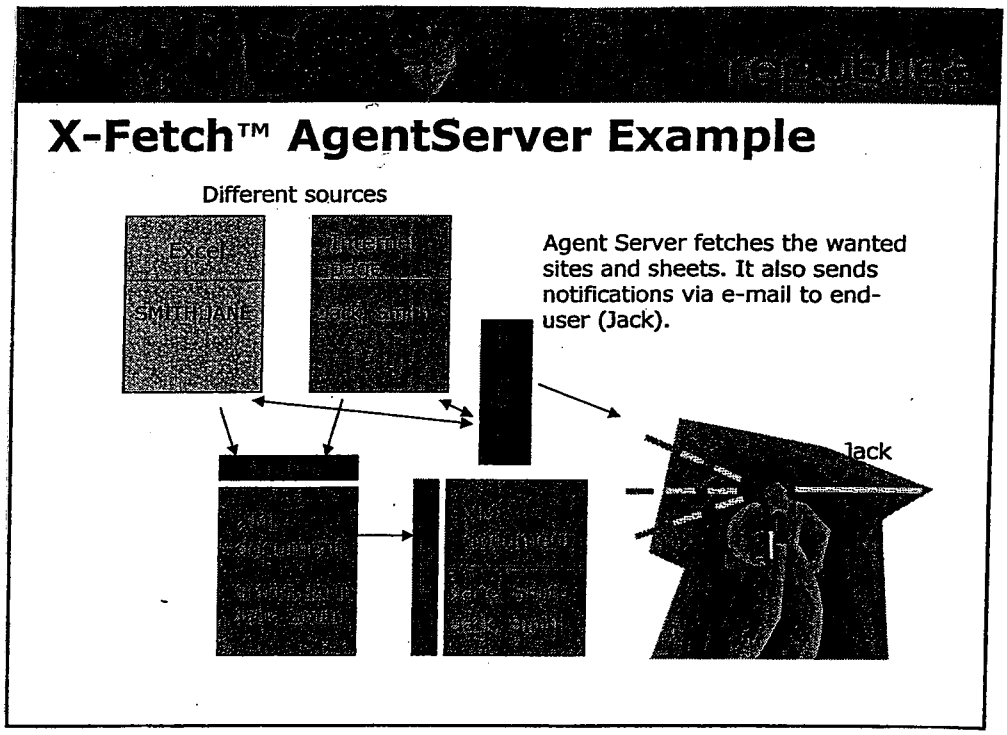
(59) "ei monimutkainen sitten kun pääsi vauhtiin"
"not complicated once one got started"

(60) "oli vaikeata ymmärtää miksi mitäkin piti tehdä"
"it was difficult to understand why each thing had to be done"

(61) "nyt jos kysyisit mitä tein, niin en osaisi sanoa mitään"
"if you would ask now what I did, I wouldn't be able to say anything"

APPENDIX 2

Material for familiarising the test persons with the test subject



Manuaalien käytettävyystesti

Republican X-Fetch Suite–paketti sisältää seuraavat ohjelmistot:

X-Fetch Wrapper, Unifier ja AgentServer.

Testissä käytetään X-Fetch Wrapper Rule Generator –ohjelmaa. Ohjelmalla voidaan valita tietoja erilaisista lähteistä: Internetistä, tietokannoista ym. Ohjelman avulla tiedot konvertoidaan XML-kielelle. Sääntöjä, joiden mukaan konvertointi tapahtuu, voidaan muokata, jotta tulos olisi sellainen kuin halutaan.

Sanastoa:

ASCII = tavallinen tekstiformaatti (ei rakenteinen)

HTML-language = Hypertext Mark-up Language
Internet-sivut kirjoitetaan HTML-kielellä. HTML on myös rakenteinen kieli, jossa tagit merkitsevät eri tekstin osia, itse teksti on tagien sisällä.

XML-language = Extensible Mark-up Language
XML on rakenteinen kieli, jossa käytetään tag'ejä merkitsemään tekstin eri osia. XML on joustava ja monikäyttöinen kieli, joka soveltuu sähköiseen julkaisemiseen. Tag-valikoima ei ole niin rajattu kuin muissa rakenteisissa kielissä.

converting data into XML = valitut tiedot konvertoidaan eli muutetaan XML-kielelle, jotta ne voidaan hyödyntää monella tavalla ja käyttää eri ympäristöissä

DEL = Data Extraction Language, kieli joka määrittelee kuinka alkuperäinen teksti muunnetaan XML-tekstiksi

Java = olio-ohjelmointikieli

user interface = käyttöliittymä

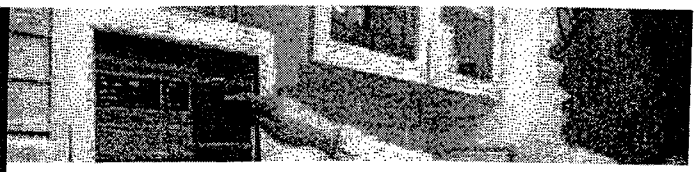
marking up/selecting data = tekstiä valitaan tietokoneen näytöltä maalaamalla

data area, entry area (=all tags area), fields = alueet, jotka näytöltä valitaan on nimetty näin, data alue on näistä suurin ja field alue pienin

to analyse data/to generate rules for data conversion = datan analysointi, ohjelma lukee valitun alueen taustalla olevaa HTML-koodia ja luo säännöt XML-konversiota varten

to extract data = datan 'eristäminen', valittu tieto irrotetaan alkuperäisestä lähteestä ja muunnetaan XML-kielelle

JYVÄSKYLÄN YLIOPISTO
University of Jyväskylä



HAKU

in English

Opiskelu
Hae opiskelemaan
Tutkimus
Tiedekunnat
Yrityspalvelut
Alumnit
Yhteydet muualle
Henkilöstölle

TÄRKEITÄ LINKKEJÄ ■

Atk-keskus
Avoin yliopisto
Chydenius-instituutti
Kirjasto
Museo
Kielikeskus
Soveltavan kielentutkimuksen keskus
Koulutuksen tutkimuslaitos
Täydennyskoulutuskeskus
Ympäristöntutkimuskeskus

NettiTV Visio
Virtuaaliyliopisto

Ajankohtaista ■

- Ranskan kielen ja ruotsin kielen lehtorin virat (kielikeskus)
- Sähköjyvä 14
- Jaakko Pehkosesta Kansantaloudellisen yhdistyksen esimies
- Parinvalinnasta voikin olla geneettistä hyötyä
- 2 professorin ja 2 lehtorin virkaa (informaatioteknologian tdk.)

Lisää ajankohtaisia asioita...

TAPAHTUMAKALENTERI ■

- THE CMCM SEMINAR "Complex Modeling, Computation and Statistical Inference" 09.04.2001
- Vierailuluento, Suomen Rooman-insitituutin säätiön sihteeri, FT Antero Tammisto luennoi 9.4.2001 09.04.2001
- Ylioppilasteatterin Kuolleet sielut -näytelmä 21.03.2001 - 29.04.2001
- Museon kulttuurihistoriallisen osaston näyttely 07.02.2001 - 31.08.2001
- Galleria Pinacothecan näyttely 07.02.2001 - 31.05.2001

Lisää tapahtumia...

PL 35, 40351 Jyväskylä
puh. 014 260 1211
fax 014 260 1021
Yksiköiden osoitteet

www.jyu[1]

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"
"http://www.w3.org/TR/html4/loose.dtd">
<HTML>
<HEAD>
<TITLE>Jyväskylän yliopisto</TITLE>
<META HTTP-EQUIV="Content-Type" CONTENT="text/html; charset=iso-8859-1">
   <link rel="stylesheet" type="text/css" href="textstyles.css">
        <script language="JavaScript" type="text/javascript"
src="jyskripts.js"></script>


</HEAD>
<BODY BGCOLOR="#FFFFFF">
<center>
<TABLE WIDTH="620" BORDER="0" CELLPADDING="0" CELLSPACING="0">
        <TR>
                <TD align="left" valign="top" width="50%" bgcolor="#002d5f">
                        <IMG SRC="images/paasivu4_640x_01.gif"
alt="Jyv&auml;skyl&auml;n yliopiston logo"></td>
                <td>
                        <img src="images/kuvaban4_pc.jpg" border="0"
alt="kuvituskuva" name="kuvax"></TD></TR>
        <TR>
                <TD align="right" valign="top">
                        <IMG SRC="images/paasivu4_640x_04.gif" alt=""></TD>
                <TD align="left" valign="middle" bgcolor="#002d5f">
                <a href="palaute.html" target="_top" class="ylalinkki"><font
color="#ffffff">Palaute |</font></a>
                <a href="http://www.jyu.fi/tiedotus/wwwkartat/" target="_top"
class="ylalinkki"><font color="#ffffff">Kampuskartta |</font></a>
                <a href="persoona/vuolle.html" target="_top"
class="ylalinkki"><font color="#ffffff">Persoona</font></a>
        </TD>
        </TR>
        <TR>
                <TD bgcolor="#ffffff" valign="top" align="left" width="50%">
        <!--haku-->
                        <a href="http://www.jyu.fi/haku.html"><IMG
SRC="images/paasivu4_640x_06.gif" border="0" alt="Haku"></a></TD>
                <TD bgcolor="#4C6D93" background="images/jopo_bg.gif">
                        <span class="etusivu-otsikko">
     Ajankohtaista <img src="images/merkki_cl.gif"
alt=""></span></TD>
        </TR>
        <TR>
                <TD bgcolor="#ffffff" background="images/logo_bg.gif"
width="50%">
                        <p class="paasivu-vasen"><br>
        <a href="http://www.jyu.fi/indexeng.shtml" class="paasivu-vasen"
target="_top">in English</a><br>
        <br>
        <a name="paasivu-vasen"></a><a href="opiskelu.html"
class="paasivu-vasen" target="_top">Opiskelu</a><br>
        <a href="tdk/hallinto/op/valintaopas.html" class="paasivu-vasen"
target="_top">Hae
        opiskelemaan</a> <br>
        <a href="tutkii.html" class="paasivu-vasen"
target="_top">Tutkimus</a> <br>
        <a href="tiedekunnat.html" class="paasivu-vasen"
target="_top">Tiedekunnat</a> <br>

