231

# An integrated multiple media news portal

Caj Södergård [*,1], Matti Aaltonen [2], Christer Bäckström [1], Ari Heinonen [4], Timo Järvinen [1]
Timo Kinnunen [1], Pauliina Koivunen [3], Sari Lehtola [2], Ville Ollikainen [1], Katja Rentto [2],
Mikko Seppälä [4], Antti Tammela [2]

[1] *VTT Information Technology, P.O. Box 1203, FIN-02044 VTT, Finland*
[2] *VTT Information Technology, P.O. Box 1206, FIN-33101 Tampere, Finland*
[3] *University of Jyväskylä, Information Technology Research Institute, P.O. Box 35, FIN-
40351 Jyväskylä, Finland*
[4] *Journalism Research and Development Centre, FIN-33014 University of Tampere,
Finland*

[*] *Corresponding author*

**Abstract:** The emerging multiple media portals accessed by a variety of terminals
require semi- and fully automatic procedures for managing the material. The IMU
(Integrated Publishing in Multimedia Network) trial system, developed in this work,
automatizes parts of the news content acquirement and processing work of the portal
web master. The IMU active proxy server extracts the metadata from news web sites
and from the television news broadcasts through video analysis making an automatic
classification and linking of related articles and TV clips possible. The deeply
integrated material is partitioned into news composites called channels, which can
be personalised by the user. The automatically computed event and media calendar
allows for a new type of integration of news and up-coming events. The news
content is refined by setting up filters for monitoring of the business environment.
Through our interfaces for PC, TV, WAP and MP3 terminals, the user accesses the
same news content at work, at home in the living room and on the move. To balance
the automatic procedures with journalistic judgement, we created web tools for
human editors to control and override the automatic operations and for creating new
content. The community feature enables groups to share news and to discuss topics
internally. The trial including closely 400 users with PC, TV and WAP terminals
showed a stable interest in the service. The typical user retrieved a few fresh articles
including a TV clip at prime time in the evening. The television set user retrieved
two times more material than the PC user, but proportionally less news. The most
popular channels for the TV-user were TV programme schedules and TV clips. The
community channels attracted the TV set users. Personalization was used scarcely
and searches even more seldom. The interviews showed that the system was well
accepted, except for navigating in the TV-IMU application with the remote
controller.
**Keywords:** Portals, Integrated publishing; Multiple media publishing;
Personalization; Video analysis; Semiautomatic editing; Community
building; Environmental scanning

## 0. Introduction

Portals, like Yahoo[1] and MSN[2], are common entry ports into the WWW. To gain in
popularity, they are including more and more features like common and localised news,

---

[1] http://www.yahoo.com/
[2] http://www.msn.com/

search engines, chat, communities, calendars, personalization and e-commerce. Audio and video multimedia material are increasingly being added creating *multiple media portals* [1]. In this environment, content – even if primarily intended for a certain medium – is repurposed for a multiplicity of distribution channels and receiving terminals. A goal is, that the media consumer can get content versions independently of the terminal device at use at the moment, be it a PC, TV, WAP telephone or a PDA.

The management of these multiple media and personalisable portals poses a significant challenge. The editors and web masters must use a wide variety of semi- and fully automatic procedures. In a recent paper [2], we presented the IMU trial system[3] for integrating newspaper and television news on personal channels. In the extensive user studies of the IMU system, the value of this deep source integration was clearly recognised and a even broader integration with more – and fresh - information sources was wished for. Editorially prepared feature articles interested the readers significantly more than ordinary news and more editorial touch was in general hoped for. The television set was by many – especially the media house people – seen as the preferred terminal for this type of content.

In the second phase of the IMU work reported here, we have addressed these observations by including constantly updated news sources as well as an editorial system for human editors. A journalistic team was founded for doing the editorial work. The television set is included as a terminal on equal footage with the PC. Personalization is tried in a more professional application for scanning of the business environment. In addition, communities have been integrated creating a link between the large-scale news creation in major media houses and the small-sized content production and discussions in local communities.

Maintaining an event calendar is tedious job for the web master. In this trial, we have created an automatic procedure for setting up calendars.

