

**Two Approaches for Implementing Intranet Content Management  
A Case Study for Standards and Regulations  
in the Tobacco Industry**

Name : Budi Santoso  
Student Number : 15002  
Address : Kasernenstraße 12, 21077-Hamburg  
Place : Hamburg – Germany  
Date : 7.11.2001

Supervising Professor : Prof. Dr. rer. nat. Florian Matthes  
Second Professor : Prof. Dr. Friedrich H. Vogt  
Supervising Assistant : Dipl.-Inf. Axel Wienberg  
Company Supervisors : Mr. Dirk Bendixen, Mr. Peter Moderegger,  
Mr. Herbert Buchholz, Mr. Frank Twesten

Information and Communication Systems  
Technische Universität Hamburg-Harburg  
Germany

## **Abstract**

As a case study for Intranet Content Management, this thesis discusses the implementation of a content management system in Hauni Maschinenbau AG to manage its external standards and internal regulations. A new system which requires a smaller number of employees is needed, taking into account of the restructuring policy of the company in the Quality Assurance department. The following subsections will explain, respectively, the organization of this thesis, an overview about Hauni Maschinenbau, the standards which have to be managed, and the project requirements by Hauni.

# Table of Contents

<b>Abstract</b> .....	<b>i</b>
<b>Table of Contents</b> .....	<b>ii</b>
<b>Table of Figures</b> .....	<b>iv</b>
<b>Chapter 1. Introduction</b> .....	<b>1</b>
1.1. Overview .....	1
1.2. Hauni Maschinenbau AG .....	1
1.3. Standards.....	2
1.3.1. Definition of Standard .....	2
1.3.2. External Standards .....	2
1.3.3. Internal Regulation.....	5
1.4. Project Requirements by Hauni.....	6
<b>Chapter 2. Project Concept</b> .....	<b>8</b>
2.1. Decentralization of Regulation Creation .....	8
2.2. Workflows.....	9
2.2.1. External Standard .....	9
2.2.2. Internal Regulation.....	10
2.3. Analysis of Standards Documents .....	12
2.3.1. External Standards .....	12
2.3.2. Internal Regulations.....	13
2.4. Design and Functionality of Web Interface.....	15
2.4.1. Login Page.....	17
2.4.2. Main Page.....	18
2.4.3. Full Text Search.....	19
2.4.4. Sitemap.....	21
2.4.5. Support Pages .....	21
<b>Chapter 3. Implementation in VIP</b> .....	<b>24</b>
3.1. The VIP System by Gauss Interprise AG .....	24
3.1.1. System architecture of VIP'ContentManager .....	24
3.1.2. VIP Three-Server Concept.....	26
3.1.3. Structure of a VIP Object .....	27
3.1.4. Creating an Object in VIP .....	28
3.1.5. Editing an Object in VIP .....	29
3.1.6. Releasing or Rejecting an Object in VIP .....	30
3.1.7. Setting Access Rights .....	31
3.1.8. Inheritance of Access Rights .....	32
3.2. Architecture of VIP Solution in Hauni .....	32
3.3. Roles and Rights.....	34
3.4. Directory Structure .....	35
3.5. Login Process .....	38
3.6. Main Page.....	40
3.7. Full Text Search .....	40
3.8. Migration from Old System into VIP .....	41
3.8.1. External Standards .....	41
3.8.2. Internal Regulations.....	42
<b>Chapter 4. Implementation in CoreMedia</b> .....	<b>43</b>

4.1. The CoreMedia Application Platform by CoreMedia AG .....	43
4.1.1. CoreMedia Content Application Platform Architecture .....	43
4.1.2. CoreMedia Production and Live Servers Concept .....	45
4.2. Web Content Management Systems implementation.....	46
4.2.1. Architecture of CoreMedia Web Content Management System Solution .....	47
4.2.2. Roles and Rights .....	48
4.2.3. Directory Structure .....	49
4.2.4. Additional User Administration .....	50
4.2.5. Full Text Search.....	50
4.2.6. Migration from Old System .....	50
4.3. Document Management Systems implementation .....	51
4.3.1. Architecture of CoreMedia Document Management System Solution .....	52
4.3.2. Roles and Rights .....	54
4.3.3. Directory Structure .....	54
4.3.4. Additional User Administration .....	54
1.1.1. Full Text Search.....	54
4.3.5.....	54
4.3.6. Migration from Old System .....	54
<b>Chapter 5. Comparison .....</b>	<b>55</b>
5.1. Installation .....	55
5.2. System Components.....	56
5.3. Server .....	56
5.4. Access Rights .....	57
5.5. Editing Object.....	59
5.6. Caching.....	59
<b>Chapter 6. Summary and Conclusions.....</b>	<b>60</b>
<b>Appendix A. Project Description from Hauni in German .....</b>	<b>61</b>
<b>Bibliography .....</b>	<b>63</b>

# Table of Figures

- Figure 1. Workflow of Internal Regulation before Restructuring ..... 6
- Figure 2. Workflow of External Standards ..... 6
- Figure 3. Workflow of importing external standard ..... 10
- Figure 4. Workflow of importing external standard in xml format ..... 10
- Figure 5. Workflow of creating regulation ..... 11
- Figure 6. Workflow editing of regulation ..... 12
- Figure 7. Letterhead example of the DIN ..... 13
- Figure 8. Letterhead example of a Report ..... 14
- Figure 9. Letterhead example of a Regulation ..... 15
- Figure 10. Footnote example of a Regulation ..... 15
- Figure 11. Website structure ..... 16
- Figure 12. Dialog sequence of Login or Logout ..... 17
- Figure 13. Login page ..... 18
- Figure 14. Hauni Mainpage ..... 19
- Figure 15. Full Text Search - Simple ..... 20
- Figure 16. Search field in advance Full Text Search ..... 20
- Figure 17. Results of advance Full Text Search ..... 21
- Figure 18. Standards Sitemap ..... 21
- Figure 19. How To page ..... 22
- Figure 20. Frequently Asked Questions page ..... 23
- Figure 21. Technical Support page ..... 23
- Figure 22. Overview of architecture of VIP ContentManager ..... 25
- Figure 23. Server process tasks ..... 26
- Figure 24. VIP Three-Servers Concept ..... 27
- Figure 25. Structure of a VIP object ..... 28
- Figure 26. Settings options for individual access rights ..... 31
- Figure 27. Project Architecture in VIP ..... 33
- Figure 28. Website Roles and Rights ..... 35
- Figure 29. VIP GUI Applet for Navigation ..... 36
- Figure 30. Activity Diagram of Header1.jsp ..... 37
- Figure 31. Document Header ..... 37
- Figure 32. Activity Diagram of Navigation1.jsp ..... 38
- Figure 33. Activity Diagram for Check Login.jsp ..... 39
- Figure 34. Activity Diagram for Full Text Search.jsp ..... 41
- Figure 35. CM 3.1 System Architecture ..... 44
- Figure 36. CoreMedia Production and Live Servers ..... 46
- Figure 37. Project Architecture in CoreMedia as Web Content Management System ..... 48
- Figure 38. Website Roles and Rights ..... 49
- Figure 39. Project Architecture in CoreMedia as Document Management System ..... 53
- Figure 40. Mapping VIP to CoreMedia access rights ..... 58

# **Chapter 1. Introduction**

As a case study for the Thesis, it discusses an implementation of content management system in Hauni Maschinenbau AG to manage its external standards and internal regulations. A new system which requires a smaller number of employees is needed, taking into account of the restructuring policy of the company in the Quality Assurance department. An organization of this thesis, an overview about Hauni Maschinenbau, the standards which have to be managed, and the project requirement by Hauni will be explained in the following subsections respectively.

## **1.1. Overview**

The thesis comprises six chapters, of which this introduction is the first. The general organization is to describe the old business process and define the project, propose a concept considering the company restructuring policy in Quality Assurance department, the next step is to implement the proposed concept based on two different content management systems, and finally collect differences between the systems and summarize findings.

## **1.2. Hauni Maschinenbau AG**

Hauni Maschinenbau AG is the company that heads the international Körber Group's Tobacco Division. Hauni, with its eight Division companies, has an established reputation as the world's top manufacturer of complete systems for the tobacco industry with a product range covering primary processing equipment, secondary machinery for making and packing cigarettes and a wide selection of associated services.

The Tobacco Division fosters close links with customers and pursues an interactive approach to product development and production. It operates a worldwide production, sales and service network. In 1998, the Tobacco Division companies achieved sales worth DM 1,026m with a workforce totalling 3690.

## 1.3. Standards

To understand the importance and role of standard and regulation for Hauni, in the following will be given the working definition of a standard and regulation and describe the various kinds of standard and regulation used by Hauni. External standard is categorized by source of the standard e.g. DIN ISO which means that the DIN standard is originated from ISO. While internal regulation is categorized by area of utilization e.g. Normen für Elektronik (NE) which means that the regulation governs mainly in the area of electrical machine parts.

### 1.3.1. Definition of Standard

The following definition of standard is taken from International Organization for Standardization as the biggest organization for standardization in the world: “Standards are documented agreements containing technical specifications or other precise criteria to be used consistently as rules, guidelines, or definitions of characteristics, to ensure that materials, products, processes and services are fit for their purpose.” *[ISO]*

### 1.3.2. External Standards

Hauni uses external standards to make its products compatible with other parts or components which interface has been defined in the standard, enhance the quality and reliability of its products, and reduce the cost of design. On average Hauni buys 30 external standards in a month which are taken from a huge number of standards. The standards are then communicated and distributed widely within Hauni and serve as documentation for business units to design or assemble Hauni products. Since the number of standards is huge, finding the relevant one can be a problem and requires an efficient and systematic access to the collection of standards.

The procedure to make the external standards available for the company remains the same before and after the restructuring, and is shown in Figure 1.

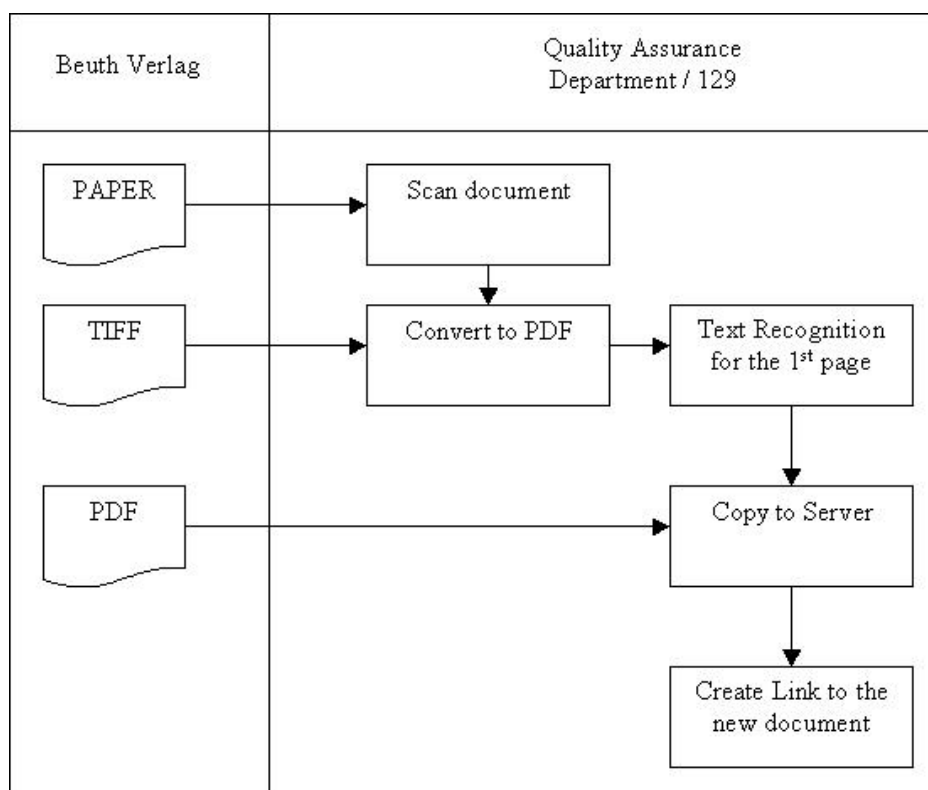


Figure 1. Workflow of External Standards

### 1.3.2.1. ISO

The International Organization for Standardization (ISO) [ISO] is a worldwide federation of national standards bodies from some 140 countries, one from each country.

ISO is a non-governmental organization established in 1947. The mission of ISO is to promote the development of standardization and related activities in the world with a view to facilitating the international exchange of goods and services, and to developing cooperation in the spheres of intellectual, scientific, technological and economic activity.

The existence of non-harmonized standards for similar technologies in different countries or regions can contribute to so-called "technical barriers to trade". Export-minded industries have long sensed the need to agree on world standards to help rationalize the international trading process. This was the origin of the establishment of ISO.



International standardization deals with technologies in diverse fields, such as information processing and communications, textiles, packaging, distribution of goods, energy production and utilization, shipbuilding, banking and financial services.

Industry-wide standardization is a condition existing within a particular industrial sector when the large majority of products or services conform to the same standards. It results from consensus agreements reached between all economic players in that industrial sector - suppliers, users, and often governments. They agree on specifications and criteria to be applied consistently in the choice and classification of materials, the manufacture of products, and the provision of services. The aim is to facilitate trade, exchange and technology transfer through:

- enhanced product quality and reliability at a reasonable price.
- improved health, safety and environmental protection, and reduction of waste.
- greater compatibility and interoperability of goods and services.
- simplification for improved usability.
- reduction in the number of models, and thus reduction in costs.
- increased distribution efficiency, and ease of maintenance.

Users have more confidence in products and services that conform to International Standards. Assurance of conformity can be provided by manufacturers' declarations, or by audits carried out by independent bodies.

### **1.3.2.2. DIN**

DIN, the German Institute for Standardization [*DIN*], is a registered association, founded in 1917. Its head office is in Berlin. Since 1975 it has been recognized by the German government as the national standards body and represents German interests at international and European level.

DIN offers a forum in which representatives from the manufacturing industries, consumer organizations, commerce, the trades, service industries, science, technical inspectorates, government, in short anyone with an interest in standardization, may meet in order to discuss and define their specific standardization requirements and to record the results as German Standards.

Standardization as undertaken by DIN is a service that aims to benefit the entire community. The results of its work have a significant influence on economic performance at both company and national level.

DIN Standards promote rationalization, quality assurance, safety, and environmental protection as well as improving communication between industry, technology, science, government and the public domain.

In DIN, standards work is carried out by some 26,000 external experts serving as voluntary delegates in more than 4,000 committees. Draft standards are published for public comment, and all comments are reviewed before final publication of the standard. Published standards are reviewed for continuing relevance every five years, at least.

### **1.3.3. Internal Regulation**

Standardization is of value not only between different organizations, but also within an organization, in this case between the business units of Hauni. Inside Hauni, they are used to communicate different areas of component design, which are reflected by the business units, to build one end product. The regulations are held confidential because they represent the experience gathered by Hauni as a competitive advantage.

Hauni Internal Regulation is created by the coordination between its internal business units. The proposal, which comes from one or more business units, will be discussed through several meetings with interested business units until the topic is mature. If the draft of the regulation has been agreed and passed the objection period the regulation will be applied to all business units.

As can be seen in Figure 1, the standardization commission used to have a lot of influence and responsibilities on the content of internal regulations. This commission coordinated business units in the making of an internal regulation.