        <a href="palvelut.html" class="paasivu-vasen"
target="_top">Yrityspalvelut</a> <br>
        <a href="alumni/" class="paasivu-vasen" target="_top">Alumnit</a> <br>
        <a href="yhteydet.html" class="paasivu-vasen" target="_top">Yhteydet
muualle</a> <br>
        <a href="henkilosto.html" class="paasivu-vasen" target="_top"
                                onmouseover="show('hkuntal');"
```

APPENDIX 3

TEST TASK

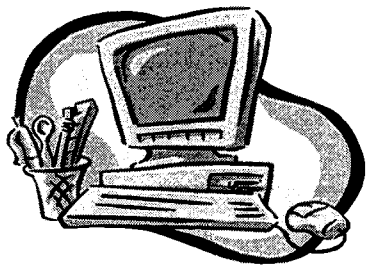
In the following task you will extract data from the Internet and convert it to XML by using X-Fetch Wrapper Rule Generator Program.

An HTML-page has already been selected (www.backtrackrecords.com) and the program is ready for use.

See the manual for instructions on performing the following subtasks:

1. Change the program from Browse mode to Edit mode.
2. Select data area (select five rows of your choice for XML conversion from the records list).
3. Select entry area.
4. Select and name the fields appropriately (e.g. Artist) (all the fields inside the entry area).
5. Generate the rules for data extraction (=analyse the data you have selected).
6. Extract the chosen data.
7. Preview the XML output in Internet Explorer.
8. View the conversion rules that were created.

Thank you!



APPENDIX 4

QUESTIONNAIRE

Please answer the following questions. You can answer them in Finnish.

1. How difficult was the task (in general) in your opinion? Circle your choice (1-4).

   easy 1 2 3 4 difficult

   Reasons for your answer:

2. Did you have problems in performing the subtasks?

   What kind of problems? What was the cause for the problems in your opinion? (E.g. name the subtasks that were difficult and explain why they were difficult.)

3. Were the instructions in the manual easy to understand, i.e. was it immediately clear for you what you had to do at each point? Give reasons for your answer.

4. How easy/difficult to understand was the language in the manual? Give reasons for your answer.

5. How would you grade the manual (1-5)? ___

6. Where do you think the manual failed to give the necessary information?

7. Other comments on the task/on the manual.

8. Do you think the brief 'training' before the test was necessary, i.e. was it useful to explain terms like HTML and XML beforehand, or where they already familiar to you?

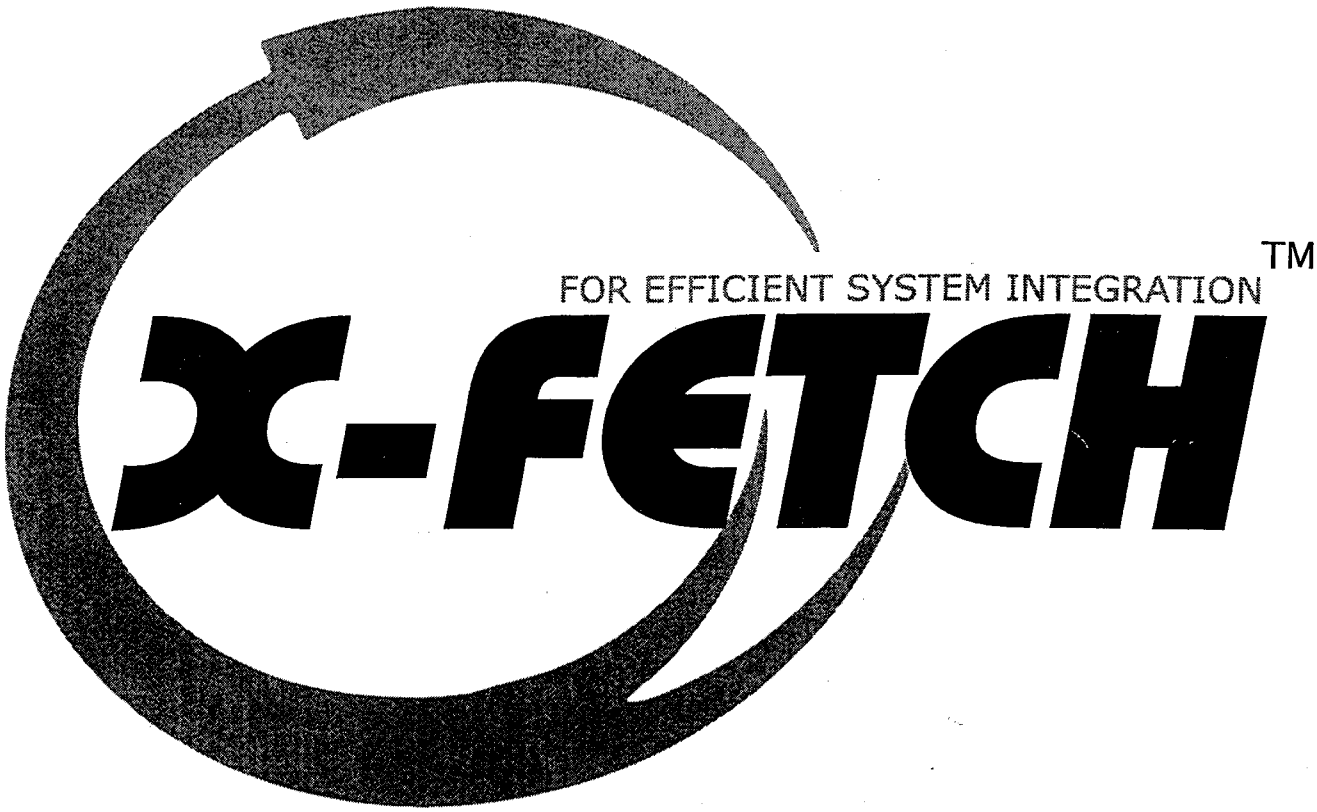
9. How would you assess yourself as a computer user? Circle your choice (1-4).

   novice 1 2 3 4 expert

10. Do you use a computer regularly? What do you use it for?

APPENDIX 5

Manual version A



Rule Generator Description

Wrapper 1.4

TABLE OF CONTENTS

1 INTRODUCTION

Manual summary

This manual describes the functions of X-Fetch Wrapper Rule Generator, version 1.4.

Definition

X-Fetch Wrapper Rule Generator enables you to select the data from actual source data for XML conversion. You can also modify the rules according to which the data is converted into XML.

Republica Corporation has a patent pending for Rule Generator as a unique solution for creating extraction rules from just marking up the desired fragments of information in the source data.

Operating principle

Using Rule Generator, data is selected from source page and marked up. Then DEL extraction rules are modified. Finally the data is extracted and converted to desired XML.

For example, you can pick pieces of information on a record sales site and mark them up as separate fields in XML. Then, with the help of X-Fetch AgentServer™, you can set the system to send the information to your WAP-enabled cell phone, or with X-Fetch™ Unifier, you can control the quality of the content inside XML elements.

Road map

To learn about the basics of Rule Generator:

- Chapter 2 lists the running requirements for Rule Generator.
- Chapter 3 describes the Editor's version compatibility with other X-Fetch Wrapper components.

To start working with Rule Generator:

- To get familiar with the Editor's user interface, move to chapter 4.
- To start up the Editor, start from chapter 5.
- To start extracting data from 'live' sources with Rule Generator, move to chapter 6.

The final chapter gives the Authors' addresses for sending your ideas and bug reports.

2 RUNNING REQUIREMENTS

Requirements

- Windows NT Server 4.0 (with Service Pack 5 or later) or Windows NT Workstation (with Service Pack 5 or later).
- Windows Internet Explorer version 5.01 and Service Pack 1 or later.
- Sun Java Runtime Environment version 1.2.2.

NOTE: X-Fetch Wrapper may work also with Windows 95/98 and Windows 2000. However, this has not been adequately tested.

3 VERSION COMPATIBILITY

Compatibilities

This version of X-Fetch Wrapper Rule Generator (1.4) supports the following versions of X-Fetch Wrapper Engine, DEL and DEL Editor:

COMPONENT	VERSION
Engine	1.4
Data Extraction Language	1.4
Data Extraction Language Editor	1.4

4 EDITOR APPEARANCE

Chapter summary X-Fetch Wrapper Rule Generator's user interface is presented in this chapter.

4.1 Bars

Bars The user interface contains the following bars:

- **Title bar** The title bar shows the name of the page.
- **Scroll bars** There are scroll bars in each frame.
- **Status bar** The status bar at the bottom has the following five panels:

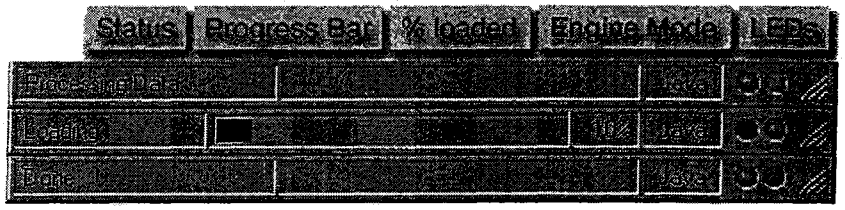


Figure 1. Status bar panels. The figure shows the three possible statuses in loading ("Processing Data...", "Loading..." and "Done").

1) Status panel Status panel shows the state of page loading and processing as follows:

MESSAGE	STATE
"Loading..." (Browse mode)	Page loading in progress.
"Processing Data..." (Edit mode)	Page processing or data analysis in progress.
"Done"	Page loading complete.

2) Progress Bar Shows the progress of page loading graphically.

