Definition of terms

Adaptation model
A model of how an *adaptive system* should behave.

Adaptive Hypertext System
A system which offer to tailor hypertext *documents* dynamically according to the knowledge, needs and preferences of each *user.*

Adaptive System
A system that adapts to the user. It is more general than an adaptive hypertext system.

AE
Adaptation engine responsible for performing the adaptation according to rules in the *adaptation model*.

AI
The field of Artificial Intelligence.

AHS
See *Adaptive Hypertext System*

AM
See *adaptation model*.

Author
A person that have written and structured a collection of hypertext *documents*.

Browser
The environment software in which the *user* is exploring *hypertext documents*.

C
Programming language.

Candidate concept
In the process of *conceptualization*, all *terms* surviving *lexical analysis* are candidates to be chosen as the *concept*.

Concept
A concept is a descriptor of the content of some particular chunk of information, or *knowledge source*. In this thesis, the concepts are abstracted from their respective knowledge sources. See also *document concept* and *element concept*.

Conceptualization

The process of abstracting *concepts* from the *documents* and *elements* of the domain.

DFA

A Deterministic Finite Automaton object, or DFA, represents words in a network of interconnected characters. Whole words can be found by following paths of linked characters.

DM

See *Domain model*.

Document

By a document, we think of a collection of formatted information, e.g. including *text*, graphics, tables and the like. For this thesis, we assume the documents are published on the *Web*, if not otherwise noted.

Document concept

Concept abstracted from a *document*.

Domain

A domain is a collection of *documents*, where the documents are somehow related around one or more abstract, higher level *concepts*.

Domain model

A domain model is a description of how the domain looks like regarding content and structure. The domain model is built from *concepts* and *relations*.

DSL

Domain specific list used by the *system* in the process of *conceptualization*, that is a list of *terms* that the *author* has decided to be important for the *domain*.

Element

A *document* consists of elements of different types. HTML provides easy integration of elements like graphics, *text*, tables, lists and the like. In HTML, the form of the elements are determined by the browser by use of *tags*.

Element concept

Concept abstracted from an *element*.

Emphasizer element

Common notion for *elements* surrounded by the <B>, <I>, <EM>, <STRONG>, <U> and <CITE> tags. Such elements get emphasised by the browser, so they are likely to attract the attention of the *user.*

HCI

The field of Human Computer Interaction aims at achieving user friendly

systems. HCI concerns both the design of user *interfaces* and tools that support the developement and implementation of interfaces.

Heuristic

Rules based on observations from a small set of data. We use two types of Heuristics, both for finding *concepts* and for finding *relations*. Each Heuristics employ a *value* that, when the Heuristic decides to fire on a *term*, is added to the total *score* of the term.

High quality

When we speak of a *domain model* of high quality, we think of a resulting model that would please the *author* when compared to his view of the *domain*.

HTML

The Hyper Text Markup Language (HTML) is the language in which most *documents* on the *Web* is written today. HTML uses mark up *tags* to specify the formatting of documents.

Hyper reference

A link between two *documents* on the *Web* provides the *user* the option to jump from the first to the other when selecting the link.

Hypertext

Documents written in the *HTML* language.

Information source

IR-techniques produce different sets of terms. These techniques are referred to as information sources. *DSL* and *lexical analysis* are examples of information sources.

Interface

The interface is what is between the *user* and the *system*.

Internet

A network of networks, providing several services like e-mail, news, the *Web*, etc.

IR-techniques

Well known methods from the field of Information Retrieval that basically process *text*.

IUI

The field of Intelligent User Interfaces (IUI) constitute the intersection of *AI* and *HCI*, and concerns how to enhance the usability, simulating intelligent behaviour in the dialogue with the user.

Knowledge source

Elements or *documents* filled with content in *HTML*, we call knowledge sources.

Lexical analysis (LA)

The process of transforming a stream of characters into a list of lower case *terms* or “tokens”. When combined with stopword removal, the idea is that any text will be transformed into a list of terms where the most frequent words in the english language, are removed.

Models

Models describe something so that the system understands it. In this thesis, models for *domain* and *users* are represented in terms of *concepts* and *relations*.

PHP

Programming language suitable for server side programming on the Web.

PROLOG

The declarative programming language PROLOG (programming in logic) is often used when *AI* is involved.

Relation

Relates *concepts*. There are different relationship types, we use parental, prerequisites, deeper explanations and synonyms.

Score

When *conceptualising* some text, the *terms* are given *values* from different *Heuristical* rules. The score is the sum of all these values.

Sets on the form $\mathbf{C_i}$

The union of all sets produced by the *information sources*. The result constitute the *candidates* that compete for presidency as the *concept*.

Sets on the form K_{DOC}

Holds all the *element concepts* and the *document concept* for a specific *document*.

Sets on the form $\mathbf{S_{P\ LA}}$

$S_{P\ LA}$, $S_{P\ DSL}$, $S_{P\ EMP}$ and $S_{P\ TF}$ are sets of *terms* that are produced when a paragraph *element* is exposed to the different *information sources*.

Sets on the form $\mathbf{R_{type}}$

All relations in the domain of the specified type

Stemming

The automatic fusion of *term* variants into one common stem.

Stopword removal

See *lexical analysis*.

System

With system, we mean the software running on a computer.

Tag

The *author* can use tags to mark up and format *documents*. In order to make a sentence appear in bold, the <B> tag is placed in front of the sentence, and it is ended with an </B> tag.

Term

A term is a word that has gone through the process of *lexical analysis*.

Term frequency

Indicates how many times a *term* occurs in a *text*.

Text

A collection of words.

TF

See *term frequency.*

UM

See *user model*.

User

The user is a person that is using a computer system. The user is communicating with the system through the *interface*. Users have different skills and can be broadly classified as novices, intermediates or experts. Novices are assumed to have no knowledge at all in the domain at the time of starting the learning process, intermediates have the basic knowledge, while experts are assumed to have deeper knowledge.

User model

A model of preferences and the knowledge held by a user.

Value

Attached to the *Heuristics* are values, which are numbers to be given to each *candidate concept* that leads the corresponding Heuristic to fire.

Web

The biggest service on the *Internet* is the World Wide Web. People often mistake the Web for the Internet, but the two are not the same.