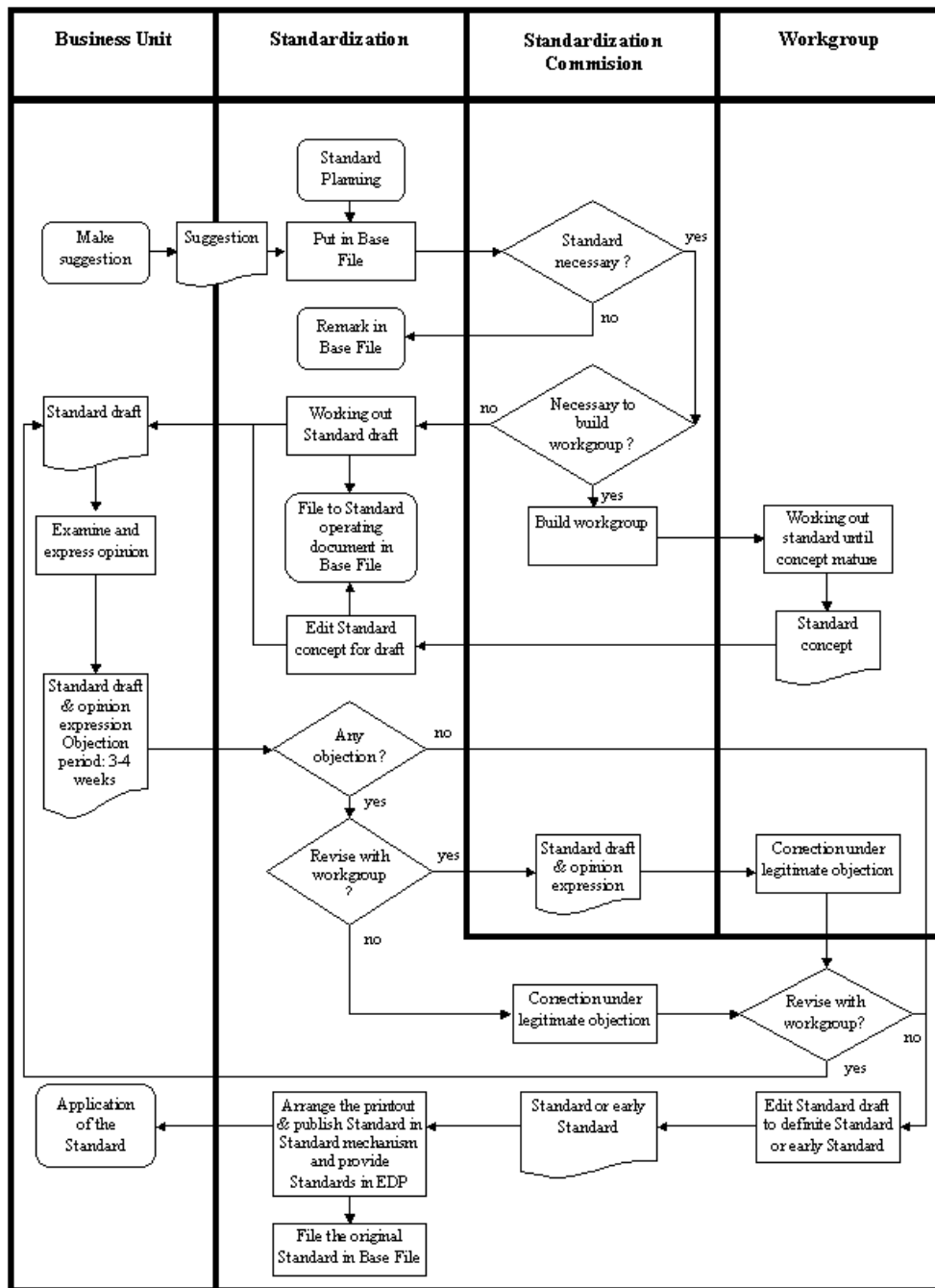


Figure 2. Workflow of Internal Regulation before Restructuring

In January 2001 Hauni restructured its standardization process. Standardization commission was dissolved. From that time the above workflow is not valid anymore and the functions of the commission were distributed to business units and Quality Assurance Department. The new workflow will be described in section 2.2.2.

## 1.4. Project Requirements by Hauni

Hauni posed the following requirements on the project which was the basis for this thesis. The original german version of this project requirements can be seen in the appendix A:

Distribution of the standards and regulations to other business units with the dissolution of the central function of the standards maintenance. Publishing of about 1300 standards and regulations at a central place in the Hauni intranet. This would supersede for the distribution of CDs and bring the data on the spot with a more up-to-date state.

Standards, which are also from external source, must be able to be seen and are personalized on the Internet for different users (suppliers, subsidiaries, technical customer service, customers). Ideally transformation of 300 internal regulations from interleaf into HTML format or embedded as a file. Editing of the regulations for the future owner in business units is using Microsoft Word.

The Quality Assurance department takes over only management of the documents, layout services, and filing the standards documents at disposal centrally in the Intranet.

Several sort and searching criteria or groups (target groups, subjects, main points) should be available. The previous full text search must be guaranteed furthermore.

## Chapter 2. Project Concept

The following sections will explain the new concept of standard and regulation management. Starting from the company restructuring policy in the dissolution of standardization commission, workflows which explain some scenarios in the procurement of standards and regulations in an up-to-date state, analysis of detail document structure, and finally some snapshots of a mock-up model for the website.

### 2.1. Decentralization of Regulation Creation

Since the dissolution of Standardization Commission, the functions of this commission are to be distributed between Business Unit and Quality Assurance Department. In a Hauni internal document [HOM01], the distribution is detailed as follows:

Functions of Business Unit related to Hauni regulation are:

- Creation of Hauni regulation.
- Coordination of regulation, which should apply to several business units.
- Maintenance (updating and modification) of existing Hauni regulations.
- Technical information on contents of Hauni regulations.

Functions of Quality Assurance department related to Hauni regulation are:

- Maintenance of Hauni regulation which is created or modified by business units in network. To this end, an intranet structure is currently being constructed. Maintenance is successful only if it is comprehensibly coordinated.
- Maintenance of external standards (DIN etc.) in the network.
- Procurement of new external standards on request.
- Distribution of information about modifications in external standards.
- Stop the archive for neglected old standards
- Assistance for the layout of regulation which is created or modified by the business units

As we can see above, every business unit is now responsible for the content of its own regulations which was intensely intervened by standardization commission before the restructuring policy. Some of the workforces from the previous commission are

positioned in the quality assurance and manage only procurement of standards and regulations for business units and provide uniform layout of the documents in the network.

The internal regulations contained in the website will also be available in the extranet especially for Hauni subsidiaries in Richmond, US. The creation of Hauni internal regulation is not only being done in the headquarter but also in its subsidiaries. Therefore any proposal for a new regulation in the extranet must be handled by the features of the new system.

## **2.2. Workflows**

This section describes conceptually the workflow of both external standard and internal regulation.

### **2.2.1. External Standard**

External standard, which is bought from Beuth publisher, is available in 3 different formats. These formats are as printed document, image file for each page, or in a pdf file. Nowadays Hauni receives the Standard in a pdf file format which simplifies the procedure of importing the external standard, but the older formats still should be supported in case Beuth publisher switches their terms of delivery again.

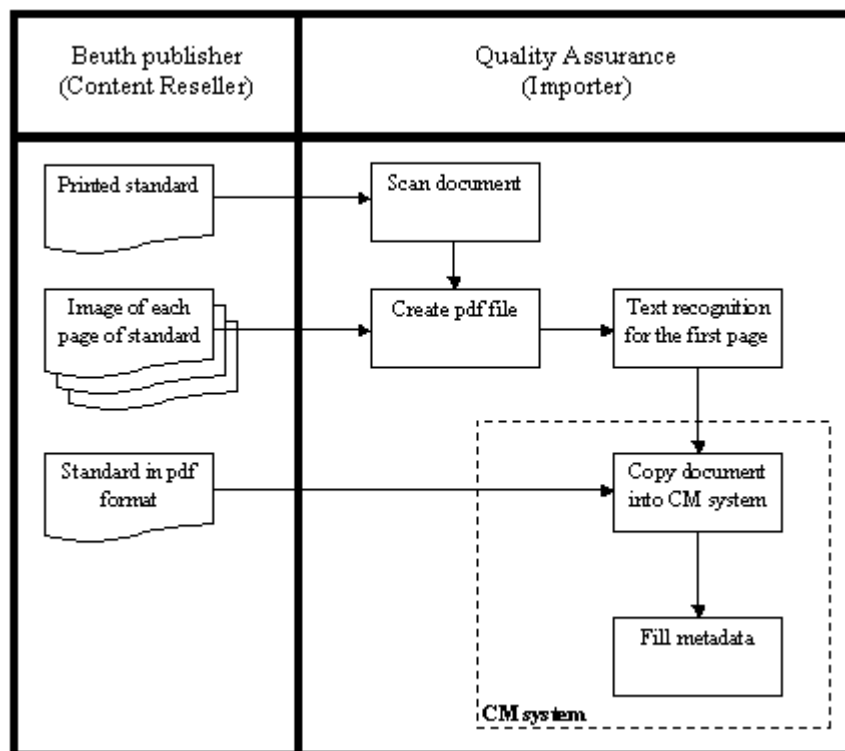


Figure 4. Workflow of importing external standard

If the external standard supplier, which in this case is Beuth publisher, provided the documents in xml format then the process would be much simpler and the metadata could be imported automatically as well. The workflow would be as follows:

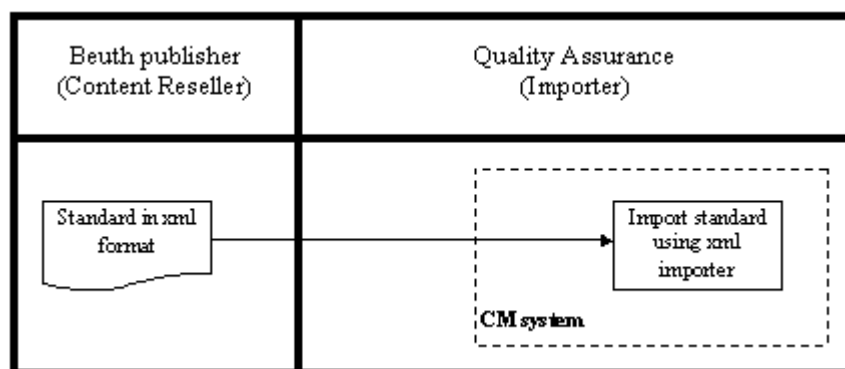


Figure 5. Workflow of importing external standard in xml format

In the case that the xml document structure differs XSLT would be used in the process of importing the documents.

## 2.2.2. Internal Regulation

Regarding the distribution of the functions the following workflows can explain how new regulation is created and old regulation is modified.

Business units are working independently in the creation of a new regulation. If the regulation has been finished the final document will be sent to Quality Assurance Department. The document layout must follow the existing regulation which must also be uniform with the other documents. If every layout check is passed then the document will be published into the network otherwise it will be returned to the author in this case is the business unit.

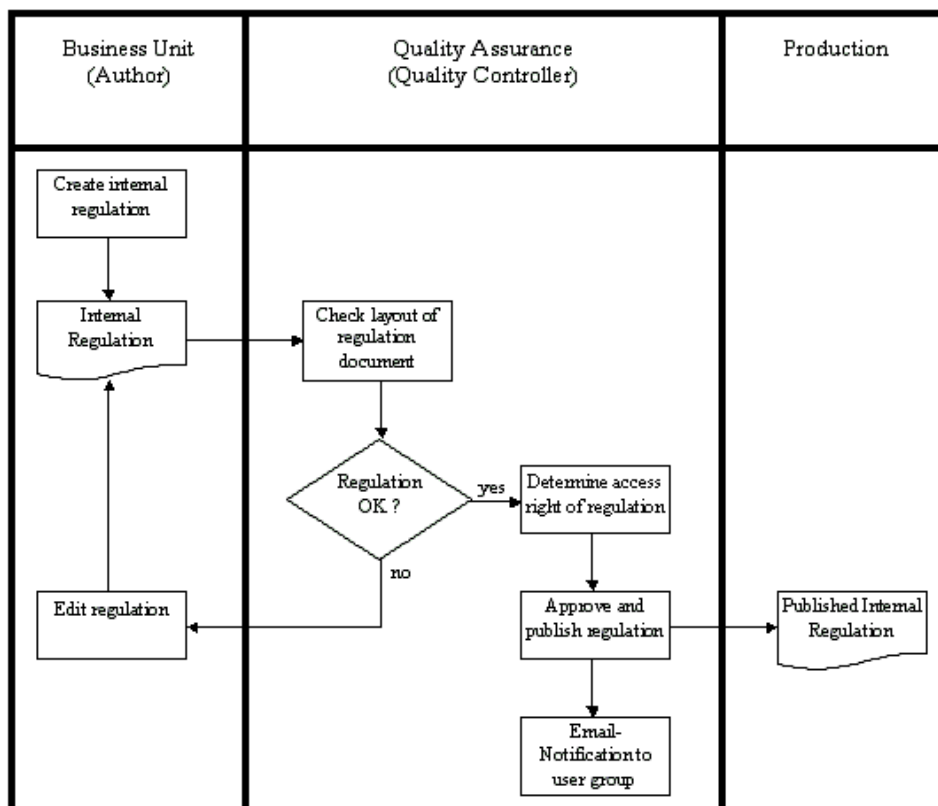


Figure 6. Workflow of creating regulation

Sometimes the regulation needs to be updated with new information or adjusted according to a new technology. In this case the business unit can edit the old published regulation and after that send the document to Quality Assurance for the checking. In order to keep the old document as backup the versioning feature of a CM system is needed.



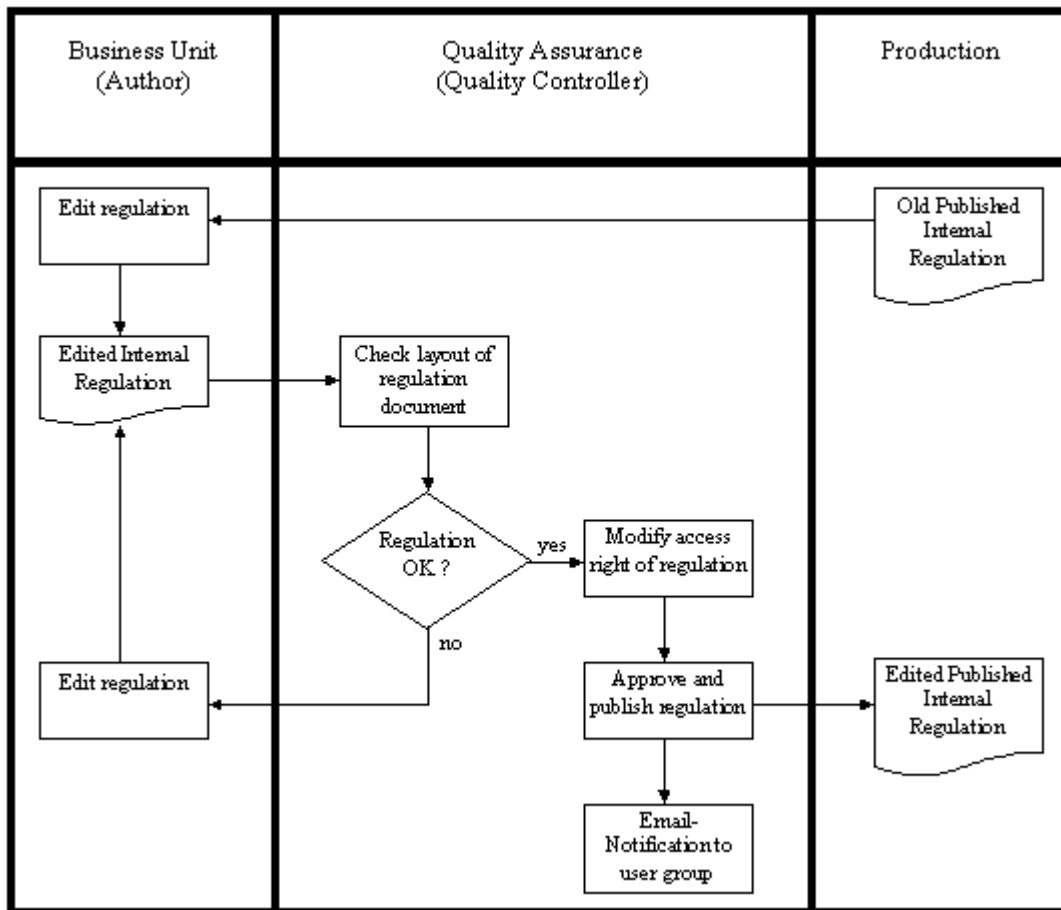


Figure 7. Workflow editing of regulation

## 2.3. Analysis of Standards Documents

This section analyzes the structure of the external standard and internal regulation documents.