3) % loaded Shows the status of page loading in percentage.

4) Engine Mode Shows the Omni/Java–engine used (dual engine version only).

5) LEDs The colors of the lights signal the status of loading/processing as well:

COLOR	MEANING
Green	Ready state
Red	Busy state

Yellow	Background process running (data analysis/extraction)

4.2 Frames

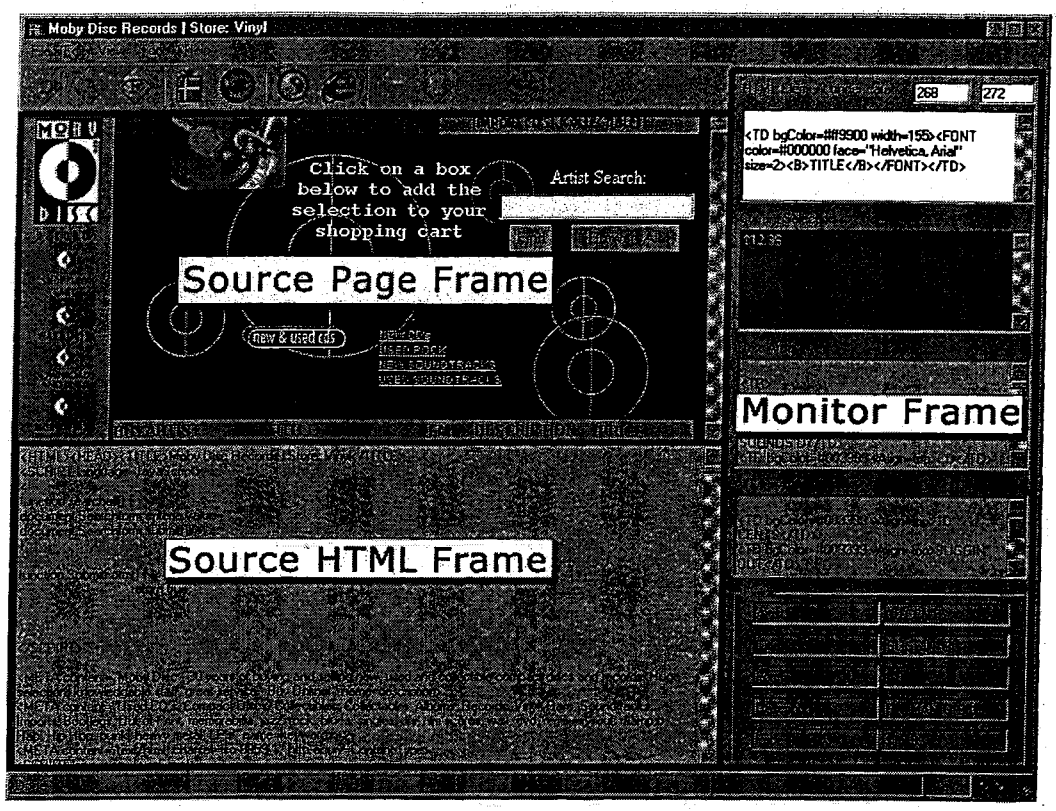


Figure 2. Frames in Edit mode. Source Page Frame on the top left, Source HTML Frame on the bottom left and Monitor Frame on the right.

Three frames There are three frames in Rule Generator:

- **Source Page Frame (top left)** This frame shows the HTML page.

  In Browse mode, it is the original HTML page. In Edit mode, it is the editing HTML page where data extraction is done.

- **Monitor Frame (right)** This frame allows you to monitor the following extraction information:

- **Current HTML element under caret**: Shows current vertical and horizontal coordinates of the HTML element under cursor (white background).
- **Currently selected HTML code** (dark green): Shows current selection in HTML code.
- **Selected Data Area** (green): Shows defined data area (if any).
- **Selected All Tags area** (yellow): Shows defined all tags area (if any).
- **Defined Fields** (lilac): Shows defined fields (if any).

NOTE: This frame is visible only in Edit mode when ticking View – Info.

- **Source HTML Frame (bottom left)**

This frame shows the HTML source of the currently loaded page.

NOTE: This frame is visible only in Edit mode when ticking View – Source HTML.

4.3 Menu commands (Edit and Browse modes)

Summary

The following menu commands are available in the menus.

NOTE: In case the menu command is only available in either Edit mode or Browse mode, it has been mentioned ("Edit mode only" or "Browse mode only").

4.3.1. File menu

- **Open** (Browse mode only)

Function Opens an HTML page.

Tip! Use to start a previously saved extraction project or for opening an HTML page.

- **Save as**

Function Saves the HTML page (both Browse and Edit modes).

Tip! Use during area selection (e.g. after choosing Data Area). This may be sensible because there is no undo. You can then load the chosen Data Area and try again.

- **Exit**

Function Closes Rule Generator.

4.3.2. View menu

- **Toolbar**

Function Tick to show/hide the toolbar in Rule Generator.

- **Status bar**

Function Tick to show/hide the status bar.

- **Refresh**

Function Refreshes the current HTML page.

- **Source HTML** (Edit mode only!)

Function Tick to open Source HTML Frame at the bottom of the window.

It is possible to edit source HTML code. To update the displayed page according to changes you made in Source HTML Frame, use **Update HTML** button.

For Source HTML Frame, see 4.2.

CAUTION! If you make any changes to the source HTML code, the data analysis becomes "out of date" and has to be redone! To reanalyze data, press Update HTML toolbar button and then the data selection indicators in the following sequence:

CAUTION! You cannot change source HTML code while doing data selection. Therefore, first finish with the selection mark-up and only afterwards make changes to the source HTML code.

- **Info** (Edit mode only!)

Function Tick to show/hide the Monitor Frame.

For description of this frame, see 4.2.

4.3.3. Tools menu

- **Use Current Rule**

Function Allows to use previously created rule set for data extraction.

Tip! Use with similar pages to avoid repeating the identical mark-up/analysis.

NOTE: **The rule set file must exist!**

4.3.4. Help menu

- **Contents** (no online help available yet)
- **Search for help on** (no online help available yet)
- **About**

Function

Shows Rule Generator version number, warning concerning misuse of source page content, and a legal disclaimer by Republica Corporation.

4.4 Browse mode toolbar buttons

Starting Browse mode

When you open Rule Generator, it is in Browse mode.

When switched to Edit mode, the Browse mode is activated by pushing **Browse** toolbar button.

NOTE: **In Browse mode, the Editor is a regular web browser (see 4.4). The actual extraction procedure is done in Edit mode (see chapter 4.5 for description of the Edit mode functions and chapter 6 for the actual extraction procedure).**

NOTE: **In Edit mode, the Rule Generator can still be used as a web browser. However, clicking a hyperlink will result in opening a new Internet Explorer browser window.**

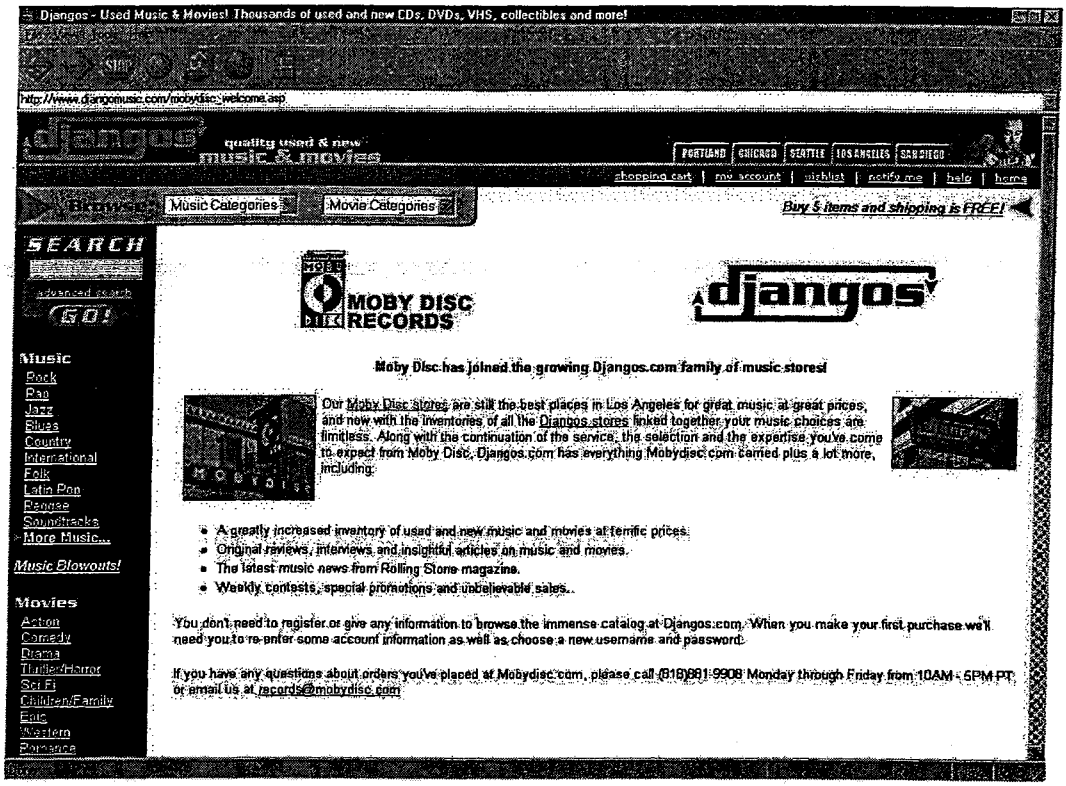


Figure 3. Browse mode. Shows Django's record sales site at http://www.djangomusic.com/mobydisc_welcome.asp.

Toolbar buttons This section lists the functions of the Browse mode toolbar buttons:

- **Back**

Function Loads the last visited web page.

- **Forward**

Function Loads a previously visited page (only available after pressing **Back**).

- **Stop**

Function Stops loading.

- **Refresh**

Function Reloads the current HTML page from cache.

- **Home**

Function Loads the page marked as default homepage in Windows NT **Internet Options.**

- **Search**

Function Loads the default search site.

NOTE: This button has the same basic function as standard Internet Explorer Search button, but it cannot be customized.

- **Edit mode**

Function Switches to Edit mode (see chapter 4.5).

4.5 Edit mode toolbar buttons

Chapter summary This chapter describes the functions of the Edit mode toolbar buttons.

The actual data selection, analysis and extraction are done in Edit mode.

Starting Edit mode Activate Edit mode (when in Browse mode) by pressing **Edit mode** toolbar button.

Toolbar buttons This section lists the functions of the Edit mode toolbar buttons:

- **D(ata Area)**