## 1.  Principles behind our work

### 1.1   Personalisation of news, event and media calendar and environmental scanning

As described in [2], IMU offers a set of channels. The channel is a stored database query that in [3] is called a semantic bookmark. One channel type are *common channels* to which the reader can subscribe (Top News, 'the weather' etc). The coverage of these channels cannot be changed by the user. In addition to these fixed channels, the user can tailor so-called *personal channels* according to his or her interests. The reader creates personal channels by selecting one or more news or event and media sections from the ones made available by the IMU content producers. The reader can, for example, make a sports channel by combining the sports sections from one TV station and one newspaper. The articles covered by the channel can be further personalised by using a query clause. A third type is the *environmental scanning channel*. Environmental scanning is defined as the acquisition of information external to an organisation, the knowledge of which would assist management in planning future courses of action [4]. Strategic planning is closely connected to environmental scanning (e.g. see [4], [5], [6]), since adaptation to environmental demands and the ability of staying ahead of environmental challenges are

---

[3] http://www.vtt.fi/imu/

vital to an organisation's success. The user can, for example, select a 'technology environment' channel and personalise it by different search terms.

*Event and Media Calendar* automatically fetches from the supplements of two web newspapers. The material consists of television and radio programme schedules, movie theatre information, theatre and movie reviews, and information on different events ( music concerts, art exhibitions etc.). The television schedules are utilised via the TV interface to set the set-top-box to record television programmes.

## 1.2  Semiautomatic editing – the IMU editorial environment

Automatic selection sometimes results in an inadequate news offering. News items are placed in wrong departments, the main news are not always necessarily the most important ones, and links are incorrect. Therefore, we have developed an editing and information managing application for the IMU newsroom.

The IMU newsroom combines the benefits of automatic news generation and the professional expertise of journalists. The human editors administer the news flow, personalise thematic news channels and create news packages from the material fetched from our news sources (Fig. 1). The news packages offer a broader view on particular topics than the separate pieces of news that are scattered all over the IMU departments.
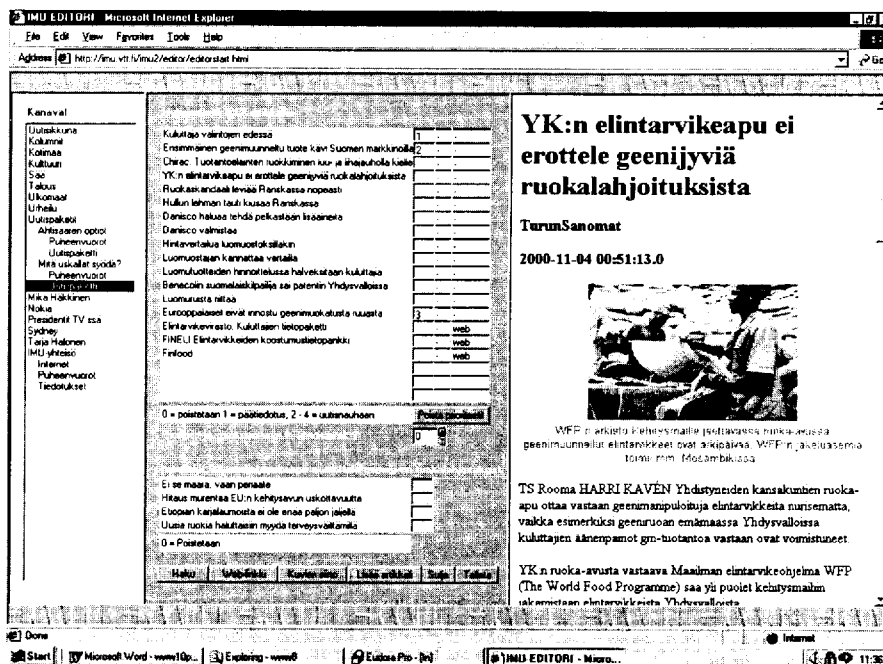


*Figure 1. The IMU editor, where a news package about food is under preparation. The package consists of an editorial introduction and links to articles in the IMU database (middle).*

News discussion forums are attached to the news packages. Thus, the audience can have their say regarding the daily news topics. The IMU editors take actively part in the discussion forums. The editors can always override the automatic choices and select the news for the main news slots (on the front page and news departments), delete repetitive news and delete incorrect links.