### 2.3.1. External Standards

In the past the documents were bought from Beuth Publisher and received in printed document. All of the pages in the document were then scanned and the images were copied into a pdf file, and the first page was transformed from image format to text with the Text Recognition feature in Adobe Acrobat. Nowadays Hauni receives a complete pdf file already in text format from Beuth Publisher.

The purpose of the text recognition for the first page of the document is to provide keywords to the search function since the first page always contains document abstract. This procedure is not done for the rest of the pages which are still in image

format. In this case the full text search which is provided in Adobe Acrobat cannot be used optimally. However, an abstract could provide some important keywords to be captured by the search engine.

The document consist of:

- Title (Kaltband aus Stahl für eine Wärmebehandlung)
- Category (Technische Lieferbedingungen)
- Subtitle (Teil 4: Federstähle und andere Anwendungen)
- Version (EN 10132-4 : 2000)
- Month of Issue (Mai 2000)
- DIN number (EN 10132-4)
- Contents / Core Information



Figure 8. Letterhead example of the DIN

External Standards are divided into several categories:

- DIN (Deutsches Institut für Normung)
- DIN EN (Deutsches Institut für Normung - Euro Normen)
- DIN EN ISO (Deutsches Institut für Normung - Euro Normen – International Standards Organization)
- DIN V (Deutsches Institut für Normung – Vornormen)
- ISO (International Standards Organization)
- VDI/VDE (Verband Deutsche Ingeieur / Verband Deutsche Elektronik)

### 2.3.2. Internal Regulations

The internal regulation is created through several meetings. For each meeting there is a report (Protokoll), which is also documented. This report is created in MsWord97 format and after approval it will be applied as the regulation and documented in pdf format.

Content of internal regulation contains text, image, table, page number, inter document reference for attachment, bullets, numbering, and formatted text.

One document will be packed into only one html file. However, attachments can be included as separate files. These attachments can be pdf or other html files.

Content of internal regulation usually contains text and images but sometimes can also contain table, intra document reference, inter document reference, bullets, numbering, and formatted text.

Text takes most of the pages. If the document consists of many pages, sections will be created and a table of contents, with references to the sections, is put at the beginning of the document. This intra file reference - also called bookmark – is the replacement for page numbers in printed documents in order to find a specific section quickly.

### 2.3.2.1. Report (Protokoll)

Protokoll vom 24.02.1997					
<b>Standardisierung / Harmonisierung</b> der Verfahrensweisen / Gerätefamilien und Dokumentation			04.04.2001 JR/Rau SH240297.DOC		
<b>Teilnehmer:</b>			<b>Verteiler:</b>		
Frau:	Deitlaff	121	Herren:	Knospe	
Herren:	Busse-Riepshoff	174		☒Frau Rohland	705
	Drabner	126		Dr. Tolasch	723
	Fritz	127		Blais	125
	Grave	174		Frank	171
	Lesch	171		Klein	121
	Meiners	171		Marsau	174
	Prien	126		Müller	751
	Rabe	Unverselle		Preuß	176
	Rudnik	126		Dr. Siems	175
	Stöcker	topack		Wedekind	Unverselle
				Dr. Wratil	723

Figure 9. Letterhead example of a Report

The document consists of:

- Meeting date (24.02.1997)
- Title (Standardisierung / Harmonisierung der Verfahrensweisen / Gerätefamilien und Dokumentation)

- Date of Issue (04.04.2001)
- Author (JR/Rau)
- Filename (SH240297.DOC)
- Participants
- Carbon Copy / List of people to receive a copy of this document
- Contents / Core Information

At the end of the document there is usually a date and place for the next meeting

### 2.3.2.2. Regulation

 <b>HAUNI</b>	<b>Sprachenkennzeichnung</b>	<b>NE0106</b>
erstellt durch: Rudnik/Borchert/126		Seite 1 von 1

Figure 10. Letterhead example of a Regulation

gültig ab: 17.10.97	Bem.: Neu aufgenommen
---------------------	-----------------------

Figure 11. Footnote example of a Regulation

The document consists of:

- Set through (Rudnik/Borchert/126)
- Title (Sprachenkennzeichnung)
- Regulation number (NE0106)
- List of Content, if the document contains of many pages
- Valid from (17.10.97)
- Remarks (Neu aufgenommen)
- Contents / Core Information

Regulations are divided into several categories:

- NE (Normen für Elektronik)
- NK (Normen für Konstruktion)
- ND (Normen für Dokumentation)
- NQ (Normen für Qualität)

## 2.4. Design and Functionality of Web Interface

Design of the website should contain at least external standard and internal regulation which are classified according to its category and a full text search feature. All files related to its category will be put in the same directory or subdirectory.

External standards which are distributed by Beuth publisher are issued for a certain company which pays for its publication rights or commercial interests. Meanwhile Internal regulations which are created by business units in Hauni give competitive advantage. Because of the importance and sensitivity of the data the website must be password-protected and a login page is necessary at the beginning of the sequence.

References to the newest standards or regulations will be created in the main page because the new documents will be accessed more frequently than the older ones. The number of references should be visually appropriate for the main page and the creation of these references must be processed automatically by the code.

In the case of a problem related to the website several stages of help will be provided such as section 'How to' which offers a first guideline to using the website, 'Frequently Asked Questions' as the collection of questions about specific problems which have been or could be frequently asked to the support staff, and 'Contacts' as the last place to ask for help.

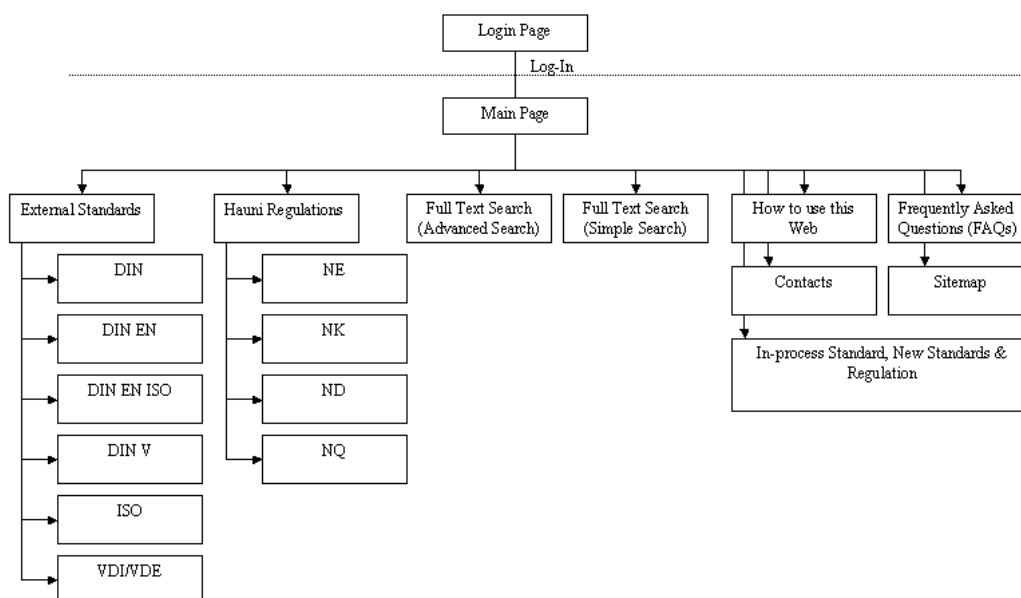


Figure 12. Website structure

Figure 12 shows a sequence of a user in the process to access the content. On every page which is password-protected a template is used to provide a user authentication for the current session. Only if the user has logged in and owns a read access to the page, the content can be shown. In the case that user pressed the logout button, then the session is terminated.

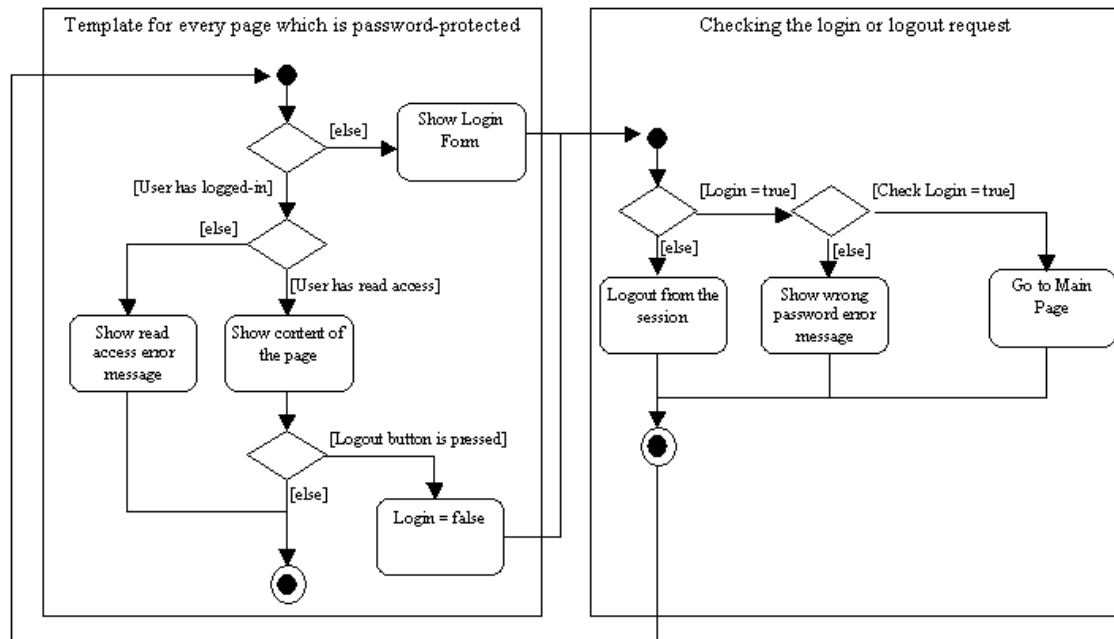


Figure 13. Dialog sequence of Login or Logout

### 2.4.1. Login Page

Before the user browses into the Standards Main Page the user will have to type in the username and password correctly. User who forgets the password must contact to the administrator, then the password will be reset. Figure 13 gives an impression of what the login page could look like.



Figure 14. Login page

## 2.4.2. Main Page

This page (illustrated in Figure 14) consists of some references to the Standards (either external or internal), reference for the Full Text Search (simple or advanced), information about certain internal regulations, ordered DIN, new purchased DIN. Depending on the DIN license, the external standards references will not be displayed for outsiders. The fast simple search will be put on the top left corner in this page. List of the Standards are categorized as following:

1. External Standards
  - DIN
  - DIN EN
  - DIN EN ISO
  - DIN V
  - ISO
  - VDI / VDE
  
2. Hauni Regulations
  - Normen für Elektronik
  - Normen für Konstruktion
  - Normen für Dokumentation
  - Normen für Qualität

Those Standards are sorted by DIN number.

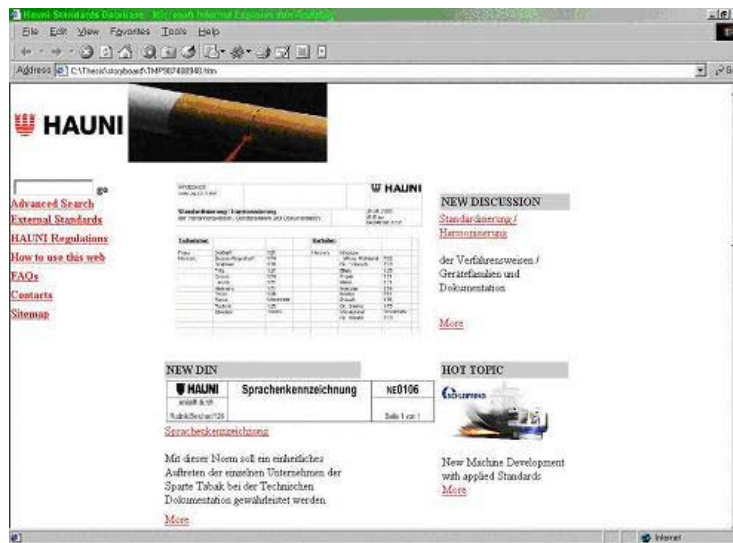


Figure 15. Hauni Mainpage

### 2.4.3. Full Text Search

There are 2 types of search functions:

- Simple search is more convenient and easier function if the user does not want to be bothered with detailed options of the search.
- Advanced search will provide filter based on category in the search. Users can search the word in specific categories (e.g. if the user only wants to search in DIN & ISO, he can click only option DIN & ISO). By pressing 'Select all' button all the categories will be selected and 'Clear all' button is on the other hand.

The result of the search can be sorted by percentage of relevance, standards category, date of issue, DIN number and ordered ascending or descending.



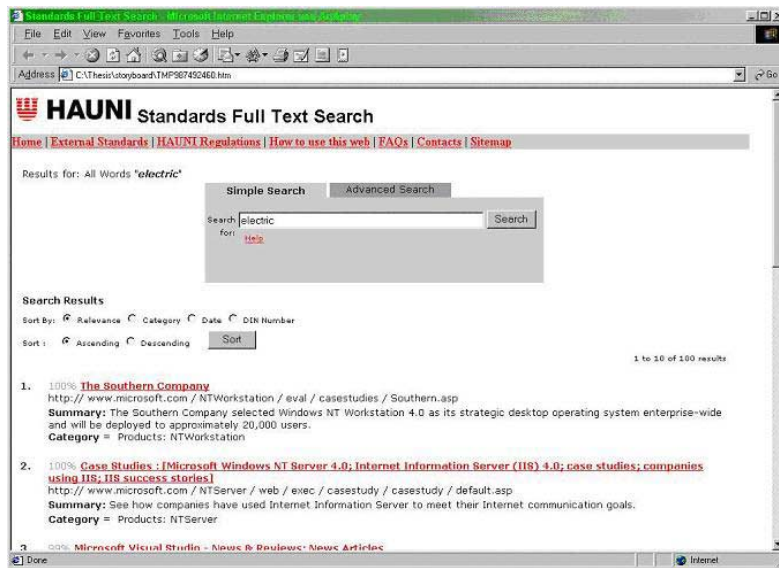


Figure 16. Full Text Search - Simple

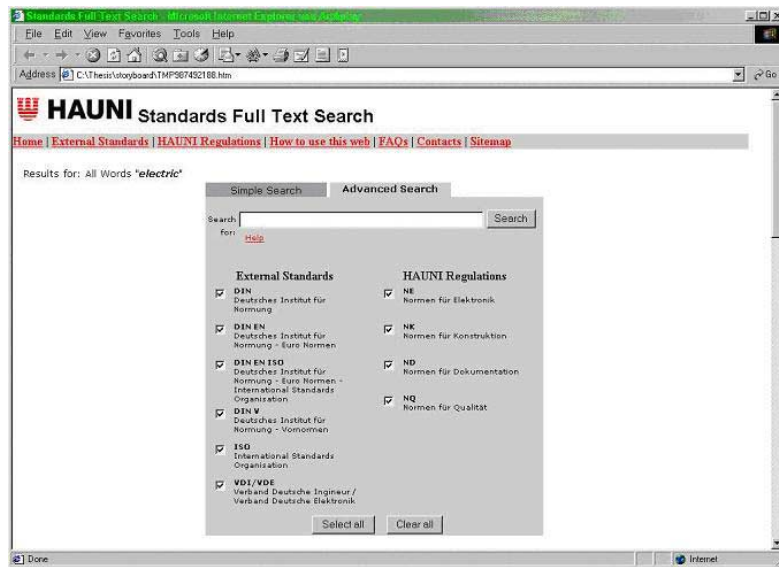


Figure 17. Search field in advance Full Text Search

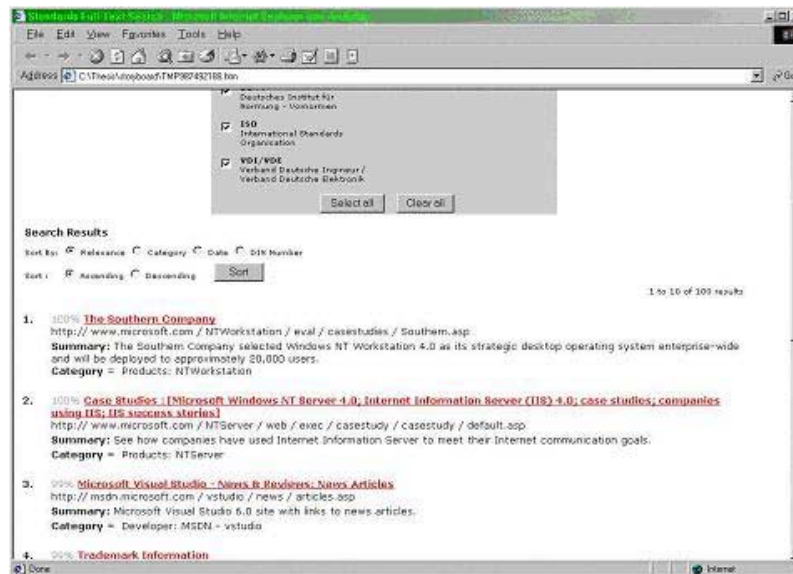


Figure 18. Results of advance Full Text Search

## 2.4.4. Sitemap

The structure of the website can be seen in this page. Changing the structure of the website will automatically update the contents of the page. Layer below standard category e.g. DIN will not be shown.