Function Starts a data analysis on Data Area.

For using this button in the extraction, see chapter 6.4.

NOTE: Data selection indicators ( ,  ,  buttons) are used to start data analysis. The appearance of the buttons also indicates the status of analysis as follows:

WHEN THE BUTTON IS...	...IT MEANS THAT THE DATA IS...
Dimmed	not yet defined
Highlighted and active	defined, but not yet analyzed
Embossed and grey	defined and analyzed.

- **A(ll Tags Area) (later referred to as Entry Area)**

Function Starts a data analysis on the Entry Area.

For using this button in the extraction, see chapter 6.4.

- **F(ields)**

Function Starts a data analysis on Fields.

For using this button in the extraction, see chapter 6.4.

- **Extract**

Function Extracts the selected data (Data Area, Entry Area and Fields) and makes an XML output.

For using this button in the extraction, see chapter 6.5.

- **Preview**

Function Shows the data selection XML in Internet Explorer browser.

For using this button in the extraction, see chapter 6.6.

- **Reload**

Function Reloads the current HTML page.

Tip! Use this button to restart the mark-up process.

CAUTION! Be aware that Reload discards all current data selections!

- **Update HTML**

Function Loads source HTML code from Source HTML Frame (code) to Source Page Frame (browser).

For details on modifying the source HTML code, see 4.2.

NOTE: Update HTML button becomes active only when changes have been made to source HTML code.

- **View Rule Set**

Function Displays the result of data analysis in XML.

For using this button, see chapter 6.7.

- **Browse**

Function

Switches to Browse mode.

For Browse mode functions, see 4.4.

5 STARTING AND EXITING RULE GENERATOR

Chapter summary This chapter tells you how to start up and exit Rule Generator after installation.

5.1 Starting

Instruction To start Rule Generator, click the application icon in your Windows **Start** – **Programs** menu.

5.2 Exiting

Instruction To exit Rule Generator, either choose **File** – **Exit** or press the  button.

Next To start working with Rule Generator, move on to chapter 6 EDITOR FUNCTIONS.

6 EDITOR FUNCTIONS

Chapter summary This chapter gives you the instructions how to perform data extraction from source material using Rule Generator.

Road map The order of presentation is chronological, starting from data selection (6.1 - 6.3) and proceeding to data analysis (6.4) and extraction (6.5). Previewing the XML output is described in chapter 6.6 and modifying the extraction rules in chapter 6.7.

6.1 Selecting Data Area (Add Data Area)

Chapter summary This chapter describes how to select Data Area.

Definition A Data Area contains the whole data selection area: all Entry Areas and all Fields.

Instructions

1. In Browse mode, select the HTML page from which you want to extract data.
2. Activate Edit mode (if in Browse mode) by pressing **Edit mode** toolbar button.
3. First, using left mouse button, select the area, which contains all the data to be extracted.
4. Then, click right mouse button and choose **Add Data Area**.

The Data Area has now been automatically analyzed.

tags appear around the selected area.

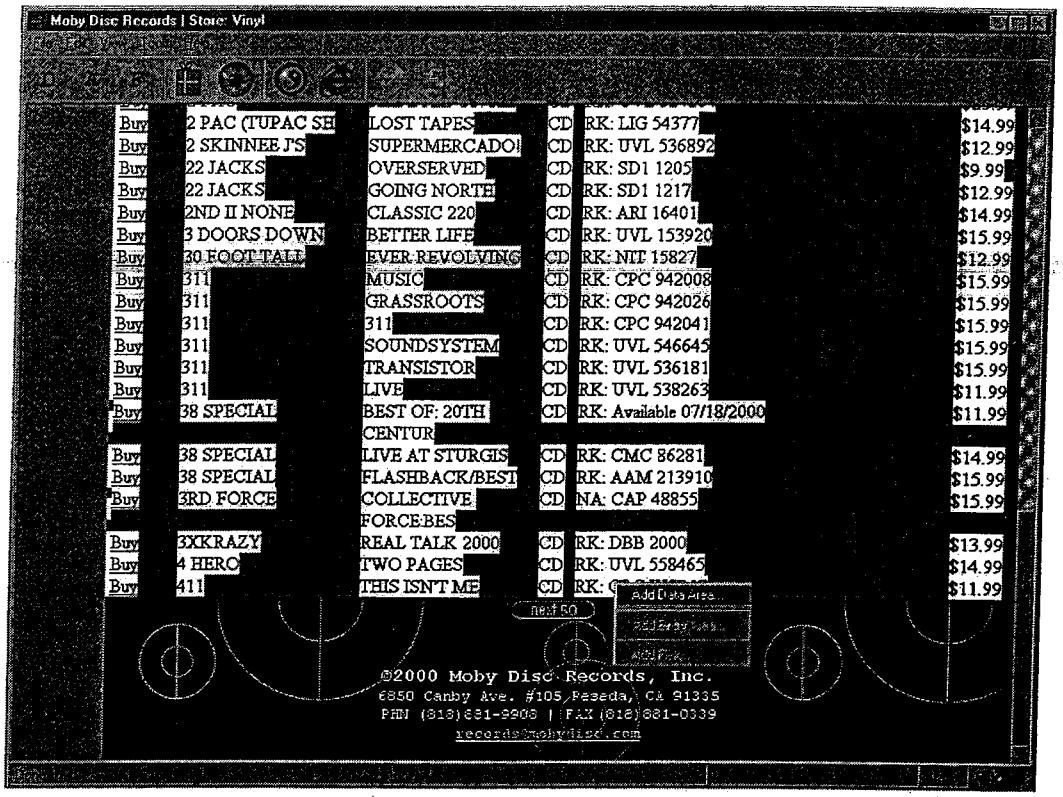


Figure 4. Adding Data Area. Select the Data Area containing all information of listed albums (shows in white). Right-click and choose Add Data Area to confirm the selection.

NOTE: No undo function is available yet. To change the Data Area definition, you have to reload the page from cache and re-create Data Area.

NOTE: Though you can select multiple Data Areas, X-Fetch Wrapper Engine does not process additional areas yet. Proceed with one data area.

Next	Proceed with selecting Entry Area.

6.2 Selecting Entry Area (Add Entry Area)

Chapter summary	This chapter describes how to select Entry Area.
Definition	An Entry Area Contains all Fields once. It is also known as All Tags Area.
Instructions	1. First, using left mouse button, select the area which contains one complete entry. In an Entry Area, all the prospective fields occur only once.

NOTE: You should not choose the first row in Data Area as Entry Area. XML output may not come out right. Rather choose the second or third row.

NOTE: **Obviously, the Entry Area must reside within the Data Area.**

2. Then, right-click and choose **Add Entry Area**.

The Entry Area has now been defined and automatically analyzed.

tags appear around the selection.

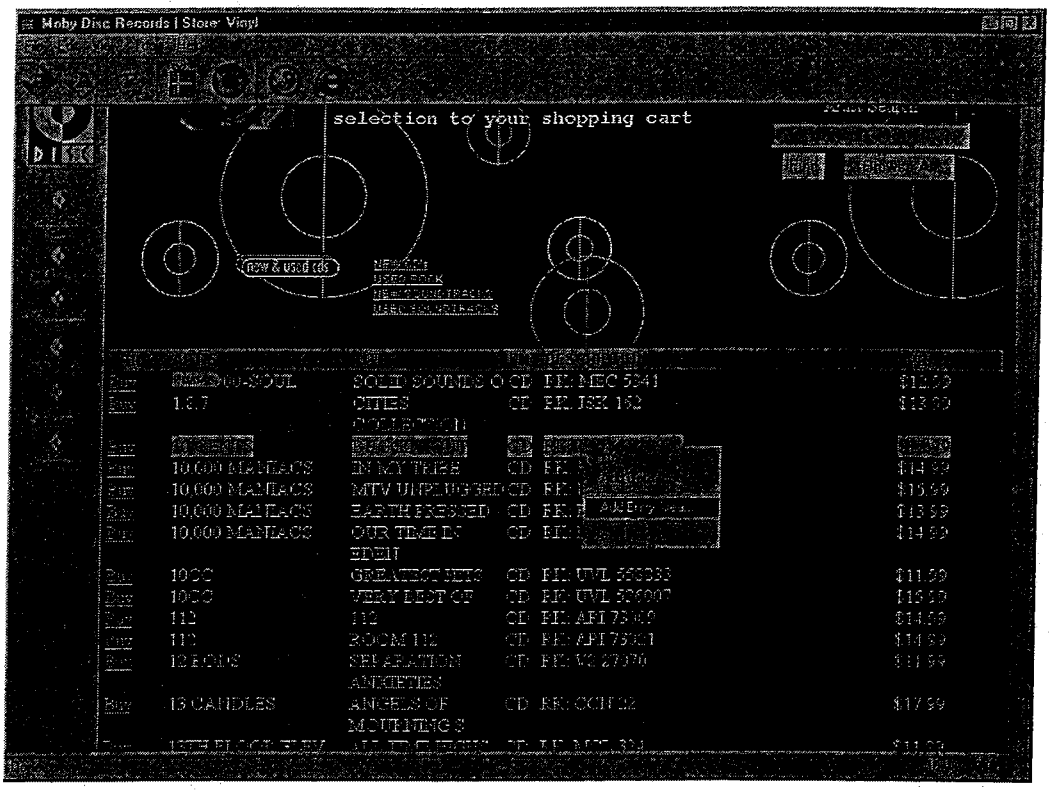


Figure 5. Adding Entry Area. Select an entire row containing the artist, album title, FMT (format), description as well as price using mouse (the area shows in white). Then right-click and choose Add Entry Area.

3. Check that the selection is how you intended. If not, redo the entire selection by pressing (**Reload**) button.

NOTE: **No undo function is available yet. To change the Entry Area definition, you have to reload the page from cache (using Reload button) and re-create Entry Area.**

NOTE: **Though you can select multiple Entry Areas, the Engine only processes the first one. Therefore, proceed with one Entry Area only.**

Next Proceed with selecting individual Fields.

6.3 Selecting Fields (Add Field)

Chapter summary This chapter describes how to select Fields.

Definition A Field is a fragment of actual data selected for extraction. In XML, it will be placed to an element of its own.

Instructions

1. Select the Field area using left mouse button.
2. Right-click and choose **Add Field.**

NOTE: Naturally, the field must reside within the Entry Area.

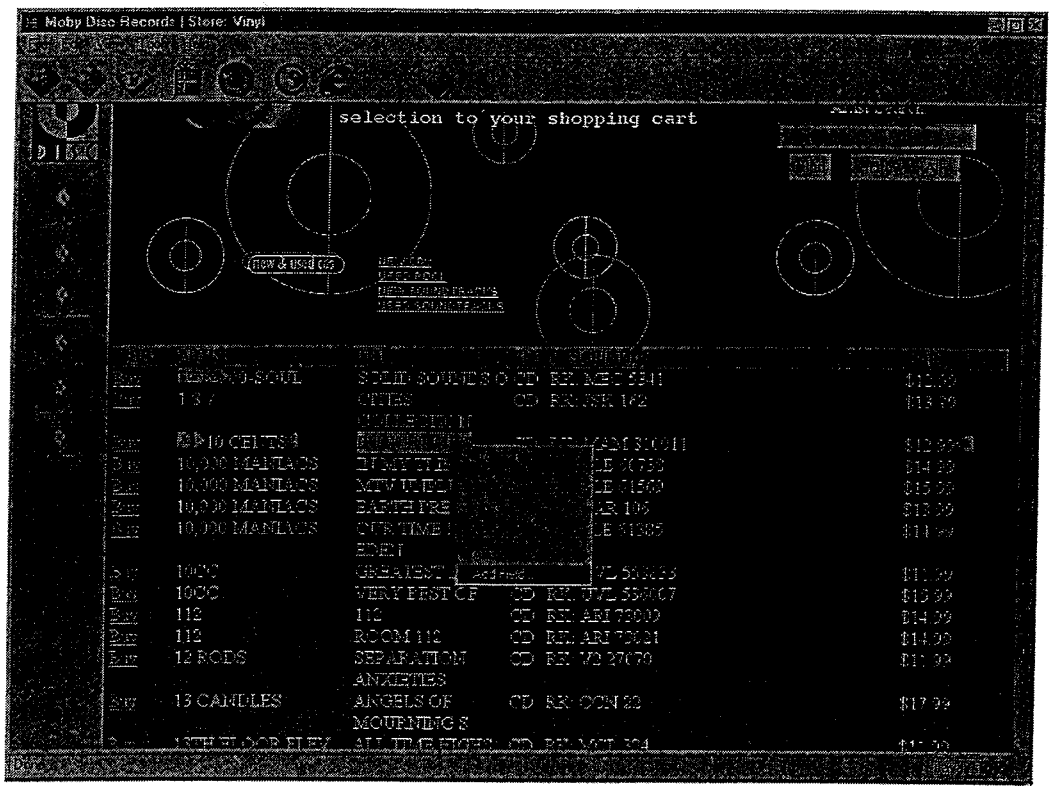


Figure 6. Selecting a Field. Select a Field ("Title": "Buggin' Out"). Then right-click and choose Add Field. Note that one field ("Artist") has been added previously (shows between ▶◀ field tags).