### 1.3   Communities

Communities link the large-scale news creation in major media houses and the small-sized content production in different communities. For the three communities in the trial we have set up communication spheres (channels) consisting of a bulletin area, a discussion forum and a personalised news channel. Some community members work as moderators producing material for their community channel. The IMU newsroom staff work in close co-operation with the moderators. The moderators use the same editing application as that in the IMU newsroom.

      The community members access their own dedicated channels through the IMU front page seeing therefore the daily news feed. The communities are also able to create their own personalised news channels. These community news channels present a filtered selection of all the IMU content.

### 2.   The functions and the user interface

The system functions and user interface were designed in a multistage process – starting from a brain storm with the project companies, proceeding through a requirement description, ending up in a user interface description that was implemented and put into the trial.

      As described in above, the IMU system combines related articles, video clips and media events into channels. Video articles, mostly TV news, and text-based articles are treated and presented in a similar way. There are three main areas on the front page (Fig. 2) (from the left) 1) *Channel selection area* where the user can select and modify the available channels. The main groups are News, Personal Channels (personalised channels and community channels) and Event and Media Calendar. 2) *News area* where the most topical news item is opened in the upper window; in the lower window are the headlines of other news items. 3*) News banner* with top news headlines, mostly with pictures. In addition to these areas, there is a *scrolling channel* – a sort of news ticker – in the uppermost part of the front page. From the front page the user can proceed to the various news departments and other services of IMU.

*Figure 2. The IMU publication, where the channels (common and personal) are presented on the left. The channel "Kotimaa" is selected. The channel contains the domestic news headlines and TV news drawn from several newspapers and TV stations. A News banner with the top headlines is displayed on the right.*

In the Event and Media Calendar the user can switch between four fixed channels: television and radio programme, movie theatre schedules, and events information (e.g. music, exhibition, and drama (Fig.3).

*Figure 3. The Event and Media Catalogue gives the user TV- and radio-programs, movie theatre offerings, and other event information.*

The PC and TV interfaces are very similar. The same information can be retrieved via the TV and the PC. Of course, the smaller resolution of the television screen affects the way in which the elements are shown. For example, the news stories cannot be scrolled as on the PC, and have therefore to be paginated. The remote control device has to be taken into account, when designing the TV interface (Fig. 4). Most of the commands should be given with the buttons on the remote control.
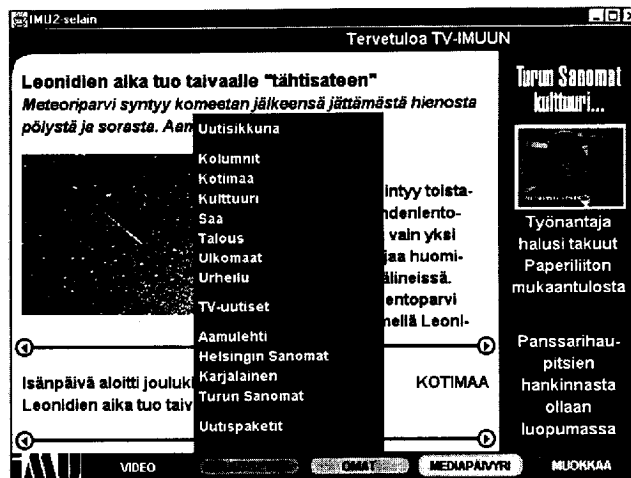
*Figure 4. The TV-IMU is used with a TV remote controller. The menu Common channels is selected with the remote control red button. The channel "Kotimaa" is selected with the arrow keys. The browser is written in Java.*

The WAP-IMU service provides personalised news for the users (Fig. 5). The metaphor behind the user interface is the channel. It is also possible to search the IMU news database with a WAP-IMU.
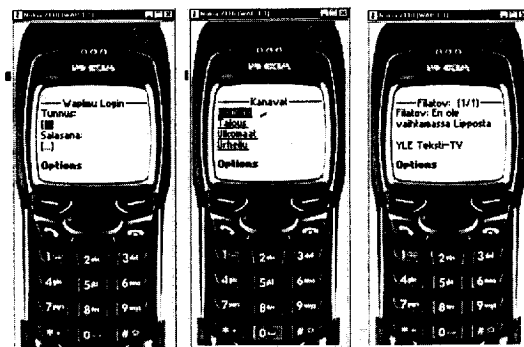


*Figure 5. The WAP-IMU publication, where the user at first logs in, selects a channel (Kotimaa), a heading and finally gets a short article.*

## 3. The system architecture

The three-tier software architecture implemented with Java is a natural, almost self-evident, choice for a modern web application. How to implement each tier is a more interesting question.