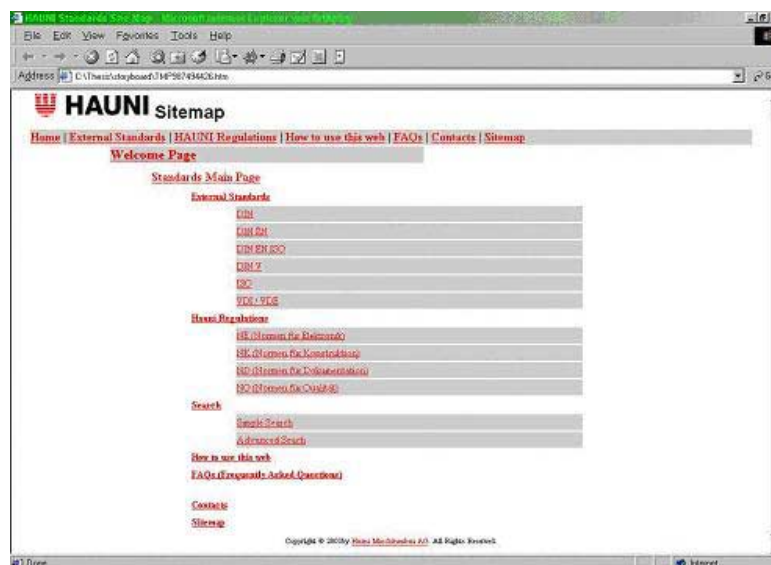


Figure 19. Standards Sitemap

## 2.4.5. Support Pages

To give support to the user who will use the website some static support pages will be provided. A short tutorial on how to use the website can give a first idea for dealing with problems in using the website. If this does not help, user can go to frequently asked questions (FAQs) page in order to find a common asked questions. Contacting the technical support would be the last option and the phone numbers or email addresses can be found in technical support page. If there are some common questions asked to the technical support the question will be put to the FAQs page.

### 2.4.5.1. How To page

Basic steps to deal with the website will be explained in this page. Only short explanation will be shown in the first page. Detail of the information can be made and a reference to the page will be created according to the topic.



Figure 20. How To page

### 2.4.5.2. Frequently Asked Questions (FAQs)

A summary of questions is written on the first section of the page and detail of the answer is in the next section.

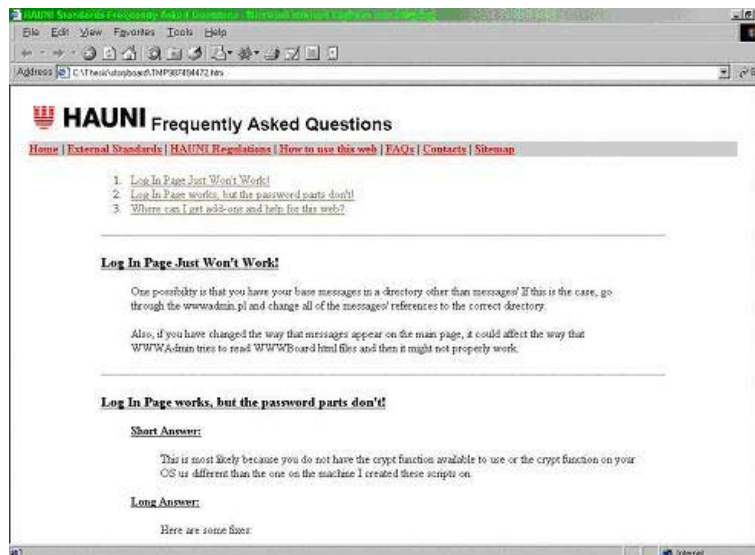


Figure 21. Frequently Asked Questions page

### 2.4.5.3. Technical Support Page

More technical support persons could be assigned if the number of standard and regulation is increasing. The division of the responsibility can be support for external standard, support for internal regulation, or network support. Information about the contact consists of only phone number and email address.

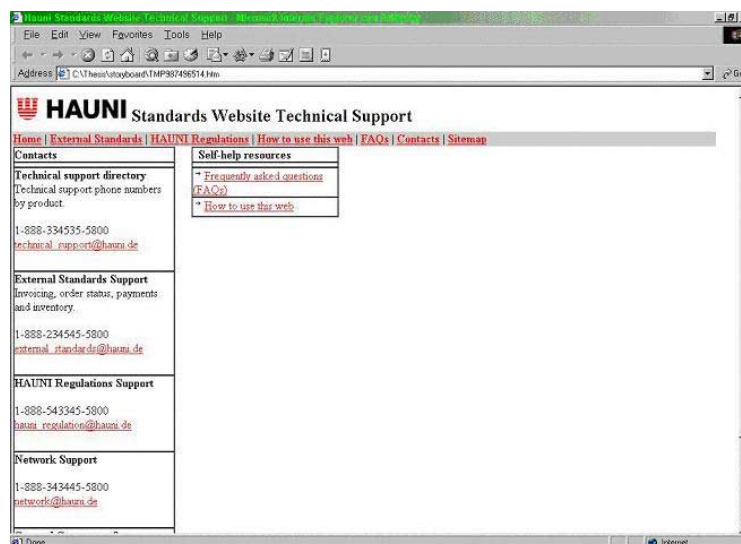


Figure 22. Technical Support page

## Chapter 3. Implementation in VIP

### 3.1. The VIP System by Gauss Interprise AG

The following subsections will describe some basic concepts and procedures for using VIP. Only those concepts and procedures that are relevant for implementing the project concept will be described.

#### 3.1.1. System architecture of VIP'ContentManager

VIP'ContentManager is a multi-tiered client/server solution, implemented entirely in Java. Client/server communication within the intranet takes place via socket connections. All activities for administration of the website are handled via these connections. Only transfers of HTML pages are performed using the HTTP protocol, which means that the web server is needed for transmission of this data only. All other activities are controlled by the VIP servers. In Internet connections, the VIP system's client/server communication is achieved by HTTP tunnelling if a socket connection is not possible.

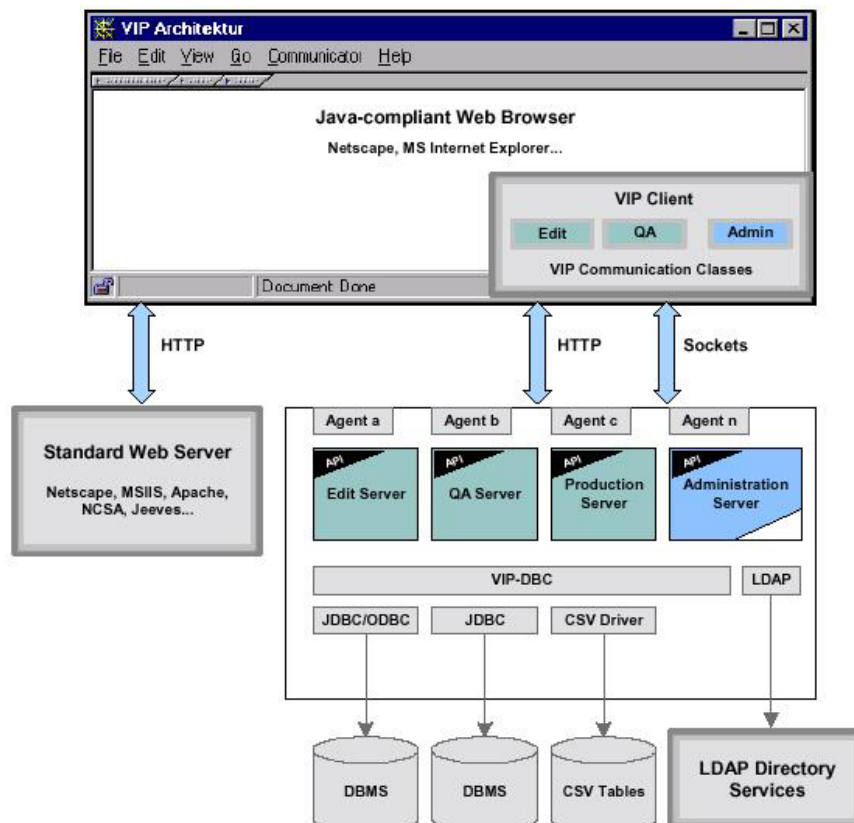


Figure 23. Overview of architecture of VIP'ContentManager

### **3.1.1.1. Database interface**

The VIP system accesses external data sources via an integrated database interface (VIP'DBC). All administrative data for the VIP system is filed via the database interface. This data may exist either in a database system (DBMS) - provided the DBMS offers a JDBC or ODBC driver - or as CSV tables.

### **3.1.1.2. Programming interface**

The programming interface (API) of the VIP servers is able to link server classes and thereby integrate agents in VIP'ContentManager. The entire functionality of the VIP system is available via the API. All functions that can be activated via the user interface are also accessible via the API.

### **3.1.1.3. LDAP**

The administration system of VIP'ContentManager offers the facility of integrating the user, group and role information of an LDAP directory service and using it in the VIP system. This permits a substantial reduction in administrative work on the configuration and maintenance of user data and improves the consistency of such data throughout the company.

### **3.1.1.4. Server processes**

In accordance with the workflow concept of VIP'ContentManager, the data for a website is managed simultaneously by the different VIP servers. These are separate server processes that can be started independently of each other. When a user logs in to one of the VIP systems, a client establishes a connection with the relevant server process.

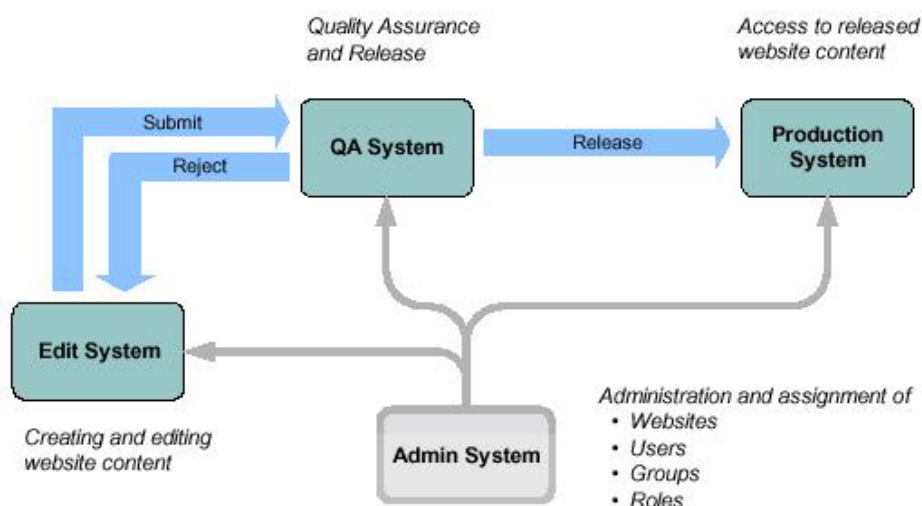


Figure 24. Server process tasks

- The Edit System manages all editorial activities in the creation and maintenance of website content.
- The QA System manages all data submitted to Quality Assurance.
- The Production System manages the “published” data for a website and “public access” to such data.
- The Administration System handles all administrative tasks such as creation of websites, user administration etc.

### 3.1.2. VIP Three-Server Concept

A VIP system consists of three separate server processes for editing, quality assurance and distribution of website contents. These are referred to as the Edit system, the QA system, and the Production system:

- The Edit system is used to create and edit the content (objects) of a website, such as HTML pages and graphics. This is where the website objects, such as HTML pages, are created and edited. After editing, the objects are submitted to Quality Assurance and thereby transferred to the QA system.
- In the QA system, the website objects are checked by qualified users for content and form. After verification, an object can be returned to the Edit system for correction (rejected), or transferred to the Production system for access by end users (released).
- The Production system allows end users to access the released pages of a website.

In addition to the three content management systems, there is an Administration system which handles all administrative tasks, including user and website administration. The Administration system is used to create new websites, and to create and assign users and user groups.

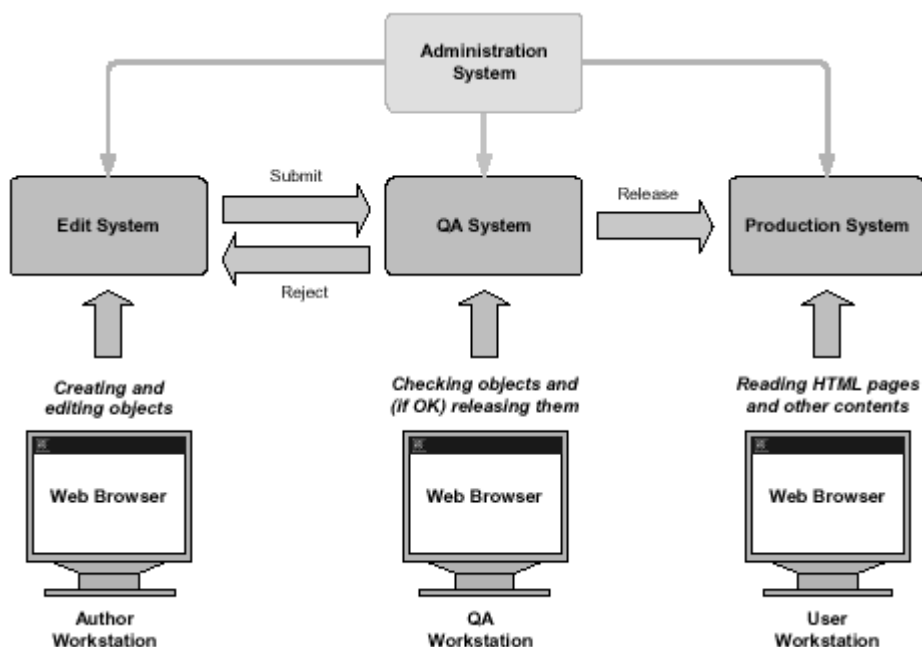


Figure 25. VIP Three-Servers Concept

To edit an HTML page, for example, the page is checked out from the web repository managed by VIP. While the object is checked out in the Edit system, it cannot be accessed for editing by other users. The page content is loaded for editing into the editor specified for this object type. After modification the page is returned to the web repository (checked in), and then submitted to the QA system. The user (or group) responsible for quality assurance is notified by e-mail that the object has been submitted. Quality Assurance then decides whether the page is to be released for publication, or whether it is to be sent back to the Edit system for correction (rejected). Releasing the HTML page transfers it to the Production system, making it available to visitors to the website. Note that these server processes reside on the same physical machine.