3. **Add new field** window appears. Assign a descriptive name to **Field Name** (e.g. "Artist", "Title") or use the default name (e.g. "Field1", "Field2" and so on).

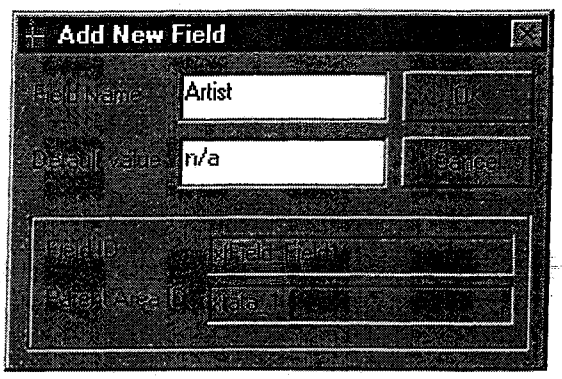


Figure 7. Defining the field. A Field with the name "Artist" is created.

NOTE: The Default value is not in use ("n/a"). The other parameters (Field ID and Parent Area ID) contain debug information and cannot be changed.

4. Click **OK** to set the selection.

   The selected Field is then added to the data selection. Field tags ⋈ now mark the field selection.

   OR Click **Cancel** to abort the field selection.

NOTE: Check that there are only two arrows per one HTML table cell. If there is another start-tag, that might not produce the correct XML. You should probably do the whole selection all over again.

NOTE: Fields are allowed to overlap, but then the tags of the first one will be included as tags in the second field.

5. Proceed by selecting another field. In most extractions, more than one field is needed.

NOTE: There is no undo function yet. Defined field(s) cannot be deleted. The only way to redefine field(s) is to restart the mark-up process (using Reload button).

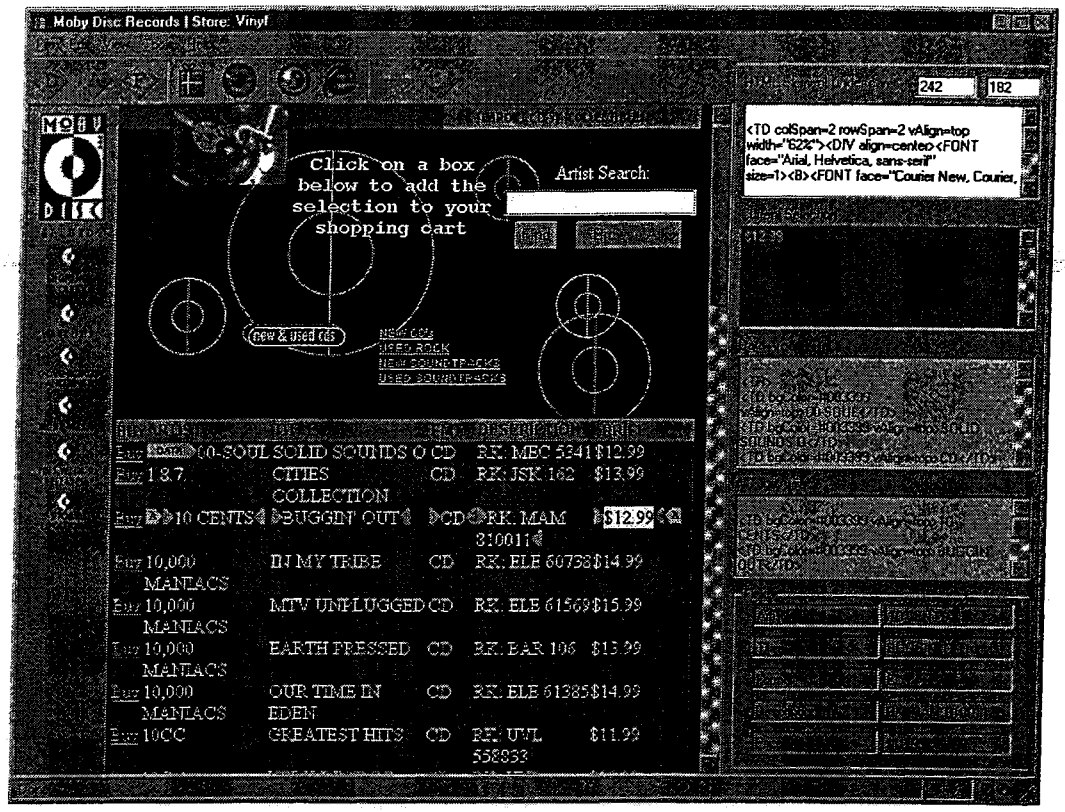


Figure 8. Once the data selection is completed, the screen should look more or less like this: tags marking the data selection.

Next Once you have selected the fields, proceed with the rule generation.

6.4 Generating the rules (F button)

Chapter summary This chapter describes how to generate the extraction rules for the data selection done above.

Instructions

1. To start data analysis on the selection, press the  **(Fields)** toolbar button. It becomes active once at least one field is defined.

NOTE: The other data selection buttons  **and**  **are not active at this stage. They are activated only in case the source HTML is modified and the data analysis needs to be done again. For general description of the data selection indicators D, A and F, see 4.5.**

2. You can check the generated extraction rules using  **(View Rule Set)** button. It becomes available when the analysis is complete.

NOTE: **Use DEL Editor to modify the extraction rules.**

Next When you want to try how the generated extraction rules work on the source data, proceed to extracting the data (see 6.5).

6.5 Extracting the data and previewing the XML output (Extract)

Chapter summary This chapter describes how to extract the data selection analyzed above.

Instructions

1. Press  (**Extract**) button (should be available after analysis is completed).

Rule Generator extracts the selected data (Data Area, Entry Area and Fields) and makes an XML output based on the generated rules.

2. XML output is saved and shown by default in text mode (see figure below).

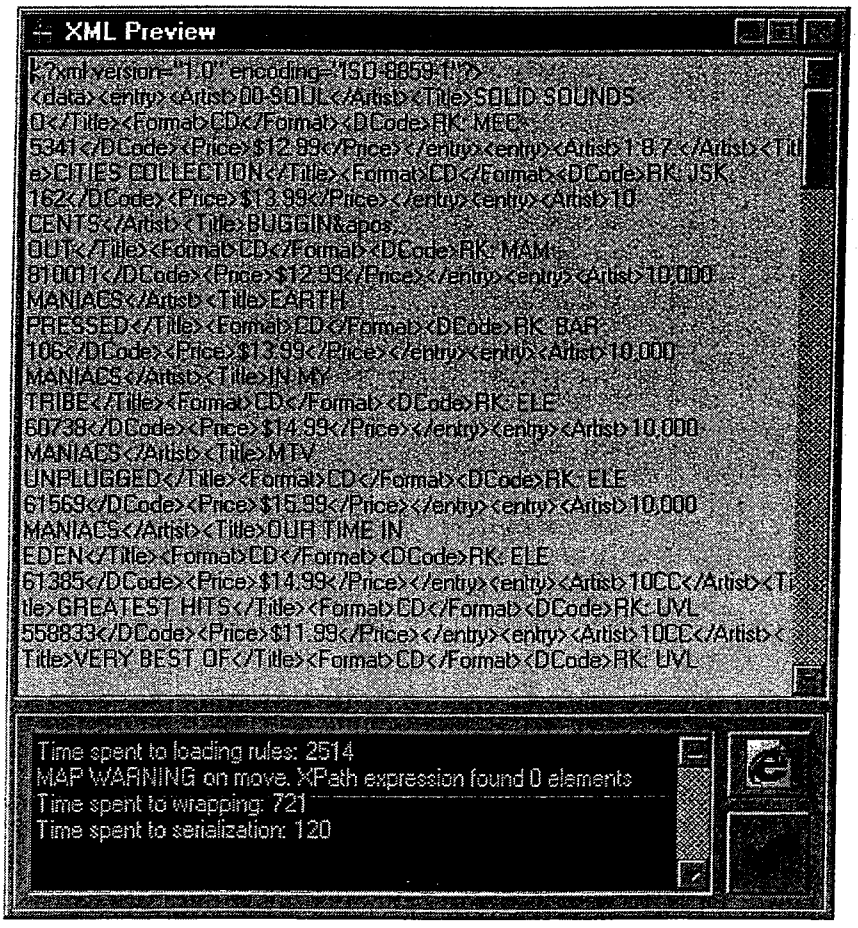


Figure 9. Text format view of your XML output.

In XML, <data> tags mark the Data Area (all information of all albums), <entry> tags close in the entries (individual albums) and

field name elements mark the album information (e.g. <Price>).

There is also a log window (at the bottom) that shows the time spent by the Engine to perform the various XML conversion stages.

NOTE: The XML output is saved automatically by the program. There is no need to save it explicitly.

3. To view the output in Internet Explorer (more visual), press  button.

OR To close the preview window, use the  button.

6.6 Previewing XML output in Internet Explorer (Preview)

Chapter summary

You can preview the XML output in Internet Explorer even after you have closed the XML preview window.

NOTE:  button does exactly the same as  button (shows XML output in Internet Explorer), but with this button, you can view it AFTER you have closed the preview window after extraction.

Instructions

1. To view the XML output from the current selection in Internet Explorer, click  **(Preview)** toolbar button (see Figure 10 below).

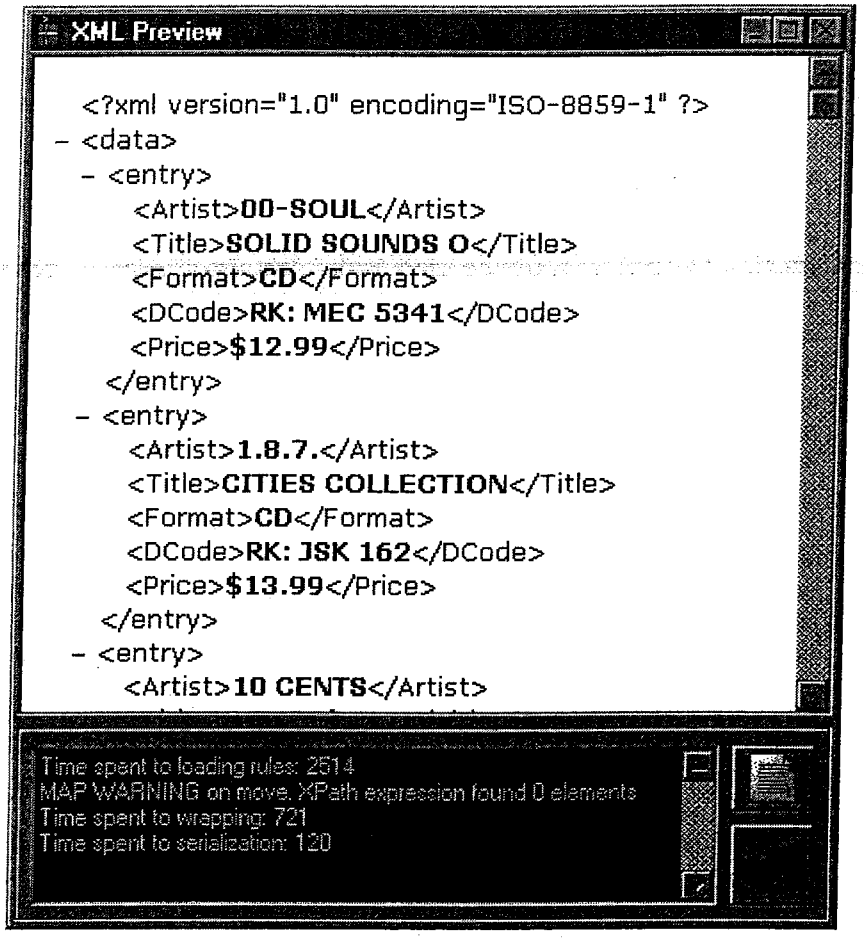


Figure 10. Internet Explorer mode shows the elements more conveniently in the XML hierarchy. Here you can also expand (+) and contract (-) the elements.

Again, the log window (at the bottom) shows the times spent by the Engine to perform the various XML conversion stages.

Otherwise it is the same as the text format view (see Figure 9 above).

2. To close the preview window, press  button.

OR To view the XML output in text format, press  button.

NOTE: Again, there is no need to explicitly save the XML file. It has already been saved after the extraction.

Next To make another extraction, return to 6.1.

To view the XML conversion rules created during extraction, go to 6.7.

6.7 Viewing the conversion rules (View Rule Set)

Chapter summary You can view Data Extraction Language rules created after you extracted the data.

Instructions

1. To view the DEL conversion rules, click  (**View Rule Set**) button.

2. To make changes to the conversion rules, use DEL Editor.

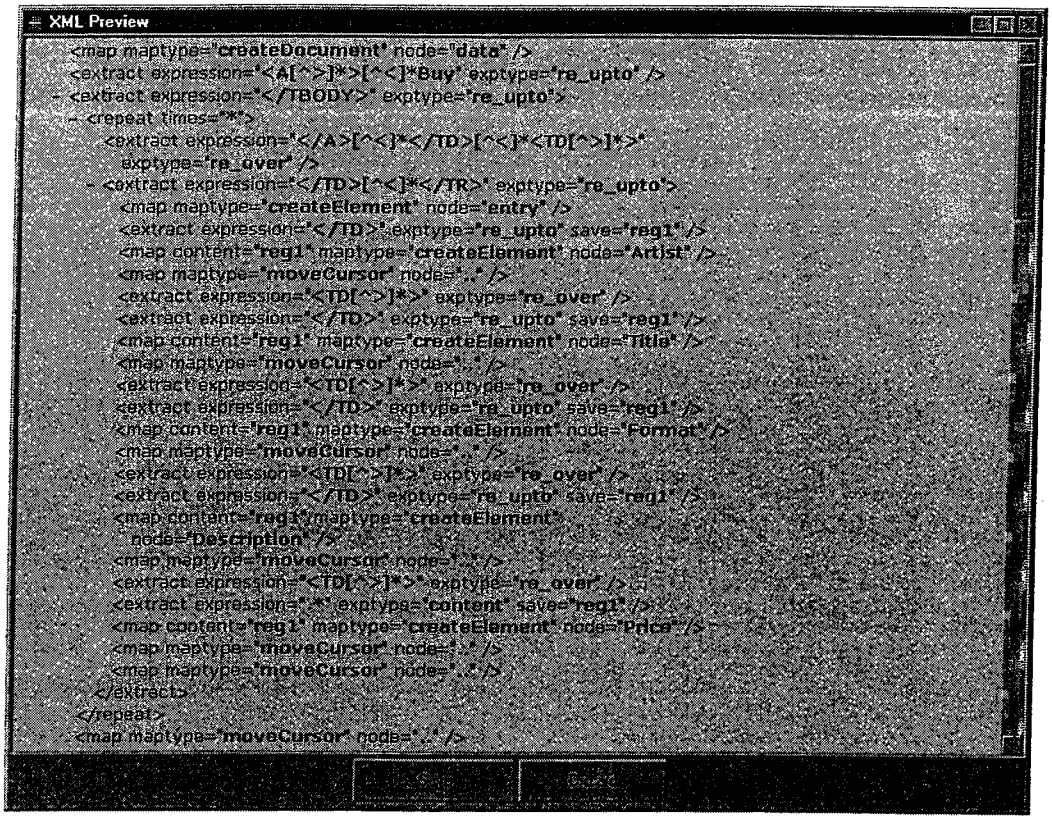


Figure 11. Previewing the conversion rules. The figure shows the rules created by Rule Generator based on the given areas and fields.

In case the extraction was not as you intended, use DEL Editor to modify the extraction rules or redo the extraction.

7 VERSION HISTORY

Chapter summary The following is the version history of X-Fetch Wrapper Rule Generator.

7.1 Version 0.8 Beta

Changes

- First official Beta release. Last known build is Build 0820 (01-26-2000).
- Used proprietary HTML parser (not fully compatible with Netscape Navigator or Internet Explorer).

7.2 Version 1.0