## 3.1   Client

In our case, where Java was chosen as the implementation language, Java applets and Java applications represent the fat client approach, and Java servlets represent the thin client solution.

### 3.1.1   PC

On the PC, we mainly developed a thin client. One servlet takes care of most user interaction tasks. All data needed for the user interface is kept in Java objects that remain active throughout the session. The model of the visible user interface is kept internally in a DOM (Domain Object Model)[4] tree structure. Normally each interaction from the user causes some parts of the DOM object to change, but as much as possible of the object is kept constant to reduce computing load. After each interaction the whole DOM tree is written out as an XML document.

The XML content is converted into HTML by the browser according to rules given in an XSLT style sheet attached to the XML file. The layout has been constructed using HTML tables in the style sheet. Separating the model handling from the graphical representation makes it much easier to generate HTML dynamically with servlets. Most of the user interface consists of HTML documents without frames. Only the scrolling channel is an applet. All selectable items cause a call to the interface servlet.

There is also a third alternative to the thin and fat client. Here, the user interface is implemented as a simple applet and the application logic is handled by a servlet that communicates with the user interface applet. This architecture has been used for the personalization wizard of the PC client. The communication between the personalization applets and the corresponding servlet is done using HTTP tunneling [7], where serialized Java objects are passed between the client and server using HTTP reducing firewall problems.

### 3.1.2   Other clients

The TV client uses a fat client architecture, where the browser is written as Java application. The WAP client is a scaled down version of the PC client. The biggest difference is that XSLT style sheets and applets cannot be used on the client side. WML documents must be generated by the server servlet directly.

The applet-servlet combination was used to implement the Editor depicted in Fig. 1. The Editor could have been implemented with servlets, but applets handled the complex user interactions better.

The Editor consists of several applets communicating with each other using a static class. One of the applets is active through the whole editing session. The other applets are invoked as needed.

---

[4] http://www.w3.org/DOM/

## 3.2   Server

### 3.2.1   Server platform

There are two NT server computers in the system: the media server and application server. The media server stores Microsoft ASF content to the file system and serves it by using Microsoft's streaming media server software. The server-side part of the IMU application runs completely in the application server. In addition to middle tier software, i.e. Microsoft's web server and JRun servlet engine, the database runs in the application server.

### 3.2.2   Middle tier

Despite the fact that the J2EE (Java 2 Enterprise Edition) technology[5] is quickly emerging as a platform for multitier enterprise applications, Java servlets were chosen as the middle tier technology. Java servlets have strong benefits in a research project.

### 3.2.3   Database

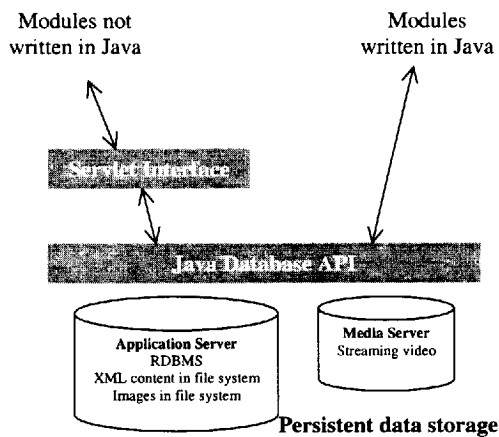The database consists of persistent data storage and database API (Fig. 6).



*Figure 6. Database architecture.*

The database uses XML only as the storage format of textual content of articles. Each article is stored in the file system as an XML file that conforms to the XMLNews-Story[6] DTD. Article, images and video clip, metadata, i.e. data needed for article selection and displaying article lists, is stored in the RDBMS. The database API, written in Java, retrieves data from the RDBMS and returns it to the client modules as Java objects that are instantiated from classes that represent IMU objects such as users, channels and articles. The Java objects contain URIs to images and videos. The textual content of the articles is passed as DOM objects.