### 3.1.3. Structure of a VIP Object

Every VIP object has the following components, which contains the different types of data for the object:



- Content
- Template
- Metadata
- References
- Access rights

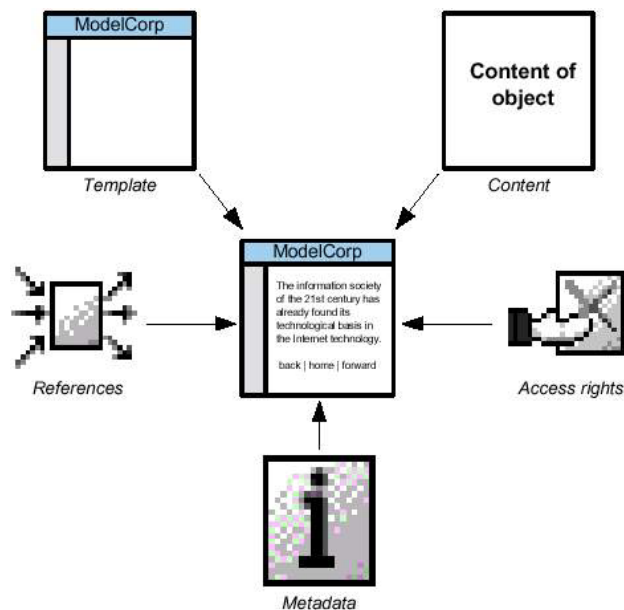


Figure 26. Structure of a VIP object

### 3.1.4. Creating an Object in VIP

There are 2 possibilities in creating an object: Creating a new object and Creating an object from an existing file.

For the first possibility, the user must select a template and fill in the metadata of the object, the object will then be created with the editor which has been configured by the administrator as the default for the object's extension, and will be saved automatically in a local directory on the client's computer. When it is checked in, the object will be uploaded to the vip server. If there are some images in the document the images are placed in a specific directory when it is uploaded. The user can override the default editor and select a different one.

As the second possibility, if the HTML object has been created locally, the uploading process will find all the files which have been defined as references in the object and

in this case a multi-import of these files is performed automatically. It is also possible to deselect the traced files. Uploading other type of object is basically copying file from client's computer into VIP server.

### **3.1.5. Editing an Object in VIP**

Before the user can edit a document the object must be checked out which means downloading the file into a VIP temporary directory on the client's computer. Afterwards the object is opened with an editor which has been set before in the preferences. If no edit program has yet been selected for a file type, a selection dialog appears to specify an edit program for the file type. After the editing process the object is then saved locally in the VIP temporary directory. In the checking-in process the object is uploaded to the server.

When an object is checked into the VIP system, a reference check is performed automatically. The content of the HTML file is compared with the metadata for the object.

The following items are checked:

- All the references contained in the object's metadata which must be present in the page (as hyperlinks or referenced graphics).
- All the hyperlinks and referenced graphics contained in the page must be present as references in the object's metadata.
- Page contains any references to files stored locally.

If this check finds that references are missing, the missing references are displayed in the results window "Check-in log".

The missing references are divided into two groups:

- The group "Following references are missing" displays references that exist in the metadata for the current object, but are not present as hyperlinks (or referenced graphics) in the content of the HTML page. This situation occurs if references are deleted while editing the VIP object.
- The group "Following references are new" displays references to VIP objects that were found in the current VIP object, but are not entered in the metadata for the

current VIP object. This situation occurs if references are inserted while editing the VIP object.

There are two ways of correcting these errors:

- The references can be corrected in the HTML editor, i.e. add all the hyperlinks missing in the object and delete all hyperlinks that are not also present in the metadata.
- Clicking the OK button in the results window (“Check-in log”) will close the window and check the object in to the VIP system. At the same time it automatically removes from the metadata all references that are not present in the object, and inserts all the references that are not present in the metadata.

When objects are checked into the VIP system, the automatic reference check also determines if the page contains references to files stored locally (e.g. hyperlinks or referenced graphics). If this is the case, these files are also incorporated in the VIP system (i.e. a multi-import is performed).

The Multi-import dialog lets the user specify which of these locally stored objects are to be included. It also allows user to move or rename individual objects, or replace them with objects that already exist in the VIP system. Accepting the multi-import will insert the selected objects in the VIP system along with the checked-out object.

### **3.1.6. Releasing or Rejecting an Object in VIP**

In the VIP, the function of Quality Assurance is defined as a control unit between the Edit system and the Production system. The task of Quality Assurance is to decide if an object meets the company’s quality standards or if it needs to be sent back to the Edit system for correction before it can be released.

The dialog window allows user to release the object, send an e-mail regarding the object's release, and add information to the content of the mail message. The default e-mail address displayed in the Recipient field is based on the e-mail address specified in the object's metadata, but may be changed or appended.

### 3.1.7. Setting Access Rights

Changing the access rights for an object can be done if permission to change access rights is granted. As with any other changes to a web object, changes to an object's access rights are subject to the quality assurance workflow, i.e. the changes must be submitted to Quality Assurance for checking and release. If Quality Assurance rejects the changes in access rights, the previous access control settings are restored.

To set a user's access rights for the current object, select the entry in the access authorization list, then click the desired access settings in the settings table.

In each row of the table the administrator can explicitly permit or prohibit the access type or leave it undefined. Access rights are also dependent on group-specific and role-specific entries.

Right	Yes	No
Read	✓	
Change	✓	
Delete	✓	
Create	✓	
Change metadata		✗
Release		✗
Change rights		✗
Tree operations		
Read (production)		

The access rights selected under "Yes" are **explicitly permitted** in this setting.

The access rights selected under "No" are **explicitly prohibited** in this setting.

The unselected access rights are **undefined** in this setting (and hence implicitly prohibited).

Figure 27. Settings options for individual access rights

Because access rights to an object may be defined differently for different groups, roles, and individual users priority rules are applied in the system:

1. Explicit prohibition takes priority over explicit permission.
2. Individual explicit permission overrides a general or group-specific / role-specific prohibition.
3. If access is neither explicitly prohibited nor permitted, an implicit ban applies which means access is not possible

Access to an object can be given to users, groups, and roles. User is an individual user, group is a collection of users classified by subject areas in organization structure (e.g. purchasing, production, sales, etc.), and role is a collection of users

classified by its function (e.g. author, QA, reader, etc.). In addition, a group-role setting can be applied in the scenario for example only “author+marketing” group-role can have a right to create object in a certain directory.

### **3.1.8. Inheritance of Access Rights**

The access settings for a directory are automatically passed on to the pages, subdirectories and other objects created under that directory. This means that the access rights can be set once for the main directories, and do not need to bother about the lower-level (child) objects within them. This transmission of access control settings to child objects is known as inheritance. Thus, the access settings for a directory are inherited by the objects created under that directory.

Because access rights are inherited by child objects, access rights for a website can be set up so that various user groups can have appropriate access to the directories and objects in the website. Every new object created inherits the access rights of the directory to which it belongs. If groups, roles or individual users access has been given to a certain directory, they will automatically have the same access rights to any new pages or subdirectories created within that directory.

Any change in an object’s access control settings breaks the chain of inheritance from the parent directory. If the object is a directory, a new chain of inheritance begins at this point. The objects created under this directory acquire the new, changed access rights.

If the access rights for a directory are subsequently changed, this change affects not only the directory itself, but also all the objects in the directory’s chain of inheritance. The change does not, however, have any impact on objects in a different chain of inheritance.

However changes in an object’s access rights can be cancelled and reintegrated the object in the chain of inheritance of the parent directory.

## **3.2. Architecture of VIP Solution in Hauni**

The following figure illustrates what other components are needed to be used with VIP system in order to completely fulfill the project requirements.

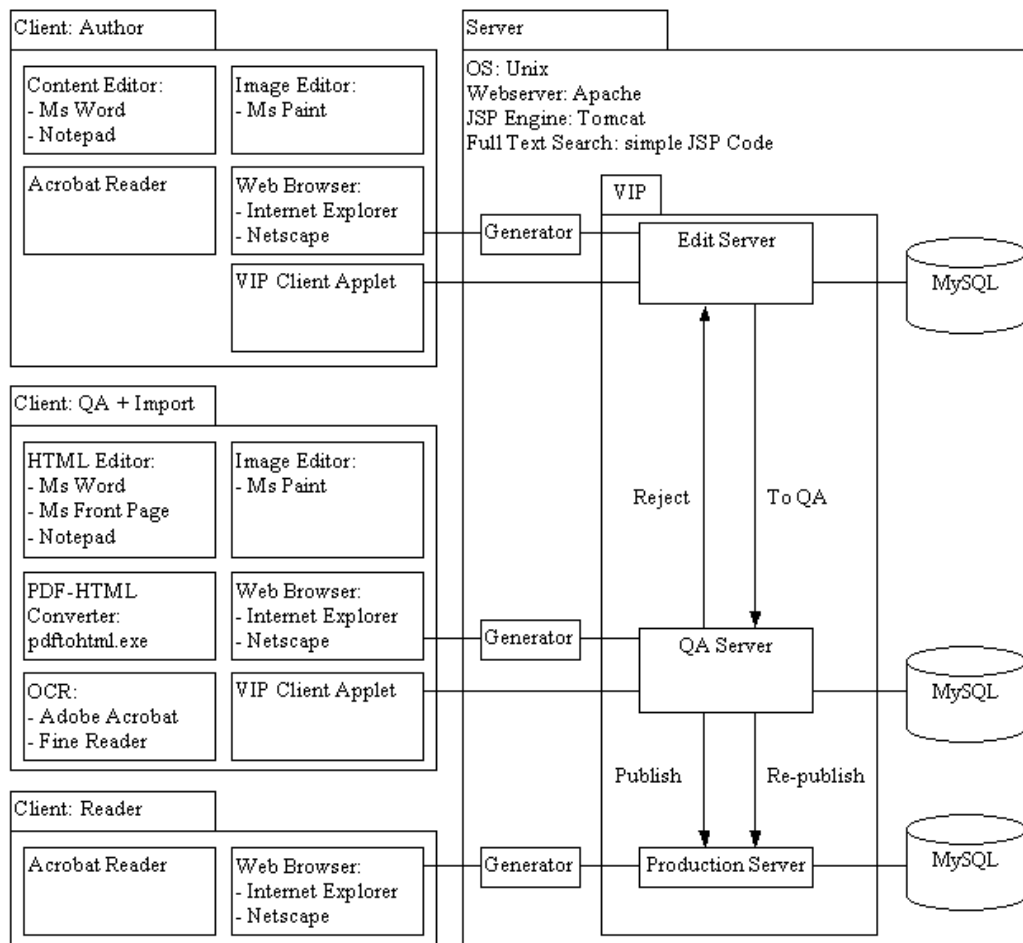


Figure 28. Project Architecture in VIP

Checking out in the VIP system means that the user copies the file into a local directory (c:\vip\temp) and edits this file with application that has been specified before. The application which will be used to edit the document can be changed in menu bar **applications>options**. The following applications will be used to create, edit, or access a certain document type:

1. Microsoft Word will handle standard and regulation documents with extension .html or .htm.
2. Microsoft Front Page will handle template documents with extension .html or .htm.
3. Notepad can be used to edit html documents. The use is limited for developers who understand html scripting.

4. Microsoft Paint as a standard installation in Microsoft Windows will be used to handle image files with extensions .gif or .jpg.
5. Internet Explorer and Netscape as web browsers will be used to access the site.
6. Fine Reader or Adobe Acrobat as OCR tool will be used to transform the first page of External Standard which is received from Beuth Verlag.
7. pdftohtml.exe and additional batch file will be used to transform pdf into html format and create necessary references in a file in order to enable multi-import in VIP system.
8. VIP Client Applet is the GUI which will be used by to access VIP environment.

In the server side flat file system is used for user database which will be replaced with MySQL. For full text search JSP code is being used to find a certain keyword in all html files. This search will be replaced in the future with httdig.

### **3.3. Roles and Rights**

Several roles are being created in order to give different responsibilities to the persons related to the management, creation, or access of the website.

Website Administrator will have all the rights as the last option to ask for solutions in case all others cannot help or are out of office for a period of time.

Document Administrator and Quality Assurance roles are assigned to one person in the Quality Assurance department since the workload is not high.

Author is technical person in every Business Unit which is related to the creation of internal regulation.

There are two kinds of Reader: Reader (extranet) who has the sufficient read access right for all documents in the website and Reader (customer) who can only access few documents. Reader (customer) has limited access to the documents because they contain important and sensitive information to Hauni competitors. Read access right will be managed by the role Document Administrator.

Roles Rights	Website Administrator	Document Administrator	Quality Assurance	Author	Reader (Extranet)	Reader (Customer)
Read	✓	✓	✓	✓		
Change	✓			✓		
Delete	✓			✓		
Create	✓			✓		
Change Metadata	✓	✓		✓		
Release	✓		✓			
Change Rights	✓	✓				
Tree Operations	✓	✓				
Read (production)	✓	✓	✓	✓	✓	✓

Figure 29. Website Roles and Rights

In order to increase the number of reader without adding more VIP license and to vary the rights of reader in accessing certain documents an additional user administration is needed. This administration will be created in a separated database different from VIP user database.

### 3.4. Directory Structure



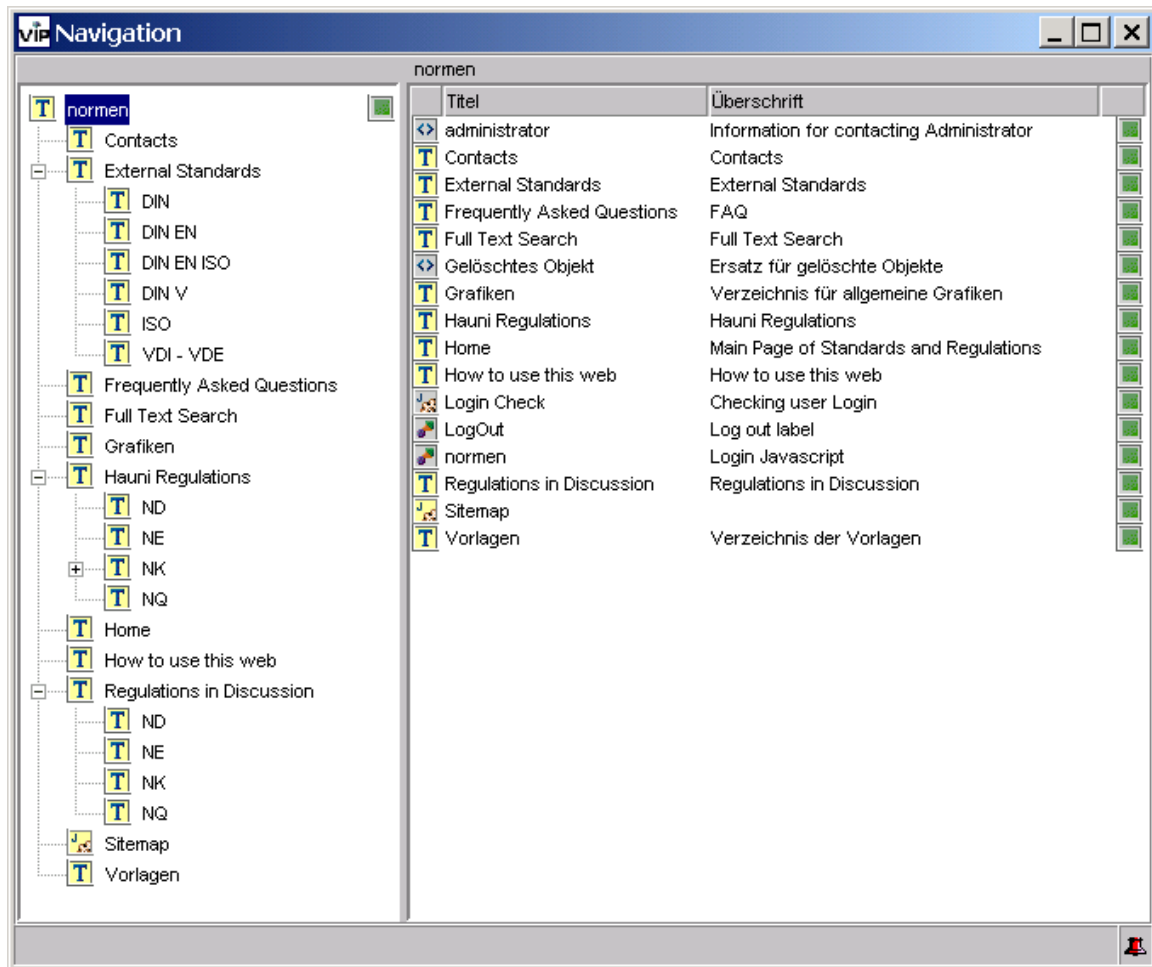


Figure 30. VIP GUI Applet for Navigation

All template files are placed in the directory **Vorlagen**. These templates are being used for the rest of the files in the website. Every template contains VIP tag `{VIPCONTENT}` to include the content of a document which is using the template. There are 5 template files which will be discussed as follows:

#### 1. Header1.jsp

Type: **JSP (template)**

Title: **Header1**

Heading: **Header1**

Template: **standard**

This template handles login form, checking read access right of a document, and document header. Result of the full text search is using this template which is cachable so that the page will not be processed twice if the user went back by pressing the **back** button to the result page of the search.