Changes

- Internet Explorer HTML parser.

7.3 Version 1.1

Changes

- Well-formed XML output.

7.4 Version 1.2

Changes

- Multiple data/entry areas support.
- New Java-based engine (version 1.3) (08-31-2000).

7.5 Version 1.3

Changes

- Dual engine support.
- New version 1.4 engine.

7.6 Version 1.3.22 (beta)

Changes

- Possibility to extract data using an already created rule file has been added.
- **Preview** function in the Browser (Internet Explorer) has been changed.

8 CONTACT INFORMATION

Contact information

All bug reports and suggestions for development should be sent to the following addresses:

Via e-mail:	Wrapper@x-fetch.com
Via mail:	Republica / Wrapper
	Ahlmaninkatu 1
	40100 Jyväskylä
	Finland

APPENDIX 6

Manual version B

X-FETCH WRAPPER:

RULE GENERATOR EDITOR

TABLE OF CONTENTS

1 INTRODUCTION

X-Fetch Wrapper Rule Generator enables a data selection from actual source data for XML conversion. The rules according to which the data is converted into XML can also be modified. Republica Corporation has a patent pending for Rule Generator as a unique solution for creating extraction rules from just marking up the desired fragments of information in the source data. Using Rule Generator, data is selected from source page and marked up. Then DEL extraction rules are modified, and finally the data is extracted and converted to desired XML. For example, pieces of information can be picked on a record sales site and marked up as separate fields in XML. Then, with the help of X-Fetch AgentServer™, the system can be set to send the information to a WAP-enabled cell phone, or with X-Fetch™ Unifier, the quality of the content inside XML elements can be controlled.

2 RUNNING REQUIREMENTS

- Windows NT Server 4.0 (with Service Pack 5 or later) or Windows NT Workstation (with Service Pack 5 or later).
- Windows Internet Explorer version 5.01 and Service Pack 1 or later.
- Sun Java Runtime Environment version 1.2.2.

Note: X-Fetch Wrapper may work also with Windows 95/98 and Windows 2000. However, this has not been adequately tested.

3 VERSION COMPATIBILITY

This version of X-Fetch Wrapper Rule Generator (1.4) supports the following versions of X-Fetch Wrapper Engine, DEL and DEL Editor:

COMPONENT	VERSION
Engine	1.4
Data Extraction Language	1.4
Data Extraction Language Editor	1.4

4 EDITOR APPEARANCE

4.1 Bars

The user interface contains the following bars:

The name of the page is shown on the title bar. There are scroll bars in each frame. The status bar at the bottom has the following five panels:

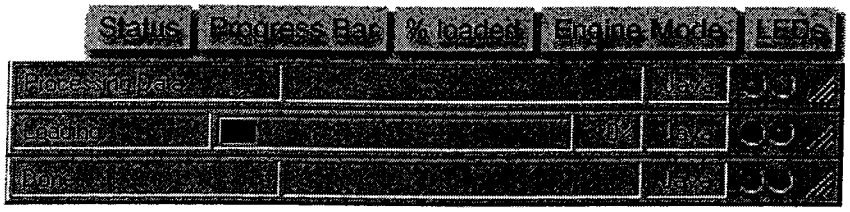


Figure 1. Status bar panels.

Status panel shows the state of page loading and processing as follows:

MESSAGE	STATE
"Loading..." (Browse mode)	Page loading in progress.
"Processing Data..." (Edit mode)	Page processing or data analysis in progress.
"Done"	Page loading complete.

Progress bar shows the progress of page loading graphically.

% loaded shows the status of page loading in percentage.

Engine mode shows the Omni/Java–engine used (dual engine version only).

LEDs: the colors of the lights signal the status of loading/processing as well:

COLOR	MEANING
Green	Ready state
Red	Busy state
Yellow	Background process running (data analysis/extraction)

4.2 Frames

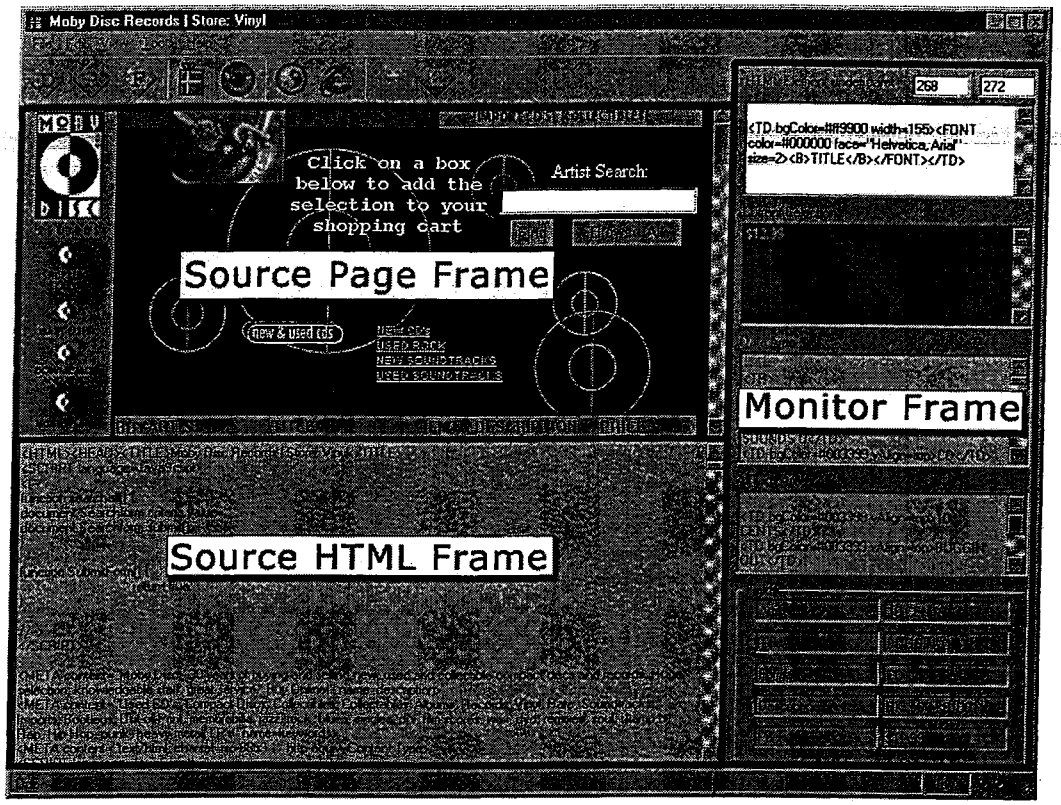


Figure 2. Frames in Edit mode.

There are three frames in Rule Generator:

Source page frame (top left) shows the HTML page. In Browse mode, it is the original HTML page. In Edit mode, it is the editing HTML page where data extraction is done.

Monitor frame (right) allows you to monitor the following extraction information:

- **Current HTML element under caret**: Shows current vertical and horizontal coordinates of the HTML element under cursor (white background).
- **Currently selected HTML code** (dark green): Shows current selection in HTML code.
- **Selected Data Area** (green): Shows defined data area (if any).
- **Selected All Tags area** (yellow): Shows defined all tags area (if any).
- **Defined Fields** (lilac): Shows defined fields (if any).

Note: This frame is visible only in Edit mode when ticking View – Info.

Source HTML frame (bottom left) shows the HTML source of the currently loaded page.

page.

Note: This frame is visible only in Edit mode when ticking View – Source HTML.

4.3 Menu commands (Edit and Browse modes)

Note: In case the menu command is only available in either Edit mode or Browse mode, it has been mentioned ("Edit mode only" or "Browse mode only").

4.3.1. File menu

- **Open** (Browse mode only)

An HTML page is opened.

Tip: use to start a previously saved extraction project or for opening an HTML page.

- **Save as**

The HTML page (both Browse and Edit modes) is saved.

Tip: use during area selection (e.g. after choosing Data Area). This may be sensible because there is no undo. The chosen Data Area can then be loaded and another attempt can be done.

- **Exit**

Rule Generator is closed.

4.3.2. View menu

- **Toolbar**

Tick to show/hide the toolbar in Rule Generator.

- **Status bar**

Tick to show/hide the status bar.

- **Refresh**

The current HTML page is refreshed.

- **Source HTML** (Edit mode only!)

Tick to open Source HTML Frame at the bottom of the window. It is possible to edit source HTML code. To update the displayed page according to changes that have been made in Source HTML Frame, **Update HTML** button is used.

For Source HTML Frame, see 4.2.

Caution: If any changes to the source HTML code are made, the data analysis becomes "out of date" and has to be redone! To reanalyze data, press Update HTML toolbar button and then the data selection indicators in the following sequence:

Caution: Source HTML code cannot be changed while data selection is being done. First finish with the selection mark-up and only afterwards make changes to the source HTML code.

- **Info** (Edit mode only!)

Tick to show/hide the Monitor Frame.

For description of this frame, see 4.2.

4.3.3. Tools menu

- **Use Current Rule**

Allows to use previously created rule set for data extraction.

Tip: use with similar pages to avoid repeating the identical mark-up/analysis.

Note: The rule set file must exist!

4.3.4. Help menu

- **Contents** (no online help available yet)
- **Search for help on** (no online help available yet)
- **About**

Shows Rule Generator version number, warning concerning misuse of source page content, and a legal disclaimer by Republica Corporation.

4.4 Browse mode toolbar buttons

When you open Rule Generator, it is in Browse mode. When switched to Edit mode, the Browse mode is activated by pushing **Browse** toolbar button.

Note: In Browse mode, the Editor is a regular web browser (see 4.4). The actual extraction procedure is done in Edit mode (see chapter 4.5 for description of the Edit mode functions and chapter 6 for the actual extraction procedure).

Note: In Edit mode, the Rule Generator can still be used as a web browser. However, clicking a hyperlink will result in opening a new Internet Explorer browser window.

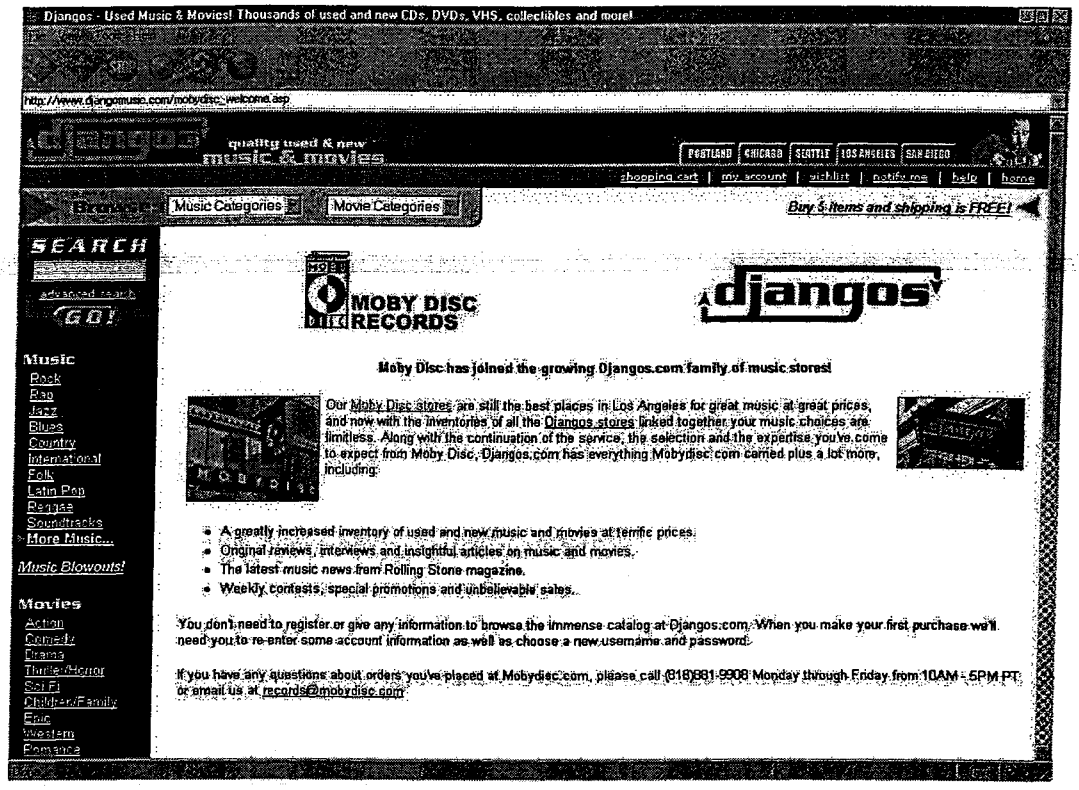


Figure 3. Browse mode. Shows Django's record sales site at http://www.djangomusic.com/mobydisc_welcome.asp.

This section lists the functions of the Browse mode toolbar buttons:

- **Back**

The last visited web page is loaded.

- **Forward**

A previously visited page (only available after pressing **Back**) is loaded.

- **Stop**

Loading is stopped.

- **Refresh**

The current HTML page from cache is reloaded.

- **Home**

The page marked as default homepage in Windows NT **Internet Options** is loaded.

- **Search**

The default search site is loaded.

Note: This button has the same basic function as standard Internet Explorer Search button, but it cannot be customized.

- **Edit mode**

A switch to Edit mode is performed (see chapter 4.5).

4.5 Edit mode toolbar buttons

The actual data selection, analysis and extraction are done in Edit mode.

Edit mode is activated (when in Browse mode) by pressing **Edit mode** toolbar button.

This section lists the functions of the Edit mode toolbar buttons:

- **D(ata Area)**

A data analysis on Data Area is started.

For using this button in the extraction, see chapter 6.4.

Note: Data selection indicators ( ,  ,  buttons) are used to start data analysis. The appearance of the buttons also indicates the status of analysis as follows:

When the button is dimmed, it means that the data is not yet defined.

When the button is highlighted and active, it means that the data is defined, but not yet analysed.

When the button is embossed and grey, it means that the data is defined and analysed.

- **A(ll Tags Area) (later referred to as Entry Area)**

A data analysis on the Entry Area is started.

For using this button in the extraction, see chapter 6.4.

- **F(ields)**

A data analysis on Fields is started.

For using this button in the extraction, see chapter 6.4.

- **Extract**

The selected data (Data Area, Entry Area and Fields) is extracted and an XML output is made.

For using this button in the extraction, see chapter 6.5.

- **Preview**

The data selection XML in Internet Explorer browser is shown.

For using this button in the extraction, see chapter 6.6.

- **Reload**

The current HTML page is reloaded.

Tip: use this button to restart the mark-up process.

Caution: Be aware that all current data selections are discarded by pressing Reload!

- **Update HTML**

Loads source HTML code from Source HTML Frame (code) to Source Page Frame (browser).

For details on modifying the source HTML code, see 4.2.

Note: Update HTML button becomes active only when changes have been made to source HTML code.

- **View Rule Set**

The result of data analysis in XML is displayed.

For using this button, see chapter 6.7.

- **Browse**

A switch to Browse mode is performed.

For Browse mode functions, see 4.4.

5 STARTING AND EXITING RULE GENERATOR

5.1 Starting

Rule Generator is opened by clicking the application icon in the Windows **Start – Programs** menu.

5.2 Exiting

Rule Generator is closed either by choosing **File – Exit** or by pressing the  button.

6 EDITOR FUNCTIONS

6.1 Selecting Data Area (Add Data Area)

A Data Area contains the whole data selection area: all Entry Areas and all Fields.

In Browse mode, a desired HTML page from which you want to extract data is selected. Edit mode is activated (if in Browse mode) by pressing **Edit mode** toolbar button. Using left mouse button, the area which contains all the data to be extracted is selected. Right mouse button is clicked and **Add Data Area** is chosen. The Data Area has now been automatically analyzed.

tags appear around the selected area.

Note: No undo function is available yet. To change the Data Area definition, the page from cache has to be reloaded and Data Area must be recreated.

Note: Though multiple Data Areas can be selected, X-Fetch Wrapper Engine does not process additional areas yet. Proceed with one data area.

6.2 Selecting Entry Area (Add Entry Area)

An Entry Area Contains all Fields once. It is also known as All Tags Area. The area which contains one complete entry is selected using left mouse button. In an Entry Area, all the prospective fields occur only once.

Note: The first row in Data Area should not be chosen as Entry Area. XML output may not come out right. It is recommended that the second or third row is chosen.