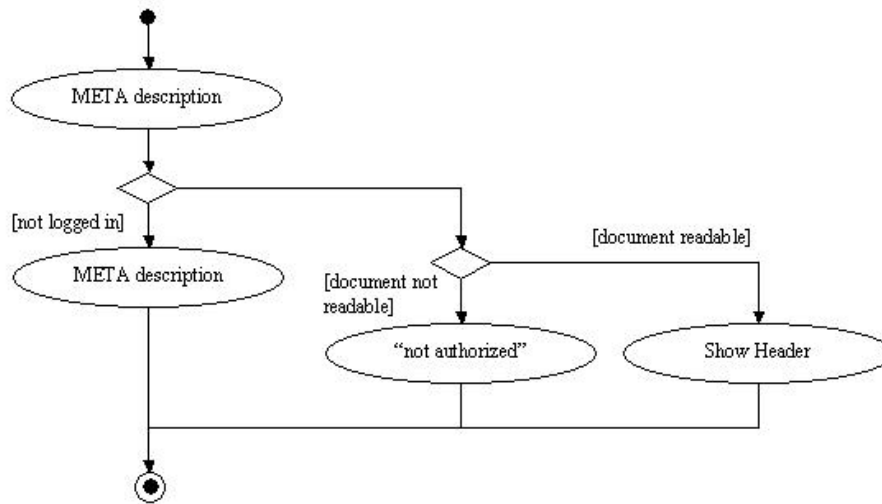


Figure 31. Activity Diagram of Header1.jsp

The header consists of hauni logo, log out button, topic hyperlinks, parent directory hyperlinks, and search form as follows.



Figure 32. Document Header

## 2. Header2.jsp

Type: **JSP (template)**

Title: **Header2**

Heading: **Header2**

Template: **Header1**

Basically the function of this template is the same as Header1.jsp. The difference is that any document which is using this template will not be cached.

## 3. Navigation1.jsp

Type: **JSP (template)**

Title: **Navigation1**

Heading: **Navigation1**

Template: **Header2**

This template constructs the middle part of the page. It lists folder/thema in which this template is being used. The list is created with JSP script which makes the

maintenance task easier. A new category of standard can be added by creating a new folder and the reference to this folder is created automatically by this template.

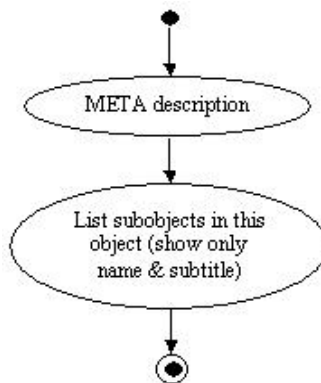


Figure 33. Activity Diagram of Navigation1.jsp

#### 4. Navigation2.jsp

Type: **JSP (template)**

Title: **Navigation2**

Heading: **Navigation2**

Template: **Header2**

Instead of showing only name and subtitle like in Navigation1.jsp, this template shows list of objects with more information (author, version, and creation date).

#### 5. Navigation3.jsp

Type: **JSP (template)**

Title: **Navigation3**

Heading: **Navigation3**

Template: **Header2**

Listing only pdf files with complete information (name, subtitle, author, version, and creation date). This template will be used in the external standards folder. There are 2 file formats (pdf and html) in this folder but since the html format is being used only for the full text search and consists only the first page any other format besides pdf will not be visible.

Index file in every folder is named automatically by VIP the same as folder name with extension htm.

### 3.5. Login Process

Request of login form in template Header1.jsp is sent to jsp file **Login Check.jsp**. The request contains 2 entries with a certain field name which are `userName` and `userPassword`. In fact The method `de.gauss.vip.portalmanager.SessionBean.checkLogin(request)` provides three possibilities for the login check:

1. the user and his password are stored in the request using the entries `userName` and `userPassword`
2. the HTTP Header in the request contains the entry "Authorization" with the user and password. Some JSP-Engines do not supply this entry.
3. the trusted-login is activated (possible since VIP5e.03). The VIP' Portal Manager logs-on the user without a password check. To activate this login, the following has to be appended at the users section or the section specified through `setLoginRepository` in the `repository.properties`:

TRUSTED\_LOGIN=true

UID\_ATTRIB=-user uid attribute in the `HTTPSession`-. This entry is necessary, if another application stores the uid in the `HTTPSession`.

This requires the Servlet API 2.2. note: Using this login, it is not possible to access to the `VipObjectHandlerRepository` or the `DirectoryRepository` or another `Repository` which needs a user authorization.

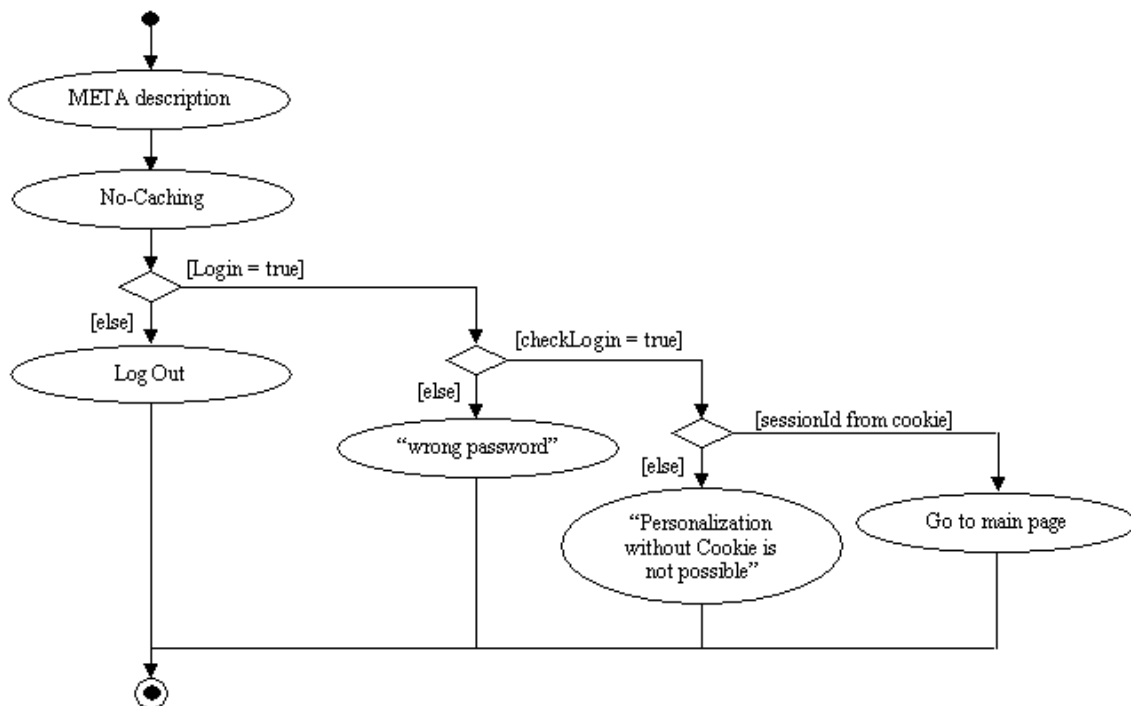


Figure 34. Activity Diagram for Check Login.jsp

### 3.6. Main Page

In the main page there is a list of 10 newly created or edited documents which is sorted descendingly by date, so that the newest document will appear at the the first row. The idea is to give a quick reference for the reader to the new documents since these documents will be accessed more often than the older ones. Information about its parents is shown in order to give more detail of the category (e.g. **External Standards > DIN > DIN 10085**). New documents can be a pdf file or an html file or even a new folder which contains a standard/regulation and its attachments. The document will not be shown if the user doesn't have a read access right to the document. The loop counter will not be increased in case there is a skipped document, so that the number of documents in the list is maintained to be 10 if possible. Every document in the list contains information about its name, title, author, version, and creation date.

### 3.7. Full Text Search

The search request form is placed in the header of all pages. This search has a different unique function. It searches a certain keyword in all html files only from the directory where the search form starts and all of its subdirectories. For example, if the user input the search keyword in the DIN page then the search proceeds to find all html files in the DIN directory and all of its subdirectories (including its subsubdirectories and so on). But if the search starts from the main page or the full text search page it will try to find all html files in the site which contains the search keyword. Not all documents can be visible to the user, it has to follow the rule of user right whether the user is authorized to read a specified document. PDF files cannot be searched directly. Section 3.8.1 describes how to extract information into an HTML file.

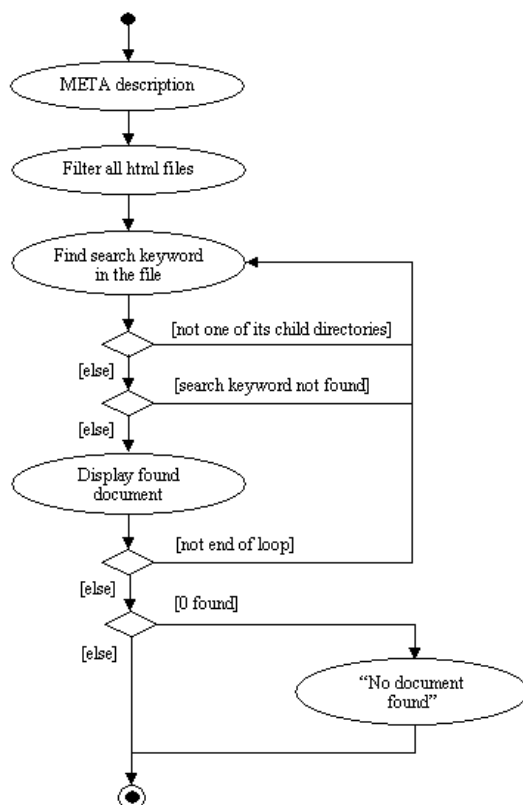


Figure 35. Activity Diagram for Full Text Search.jsp

## 3.8. Migration from Old System into VIP

This section describes migration procedures for external standard and internal regulation documents from the old system into the VIP.

### 3.8.1. External Standards

All of external standard files are in pdf format and placed on a certain directory in the old system. VIP ContentMiner which is needed to give interface to pdf files so that its content can be indexed and searched. However, this module will not be bought because of financial consideration. So a small jsp coded search is created to find the search words in the html files, and the first page of external standard file must be converted into html format. The conversion is using an open-source PDF-to-HTML converter [OD].

Uploading all the pdf and html files into VIP system at once is not possible since there is no reference between the files which is needed for multi-import. A small workaround with a batch file is required. List of the batch script is as follows:

```
echo "<html>" > link.htm
echo "<body>" >> link.htm
for %%x in (*.html) do echo "<a href=%%x>%%x</a><br>" >> link.htm
for %%x in (*.pdf) do echo "<a href=%%x>%%x</a><br>" >> link.htm
echo "</body>" >> link.htm
echo "</html>" >> link.htm
```

This script will create a file named link.htm which consists references to all pdf and html files which are going to be uploaded. Uploading only this file will trigger the multi-import of VIP to upload the rest of the files. This inconvenience has been complaint to VIP support and may be added in the next version. After the uploading process file link.htm can be deleted from the server.

### **3.8.2. Internal Regulations**

Interleaf application is used to document internal regulations. Interleaf can export its document format into html. After being converted the document can directly be uploaded into VIP system. Instead of uploading the regulation one by one the same scenario as uploading the external standard using the batch script is used. Multi-import will be done automatically after all the references are captured.

## Chapter 4. Implementation in CoreMedia

Implementation in CoreMedia will be divided into 2 types, web content management systems which is focusing in structured contents and document management systems which is focusing in management of documents. Before describing the implementation in both variants an overview of CoreMedia system will be explained in the first subsection.

### 4.1. The CoreMedia Application Platform by CoreMedia AG

This section gives an overview about CoreMedia Content Application Platform.

#### 4.1.1. CoreMedia Content Application Platform Architecture

As described in [CMDM01], The CoreMedia Content Application Platform (CoreMedia CAP) system mainly consists of a pair of servers (the production server and the live server), several editorial workstations (CoreMedia Editor), importers and exporters, and usually several web servers. The production server is used for creating and editing multimedia content, while the live server presents published content in the Internet.

The technology of CoreMedia CAP is based on open standards. All system services log on to the appropriate server via CORBA Middleware (at present, ORBacus from Object Oriented Concepts Inc. is being used). Typically, editors and importers log on to the production server in this way, while generators log on to the live server. Using a CORBA-based event model, all systems can communicate with each other. An action of a client which changes the repository can trigger freely definable actions of other clients. In this example, other actions can be triggered besides the actual publication action (e.g. static page generation or function calls of third-party systems such as CRM or personalisation tools).

Within CoreMedia, there is strict separation of content and layout. Content (multimedia documents) is stored in relational databases. Layouting is carried out through execution of Java Server Pages (JSP), in which content is linked with layout.



Layout elements can be added to the JSPs, or can be stored in the database as layout documents, which is suitable for more flexible site layout.

The databases are accessed from the production and publication sides by CoreMedia servers, each administrating its own document cache. Page generation occurs in generator modules (Dynamic Page Generator) which administrate documents and complete, generated HTML pages in caches. The generators are self-sufficient and, due to their caches, can be deployed as required: sites with low loads only require one generator, while sites with higher loads require several.

In the overall communication between the individual components, caching is used. This can be divided into three levels:

- The CoreMedia server holds results from database queries in a document cache, so as to avoid unnecessary database access.
- The dynamic page generator also saves documents in its own cache, so as to direct as few CORBA queries to the CoreMedia server as possible.
- The results of JSPs are generally complete HTML pages which are delivered to the Internet and held in the file system so that they can be delivered by the webserver without additional JSP execution.