Note: Obviously, the Entry Area must reside within the Data Area.

Then, **Add Entry Area** is chosen by right-clicking. The Entry Area has now been defined and automatically analyzed.


tags appear around the selection.

Check that the selection is how you intended. If not, the entire selection can be redone by pressing  (**Reload**) button.

Note: No undo function is available yet. To change the Entry Area definition, the page must be reloaded from cache (using Reload button) and Entry Area must be recreated.

Note: Though you can select multiple Entry Areas, the Engine only processes the first one. Therefore, proceed with one Entry Area only.

one. Therefore, proceed with one Entry Area only.

6.3 Selecting Fields (Add Field)

A Field is a fragment of actual data selected for extraction. In XML, it will be placed to an element of its own. The Field area is selected using left mouse button, then **Add Field** is chosen by right-clicking.

Note: Naturally, the field must reside within the Entry Area.

Add new field window appears. A descriptive name is assigned to **Field Name** (e.g. "Artist", "Title") or the default name is used (e.g. "Field1", "Field2" and so on).

Note: The Default value is not in use ("n/a"). The other parameters (Field ID and Parent Area ID) contain debug information and cannot be changed.

The selection is set by clicking OK.

The selected Field is then added to the data selection. Field tags now mark the field selection.

OR Click **Cancel** to abort the field selection.

Note: Check that there are only two arrows per one HTML table cell. If there is another start-tag, that might not produce the correct XML and the whole selection should probably be done all over again.

Note: Fields are allowed to overlap, but then the tags of the first one will be included as tags in the second field.

Proceed by selecting another field. In most extractions, more than one field is needed.

Note: There is no undo function yet. Defined field(s) cannot be deleted. The only way to redefine field(s) is to restart the mark-up process (using Reload button).

6.4 Generating the rules (F button)

To start data analysis on the selection, the  **(Fields)** toolbar button is pressed. It becomes active once at least one field is defined.

Note: The other data selection buttons  and  are not active at this stage. They are activated only in case the source HTML is modified and the data analysis needs to be done again. For general description of the data selection indicators D, A and F, see 4.5.

The generated extraction rules can be checked by using  **(View Rule Set)** button. It becomes available when the analysis is complete.

Note: Use DEL Editor to modify the extraction rules.

6.5 Extracting the data and previewing the XML output (Extract)


(**Extract**) button is pressed (should be available after analysis is completed). Rule Generator extracts the selected data (Data Area, Entry Area and Fields) and makes an XML output based on the generated rules.

XML output is saved and shown by default in text mode (see figure below).

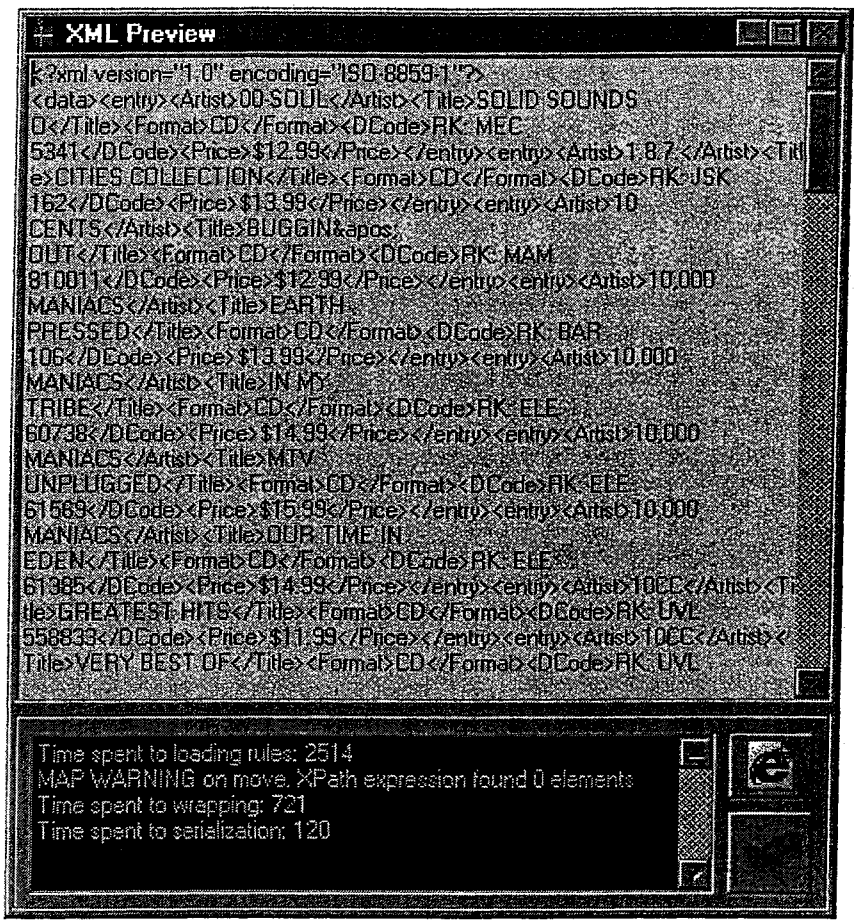


Figure 4. Text format view of the XML output.

Note: The XML output is saved automatically by the program. There is no need to save it explicitly.

To view the output in Internet Explorer (more visual),  button is pressed.

OR To close the preview window, the  button is used.

6.6 Previewing XML output in Internet Explorer (Preview)

The XML output in Internet Explorer can be previewed even after the XML preview window has been closed.

Note:  button does exactly the same as  button (shows XML output in Internet Explorer), but with this button, you can view it AFTER you have closed the preview window after extraction.

To view the XML output from the current selection in Internet Explorer, click  (**Preview**) toolbar button (see Figure 5 below).

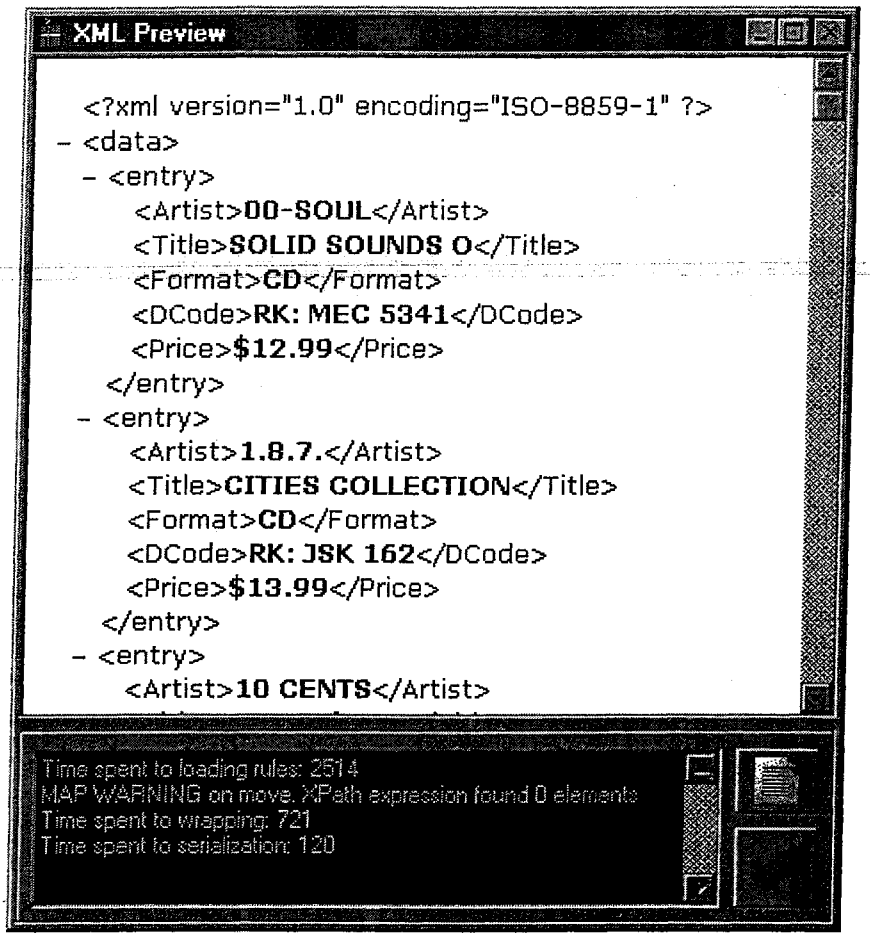


Figure 5. Internet Explorer mode shows the elements more conveniently in the XML hierarchy. Here you can also expand (+) and contract (-) the elements.

1. To close the preview window,  button is pressed.

OR To view the XML output in text format,  button is pressed.

Again, there is no need to explicitly save the XML file. It has already been saved after the extraction.

To make another extraction, return to 6.1.

To view the XML conversion rules created during extraction, go to 6.7.

6.7 Viewing the conversion rules (View Rule Set)

You can view Data Extraction Language rules created after you extracted the data.

1. To view the DEL conversion rules,  (**View Rule Set**) button is clicked.

2. To make changes to the conversion rules, DEL Editor is used.

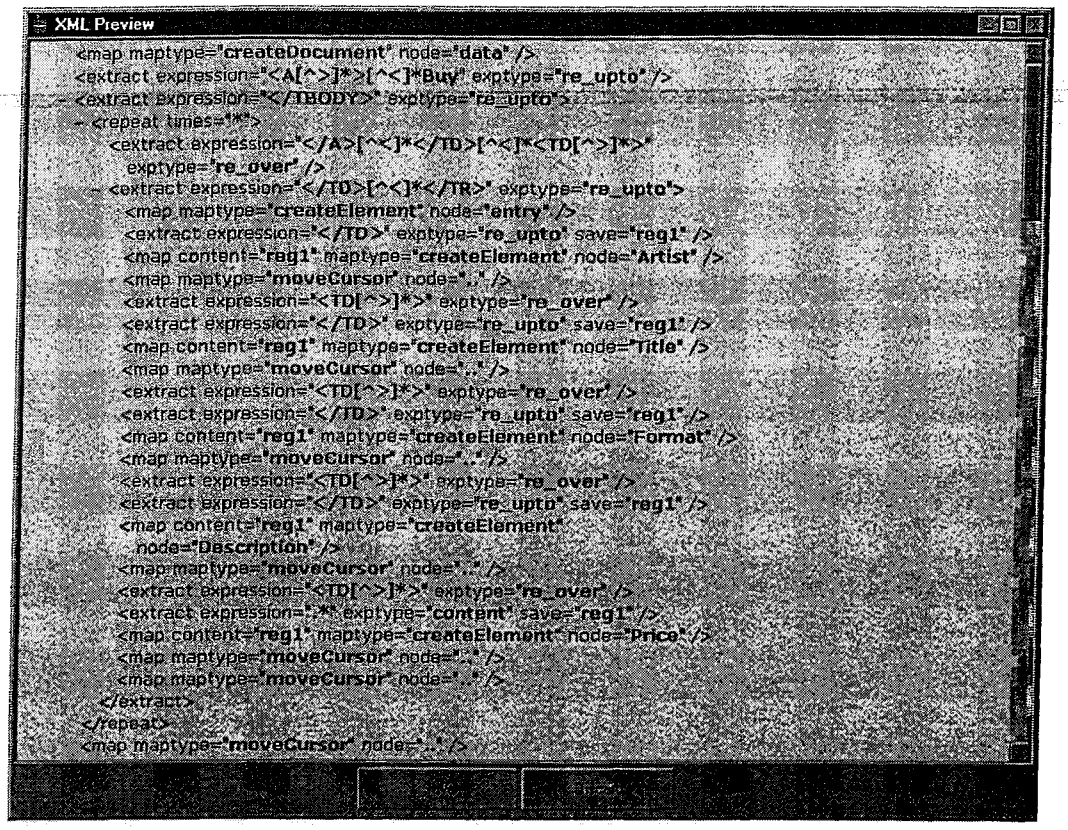


Figure 6. Previewing the conversion rules.

7 VERSION HISTORY

7.1 Version 0.8 Beta

- First official Beta release. Last known build is Build 0820 (01-26-2000).
- Used proprietary HTML parser (not fully compatible with Netscape Navigator or Internet Explorer).

7.2 Version 1.0

- Internet Explorer HTML parser.

7.3 Version 1.1

- Well-formed XML output.

7.4 Version 1.2

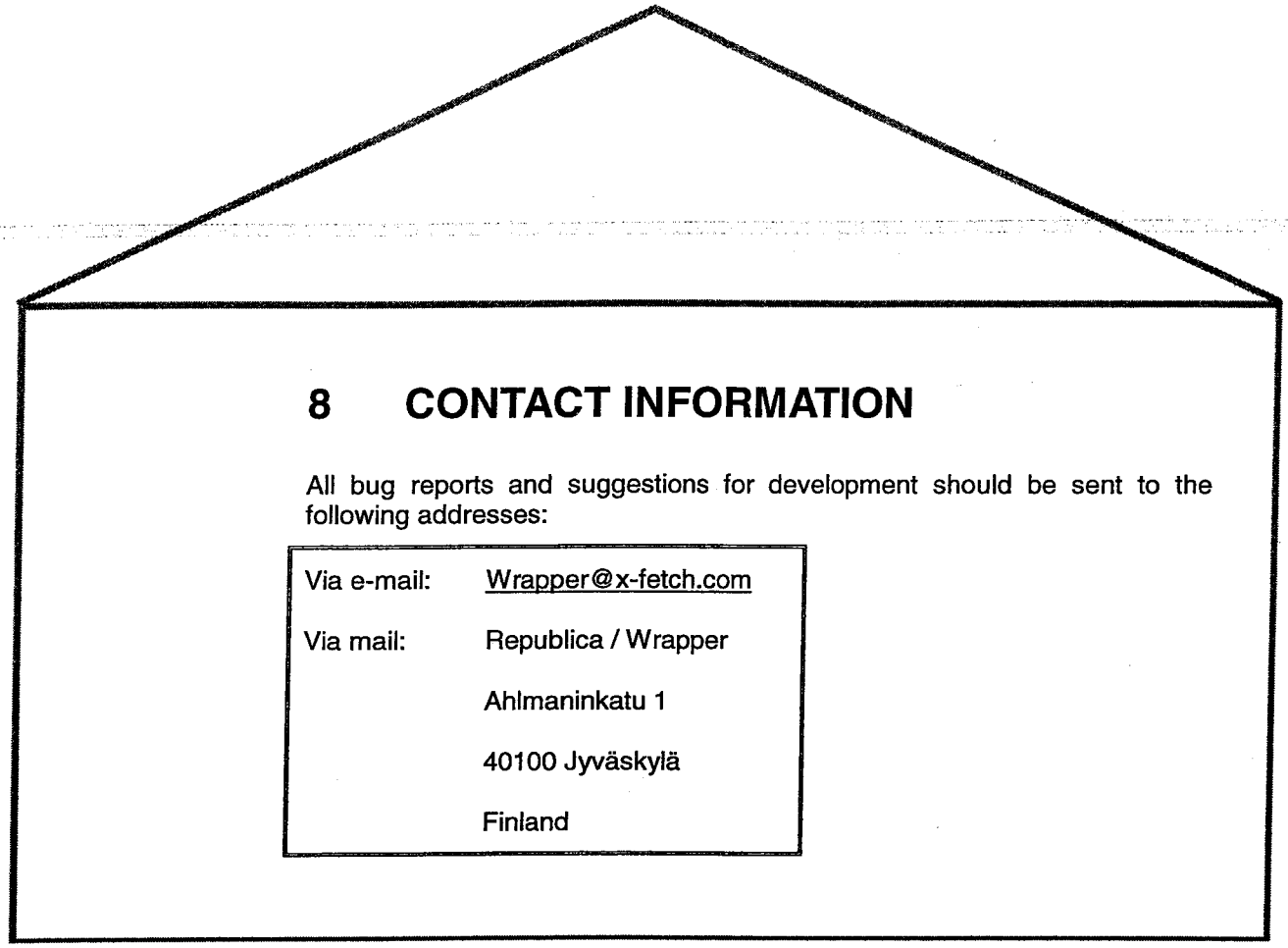
- Multiple data/entry areas support.
- New Java-based engine (version 1.3) (08-31-2000).

7.5 Version 1.3

- Dual engine support.
- New version 1.4 engine.

7.6 Version 1.3.22 (beta)

- Possibility to extract data using an already created rule file has been added.
- **Preview** function in the Browser (Internet Explorer) has been changed.



8 CONTACT INFORMATION

All bug reports and suggestions for development should be sent to the following addresses:

Via e-mail:	Wrapper@x-fetch.com
Via mail:	Republica / Wrapper
	Ahlmaninkatu 1
	40100 Jyväskylä
	Finland