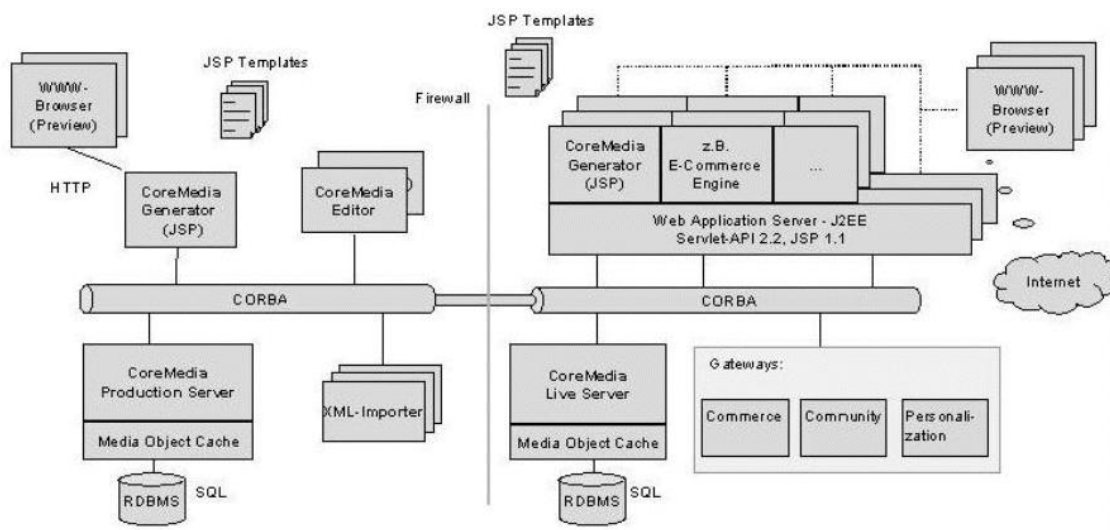


Figure 36. CM 3.1 System Architecture

The CoreMedia technical points are:

- CoreMedia corresponds to the 100% pure Java specification from Sun.

- All functional modules shown in the architecture diagram (see Figure 2.1a) can be distributed over different computers, since the modules communicate using event-based CORBA-APIs.
- CoreMedia integrates and consistently uses the following document management standards: XML, DOM, SGML and SAX.
- Through the use of JSPs, CoreMedia is J2EE conformant, so that the JSPs can be executed in web application servers in which e-commerce or ERP systems can be integrated using EJBs.

Documents can be structured as desired and are stored in relational structures, which the system creates automatically and administrates internally. A document interface to the outside is offered so that other database paradigms (object-oriented, XML-based) can be supported in future without the template interface having to be changed.

#### **4.1.2. CoreMedia Production and Live Servers Concept**

In CoreMedia a distinction is made between production and publication of content. Through the process of publication, content from the production database is transferred to the database of the so-called live server. There, the content is then available online. All operations on the databases are handled with CoreMedia software components, the CoreMedia Servers. Corresponding to the two databases, there are two CoreMedia Servers:

- the Production Server and
- the Live Server.

User works from the creation of documents to their approval on the production server. Not until the publication process are the folders and documents copied to the live server or updated there.

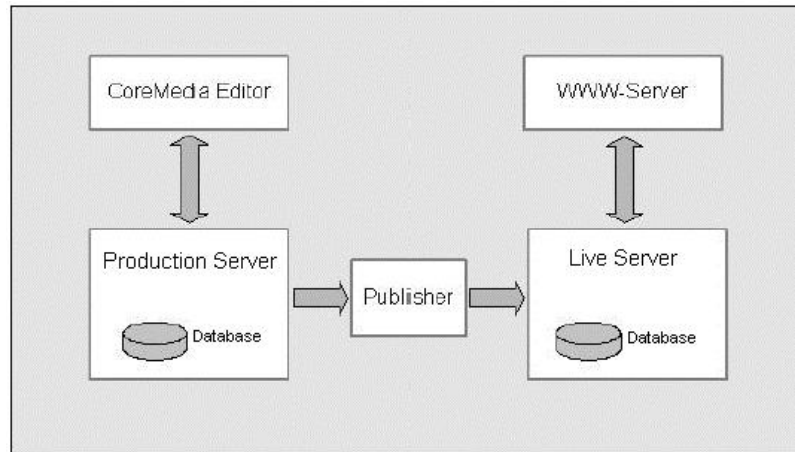


Figure 37. CoreMedia Production and Live Servers

When a document is successfully published, it is published on the live server. Only now is it visible for users of the website. However, it also remains on the production server, so that a new version can be prepared and published at any time.

Publishing means updating the state of the live server. Therefore the Publish command does not refer only to the actual publication of folders. Publication must also be used if documents on the live server have to be moved, renamed or deleted. The desired changes are carried out on the production system and replicated on the live server by a publication.

## 4.2. Web Content Management Systems implementation

Before describing the implementation, two definitions taken from [GC01] that illustrate our understanding of the term web content management systems will be given:

"Software-basiertes Content Management befasst sich mit der systematischen Sammlung und Verwaltung von Informationsbausteinen in einem einzigen (logischen) Bestand. Es stellt Anfragemethoden und Mechanismen für die sichere Arbeit ganzer Nutzergruppen mit diesem Inhaltsbestand ("Content Base") bereit."

This can be translated approximately as follows:

"Software-based content management is concerned with the systematic collection and administration of information components in a single (logical) inventory. It provides query methods and mechanisms for the secure work of whole user groups with this content base."

The second definition is as follows:

"Software-basiertes Content Management befasst sich mit der systematischen Sammlung, Erstellung, Speicherung und Veredelung von strukturierten Inhalten und Mediendaten aller Art in einem einzigen, fein granulierten (logischen) Bestand. Es unterstützt gezielt die sichere Aggregation, Veredelung, Verarbeitung, Auswertung und Wiederverwendung dieser Content Base durch ganze Benutzergruppen."

A translation would be:

"Software-based Content Management is concerned with the systematic collection, production, storage and refinement of structured contents and media data of all types in a single, fine granulated (logical) inventory. It supports secure aggregation, refinement, processing, evaluation and reuse of the Content Base by whole user groups."

The definitions emphasize the structured contents which are stored in a single, fine granulated (logical) existence and can be processed for further use afterwards. Therefore document structures which have been describe in section 2.3.1 for external standard and section 2.3.2 for internal regulation should be considered as a reference in the structure.

#### **4.2.1. Architecture of CoreMedia Web Content Management System Solution**

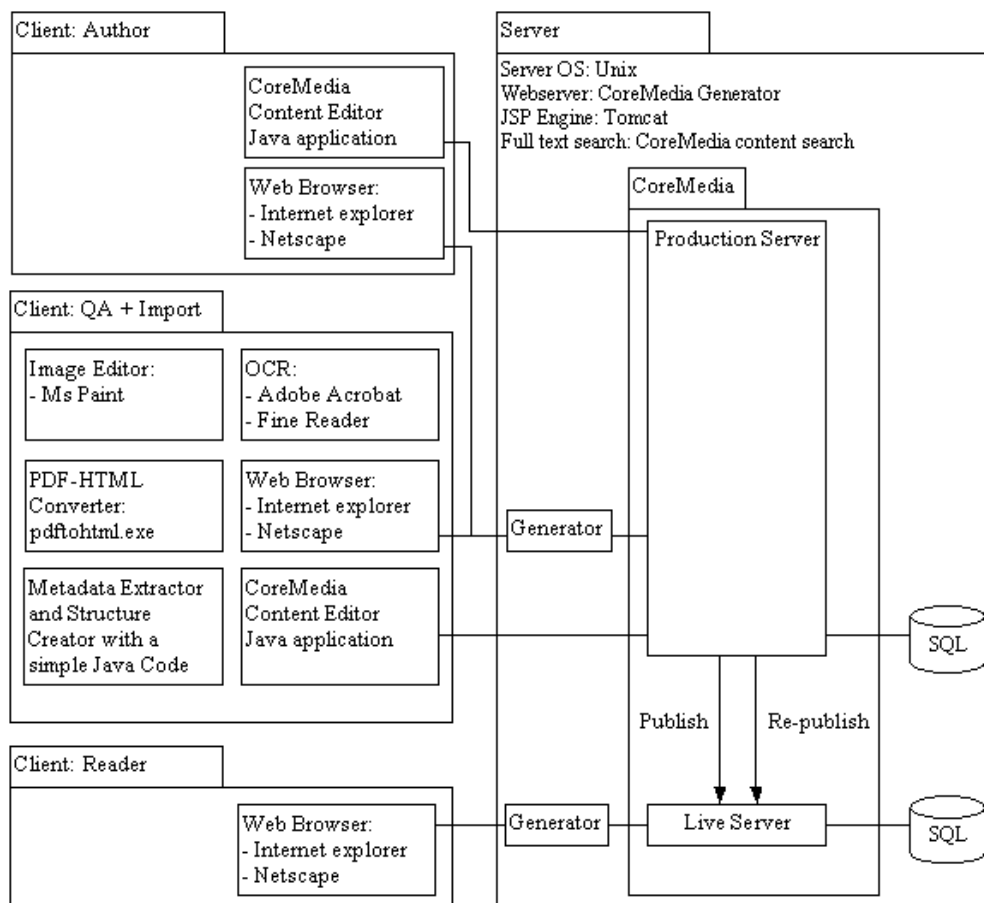


Figure 38. Project Architecture in CoreMedia as Web Content Management System

The focus of the architecture is on structured content which would get the most advantage of CoreMedia.

Previous external standard documents consist mostly of images and must also be OCR-ed with Adobe Acrobat plus OCR add-ons or Fine Reader package in order to extract the content.

Content from external source which is stored in different format (pdf) must be extracted with PDF-HTML converter and refined to correct some incorrect layout and text conversions. The correction effort would be very time-consuming since it depends on the open source program pdfhtml which does not perfectly convert pdf to html document.

#### 4.2.2. Roles and Rights

Website administrator has full access to the website and therefore will be assigned with all of the rights.

In this case the functions of quality assurance are checking the document layout by accessing the document which needs read access right, approving the document if the layout is created according to the guideline, publishing set of documents if there is no broken link, and finally granting rights to give reader (customer) a limited access to some of the documents.

Author as content contributor can read all documents published in the website in order to get inputs for a certain topic. Editing and updating document are also the functions of author and if the document reflects an old policy or never been used anymore then author can decide whether to delete or still keep it.

Reader in the extranet has a read access to all documents while customer has only limited read access to few documents. This is implemented using a separate mechanism, since customers are personalized in accessing the pages and do not contribute to the content of the website and it is not possible to integrate CoreMedia user administration on the template in order to create personalization in the website. Therefore an additional user administration is required.

Rights \ Roles	Website Administrator	Quality Assurance	Author	Reader (Extranet)	Reader (Customer)
Read	✓	✓	✓	✓	✓
Edit	✓		✓		
Delete	✓		✓		
Approve	✓	✓			
Publish	✓	✓			
Grant Rights	✓	✓			

Figure 39. Website Roles and Rights

### 4.2.3. Directory Structure

The structure of directory for this implementation does not differ from the structure in the concept which has been explained before in section 2.4.

#### **4.2.4. Additional User Administration**

Since CoreMedia access right cannot be implemented to restrict user to reading a certain object in the live server web page additional user administration will be used. Any web page which contains sensitive informations will be password protected. Therefore, a JSP template is created which shows a login page if the user has not been logged in, creates a session for the login user, and checks user read access right for a certain page.

Every reader to the web pages must be registered to this user administration. All business units can access all external standards and internal regulations in the website. Customer has a restricted read access to objects in the website. The object IDs (OIDs) of these objects must be listed as values of a field in customer setting. Manual input of these values will not take much time since only few customers can have the access to the website and access will be given only to documents which describe about the installation of the machine.

#### **4.2.5. Full Text Search**

Feature of full text search will be provided with the usual CoreMedia search since all contents are in text format. Since the contents are structured in the document, content search for a certain field can be implemented.

#### **4.2.6. Migration from Old System**

This section describes migration procedures for external standard and internal regulation documents from the old system into the CoreMedia as Web Content Management System.

#### **4.2.6.1. External Standards**

Previous external standard documents as pdf files contain text format only for the first page and the rest of it are images. If full text search is to work for external standards, these images must be OCR-ed and all the pages must be converted into html files.

There are two OCR tools owned by Hauni: Adobe Acrobat with OCR add-ons and Fine Reader. The result of Adobe Acrobat is stored in pdf format and converted to html with pdftohtml program. Microsoft Word doc is the result of Fine Reader and html format can be produced with Microsoft Word internal exporter.

Some refinements are required in the OCR process and in the conversion of pdf or doc into html format. Unrecognized characters, incorrect recognition, and minor text or table positioning in the layout occur mostly in the OCR process while incorrect table position, missing intra-file reference, and inappropriate image size occur in pdf to html conversion.

With the help of an additional java program metadata from the resulting html files could be extracted and the document could be structured in order to enable importing process in the XML importer.

#### **4.2.6.2. Internal Regulations**

Previous Internal regulation documents which are stored in Interleaf can be converted to html with its own exporter. Few refinement to the document is needed for some minor incorrect conversions such as intra-file reference with image and latin alphabet for technical variable.

The same procedure is done to extract metadata of the document and to create document structure with the java program.

### **4.3. Document Management Systems implementation**

In this implementation structure of content is not the primary concern but a document as a whole is more important to be managed (in the administration). One of the definition found to be close in the meaning [BKM92]:



"Das Dokumenten-Management hat die Aufgabe Dokumente zu archivieren, zu speichern, zu drucken, einzulesen, wiederzufinden und den Menschen bei der Bearbeitung, der Verwaltung, der Weitergabe und Ablage von Dokumenten zu unterstützen. Es hat zum Ziel, die Produktivität durch eine Verkürzung der Dokumentendurchlaufzeit und eine sofortige Bereitstellung notwendiger Informationen zu erhöhen."

And the translation is:

"The documents management has the task to archive, to save, to print, to read, and to find again the documents, and to support the people by editing, administering, circulating and filing of documents. The aim is to increase the productivity by shortening of the document processing time and an immediate response of necessary information."

### **4.3.1. Architecture of CoreMedia Document Management System Solution**

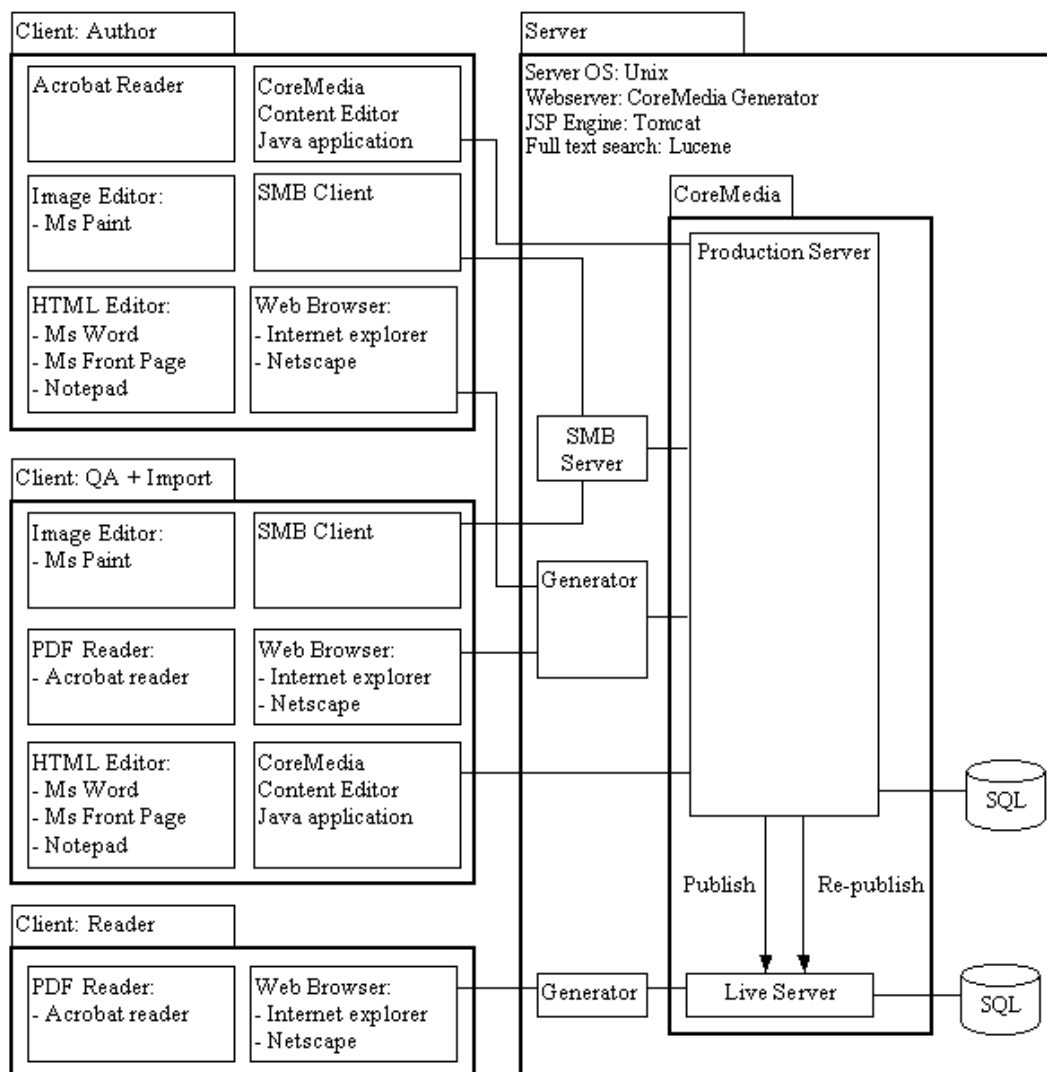


Figure 40. Project Architecture in CoreMedia as Document Management System

CoreMedia content editor is used only to administer the flow of documents. Creation of a new document and editing an old one will be using Server Message Block (SMB) protocol which is a Microsoft Windows de-facto standard protocol that lets programs make requests for files and services on remote computers over the Internet.

The author will use Microsoft Word as the editor to give contribution on the content of document while Quality Assurance (QA) will use Microsoft Front Page or even a simple one with notepad in Microsoft Windows environment to create and edit layout templates.

External standards which are in pdf format will be stored as files. These documents are not created internally by Hauni but bought from Beuth publisher instead.

Therefore editing this type of document will be prohibited. Importing this document using SMB will be done by QA who is also responsible as Importer.

### **4.3.2. Roles and Rights**

Composition of rights for each role will not be different from the one in the implementation of the web content management system which has been described before in section 4.2.2.

### **4.3.3. Directory Structure**

The structure of directory for this implementation does not differ from the structure in the concept which has been explained before in section 2.4.

### **4.3.4. Additional User Administration**

The same user administration which has been described in section 4.2.4 is used to restrict user for reading certain documents in the website.

### **4.3.5. Full Text Search**

A full text search which provides interfaces to other document formats such as pdf and MsWord as an external component must be used in this scenario since CoreMedia search can only be used for the sgml type in the structured document. In an approach similar to that described in section 3.8.1 can be used for extracting the contents of PDF files into HTML, and then using an HTML search engine to find the search keyword in the extracted contents.

### **4.3.6. Migration from Old System**

All documents are kept in the original format, external standards in the pdf format and internal regulations are exported from Interleaf document format to html format with the Interleaf exporter.

## Chapter 5. Comparison

The following subsections will explain some differences between CoreMedia and VIP system found in the implementation of the project.

### 5.1. Installation

For the server-side installation VIP Content Manager requires a Java Development Kit (JDK) and a web server as the primary components. Additionally a mail server will be required if an automatic notification to the user groups as a feature in the Edit and Quality Assurance is needed.

Client-side installation can be automatically done with the access to the start page of VIP Content Manager using a web browser at the first login with a correct password. This access always triggers a verification tasks whether the client has already a VIP GUI client and whether the current version is the latest one. In order to edit objects editors for certain types of objects must be installed in order to edit objects.

In the newest version of VIP, a web-based Content Manager is added and without an installation of client-side application an authorized user can use VIP features to a certain extend. An example of this limitation is changing metadata feature in the web-based Content Manager which is not supported.

A full-scale Installation of client-side of CoreMedia must be done manually in each client workplace or alternatively slightly limited web-based editor. Java Web Start has been developed which is used for automatic client-side installation.

CoreMedia Web-based Content Editor has some limitations such as not possible of handling user administration, there is no restricted read access on object which means by default readers can read all objects presented on the server, and no WYSIWYG feature in XML.

The automatic client-side application installation will reduce installation time at any place as long as an internet access is provided and authorization is granted for the access of the client on the server.

Zero installation with web-based application has been introduced by both systems but still with some limitations. These limitations can be overcome with additional programming or configuration which consumes more time and money.

Major issues in this development are increasing security of the access from the internet and capability to handle different behaviours of internet browsers from primary suppliers.

## **5.2. System Components**

VIP Content Manager does not provide any object editor such as HTML, PDF, or image editor. Installation of external editors is required in order to be able to edit the objects on a separated system but it provides references to run the editors and open the object. Standard setting of object editor references for in the client workstation which has been set by the server can be changed manually by the client and the settings are saved locally on clients workstation to be remembered for the next use.

A CoreMedia Content Editor can edit sgml type of content and handle limited functions of image editing. Editing image with external application can be done after the object is saved locally in the client workstation which means that the object is downloaded from the server to the file system of the client computer and after the edit the object must be uploaded back to the server. The same scenario is applied to edit microsoft word and adobe acrobat documents which must be saved as blob type. Alternatively SMB client-server application can be used to get direct access to the documents through internet and the CoreMedia Content Editor must be used to check in and out the document.

## **5.3. Server**

Three-Server concept of VIP is physically 1 server with 3 different directories which virtually behave for 3 different systems. These systems consists of edit system, quality assurance system, and production system. One released object results in having 3 identical objects stored in different directories. Submitting or releasing a new object means copying the file from one directory to another and if a new object is being submitted then the old file will be moved and versioned to a certain directory for back up.

In CoreMedia there are only 2 servers which are production server and live server. Approval in the production server means giving a certain flag to the object. Publishing the object will copy it into the live server.

## 5.4. Access Rights

The following table shows the different access rights between VIP and CoreMedia system.

VIP	VIP access right description	CoreMedia	CoreMedia access right description
Read	User is allowed to read the object	Read	User is allowed to read the object in both servers
Read (production)	User is allowed to read the object in the production system		
Create	User is allowed to create new objects within the topic.	Edit	User is allowed to create a new object, edit the object and do tree operations to the object
Edit	User is allowed to check out, edit and check in the object in the Edit system and to submit it to Quality Assurance		
Tree operations	User is allowed to copy the object and (in conjunction with the right to create) paste it in a different location within the object hierarchy. To move the object from its original location the user also requires edit rights		
Delete	User is allowed to delete the object	Delete	User is allowed to delete the object
		Approve	User is allowed to approve the object in the production

			server
Release	User is allowed to release this page in the QA system and thereby publish it into the Production system	Publish	User is allowed to publish the object into the live server
Change access rights	User is allowed to change the access rights to the object	Grant Rights	User is allowed to change the access rights to the object
Change metadata	The user is allowed to change the object's metadata. This permits changes in various properties of the object, e.g. the expiration date, the title or the template.		

Figure 41. Mapping VIP to CoreMedia access rights

VIP system provides more specialized access rights. These various access rights are usually used to maintain bigger website in order to increase efficiency with job specialization.

Both systems provide access rights inheritance which means that access settings for a directory are automatically passed to the objects and subdirectories under that directory. The difference is that changing access settings for the subdirectory in the VIP system will cancel the chain of inheritance from its parent directory.

The value of the access right in VIP system is a ternary value which is yes, no, or neutral.

VIP read access right can also be queried in JSP templates in order to make personalization in the website. A method in VIP API can verify user and password with VIP user database requested from the login page.

In CoreMedia the value of access rights is a binary value. User may access an object if s/he is a member of a group that has the right for the operation for the directory containing the file, or a parent directory thereof. Therefore, rights cannot be assigned to the individual user instead classifying users to a group and assigning the rights to the group.

Since the user management is not available in templates, CoreMedia access rights cannot be queried, forcing projects to implement or use a different authorization system.

## 5.5. Editing Object

Before user can edit an object in the VIP system, the whole object must first be downloaded into the client computer and stored temporarily in the default VIP directory. Checking the object in means moving the old object into the back up directory for versioning, transferring the edited object back to the server, and setting the new object as the latest version.

VIP system cannot edit the object with its own application, instead leaving the task to the external editor which can only open the object if the object is fully transferred to a local computer. VIP gives the references to the editors according to the extension of the object which to be opened.

Editing object in CoreMedia does not require the object to be downloaded into the client computer. The system can edit the document according to the field type in the document structure.

The procedures in VIP which take time compared to CoreMedia are saving the file to the local machine and starting external editor to edit the object. The cost is increasing as the number of external editors for a specific object type is increased.

## 5.6. Caching

Both VIP and CoreMedia systems provide cache function in order to accelerate multiple access to objects. However, in VIP the objects that are in the cache will not automatically be updated if the data on the server are changed, the user himself has to update the objects. This automatic update feature is supported by CoreMedia which is necessary for frequently changed objects.



## **Chapter 6. Summary and Conclusions**

External standard is a standardization document purchased from a content reseller, in this project the content reseller is Beuth Verlag. The document does not have an XML structure and is available in a certain application format.

Hauni is the end user of the copyrighted external standards and does not have any right to pass the document to the third party. Therefore, only Hauni user may access this document in the website.

Hauni business units are the content creators of their internal regulations. Access to these documents in the website can be given to the business units themselves and to selected customers.

Document structure for both external standard and internal regulation is less important because of the following reasons:

1. Hauni is not a content reseller but either the content author or content user/reader. Therefore, the number of transaction to export or import document to/from other content management systems is relatively lower.
2. The use of the structured document itself for Hauni is only to enable the documents to be searched categorically. The use of the categorized search is not frequent.
3. Because of the strict rule in the structured document, it is relatively more difficult to be constructed.

Therefore, the definition of document management system which has been described in section 4.3 before is more suitable for this project than that of a Web Content Management System.

Both VIP and CoreMedia system can handle Hauni external standards and internal regulations with a different degree of modification. VIP system has a direction rather to document management system application which leaves the task of document editing to the appropriate external editor and does not focus on the document structure. In contrast, CoreMedia focuses on the document structure which is more suitable for content reseller/broker.

## Appendix A. Project Description from Hauni in German

Normen und Richtlinien in der Tabaksparte

Allgemeiner Bestand:

Werknormen circa 300 Dokumente

Externe Normen circa 1000 Dokumente

### 1. Ist Stand

Normen und Richtlinien werden über Exchange dem Anwender angeboten. Für externe Anwender (Lieferanten, Tochterunternehmen) werden CD's erstellt. Normen werden mit Interleaf erstellt und dann zu pdf umgewandelt, anschließend im Netz dem Anwender zur Verfügung gestellt.

DIN - Normen werden per CD im tif Format geliefert und ebenfalls ins pdf Format umgewandelt, wobei die erste Seite eines Dokumentes (Thema, Nummer) mit einer Texterkennung bearbeitet wird, anschließend werden sie im Netz dem Anwender zur Verfügung gestellt.

- Doppelte Pflege auf Unix und NT
- Externe Verteilung aufwendig

### 2. Anforderungen

Verteilung der Normen und Richtlinien auf andere BU's, d.h. Auflösung der zentralen Funktion der Normenpflege. Veröffentlichung der circa 1300 Dokumente an einer zentralen Stelle im Hauni-Intranet. Idealerweise Umwandlung der 300 Werknormen ins HTML Format. Die Bearbeitung der Normen erfolgt vom zukünftigen Eigentümer (BU) in Word. Diese Dokumente werden ins HTML Format gewandelt oder als Datei eingebettet. Die Abt. 129 übernimmt nur noch Verwalter- und Layoutdienste und stellt die BU Dokumente zentral im Intranet zur Verfügung.

Die Normen sollen über mehrere Sortier- und Suchkriterien oder Gruppen (Zielgruppen, Themen, Schwerpunkte) zu finden sein. Die bisherige Volltextsuche muß weiterhin gewährleistet sein.

Normen, die auch von externer Stelle einzusehen sein müssen, könnten mittels Passwort über das Internet aufrufbar sein (Lieferanten, Tochterfirmen, TKD, Kunde

?). Dies würde die Verteilung von CD's ersetzen und die Daten vor Ort auf einen aktuelleren Stand bringen.

Die Dokumente werden zentral auch für andere BU's abgelegt, und für gleiche Themen werden keine unterschiedlichen Ablageorte vorhanden sein.

## Bibliography

- [BRJ99] G. Booch, J. Rumbauch, and I. Jacobson, "The Unified Modelling Language User Guide", Addison Wesley Longman, Inc., 1999, Massachusetts.
- [VIPAM01] Gauss Interprise AG, "VIP' ContentManager Administrator Manual Version 5e.04", 2001.
- [VIPUM01] Gauss Interprise AG, "VIP'ContentManager User Manual Version 5e.04", 2001.
- [VIPRM01] Gauss Interprise AG, "VIP'PortalManager VIP'ITF Reference Manual Version 5e.04", 2001.
- [VIPCM01] Gauss Interprise AG, "VIP'ContentMiner Manual Version 5e.04", 2001.
- [CMDM01] CoreMedia AG, "CoreMedia Content Application Platform Developer Manual", 2001, Hamburg.
- [CMUM01] CoreMedia AG, "CoreMedia Publisher 3.1 User Manual", 2001, Hamburg.
- [HOM01] H. Hommel, "Richtlinien Tabaksparte", Hauni Internal Regulation, 2001, Hamburg.
- [BUCH98] H. Buchholz, "Richtlinien-Erstellung Layout Vorlagen", Hauni Internal Regulation, 1998, Hamburg.
- [GUEN00] U. Günther, "Konzernmitgliedschaft im DIN", Hauni Internal Regulation, 2000, Hamburg.
- [BSW00] HJ. Bullinger, E. Schuster, and S. Wilhelm, "Content Management Systeme: Auswahlstrategien, Architekturen und Produkte ; Dokumentation", *Wirtschaftswoche*, 2000, Verlag-Gruppe Handelsblatt, Düsseldorf.
- [ISO] International Organization for Standardization, "Introduction", <http://www.iso.org>.
- [DIN] Deutsches Institut für Normung, "DIN", <http://www.en.din.de>.
- [TJP] The Jakarta Project, "Tomcat Documentation", <http://jakarta.apache.org>.
- [JT] Sun Microsystems, "Java™ Technology", <http://www.sun.com/java>.

- [JSP] Sun Microsystems, "Java Server Pages™ Dinamically Generated Web Content ", <http://java.sun.com/products/jsp>.
- [OD] G. Ovtcharov and R. Dorsch, "PDFTOHTML", <http://www.ra.informatik.uni-stuttgart.de/~gosho/pdftohtml>.
- [GC01] ROTHFUSS, Gunther; RIED, Christian: Content Management mit XML. Berlin ; Heidelberg ; New York : Springer-Verlag, 2001 (Xpert.press). - ISBN 3-540-66594-3.
- [BKM92] BULLINGER, Hans-Jörg ; KURZ, Eberhard ; MAYER, Renate: Zukunftsorientiertes Dokumenten-Mangement in Büro und Fertigung: Aufgaben, Nutzen und Trends. In: BULLINGER, H.-J. (Hrsg.): Dokumenten-Mangement. IAO-Forum 6. Oktober 1992. Berlin, Heidelberg, New York, London : Springer-Verlag, Oktober 1992 (Forschung und Praxis Band T 34). - ISBN 3-540-56146-3, S. 10-35.