---

[5] *http://java.sun.com/j2ee/*
[6] *http://www.xmlnews.org/*

## 3.3 Video segmentation

Television news broadcasts are segmented into stories. Even if the segmentation methods resemble those in the literature, see e.g. [8], we are applying them in a new way (Fig. 7). The appearance of the news anchor and various kinds of visual effects governs the setting of boundaries for a story (Fig. 8). A story is further segmented into separate scenes with associated stills determined from the first scene following the news anchor. The segmented news stories are labelled with news transcripts retrieved from closed caption texts.
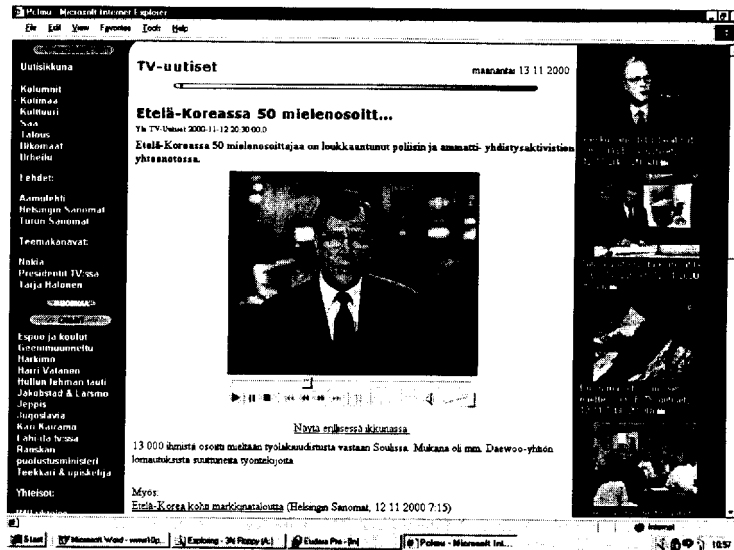


*Figure 7. The channel "TV news" is selected. The news transcript retrieved from closed captions is in the middle. The news topics found with video analysis are listed together with their representative stills at the right. Links to related newspaper articles are below.*
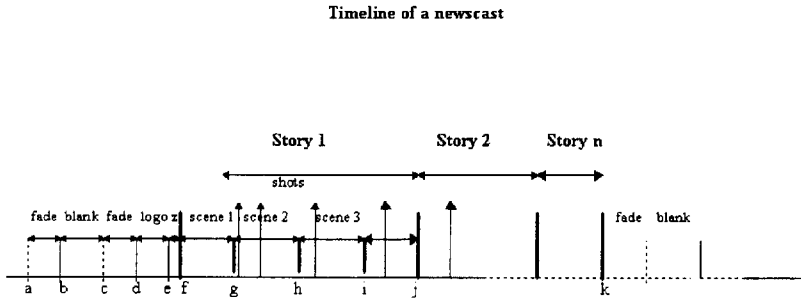
**Timeline of a newscast**



*Fig 8. Time representation of typical video segmentation. The above timeline represents events within the video. Fade-out is represented by a-b, blanks by b-c, fade-in by c-d. During d-e the news show logo is displayed. During e-f the news anchor is zoomed in. The news program and the first story begin at transition f. The periods f-g, g-h, h-i and i-j signify scenes in the first story. At j starts a new story usually found by similarity search. Event k represents the beginning of a fade-out, this also represents the end of the news program.*

## 3.4 TV Client

A PC-based Set Top Box (STB) was built to study, how an integrated media terminal – mixing computer and television - could be used as an integral part of everyday life in a living room. The STB has DVD playback, facilities to record and replay television programmes and an infrared remote controller, which complies with the NorDig specifications [9].

The televison schedules are used for interactive programming resembling the services available for digital recorders TiVo [10] and ReplayTv [11]. Television trailers, MPEG2 ML@MP compressed at up to 8 Mb/s, are. It is also possible to use STB as television tuner with an instant recording capability. STB has a Nicam [12] tuner to receive and record digitally transmitted stereo audio.

## 3.5 Environmental scanning

A telephone interview concerning strategies of business environment scanning was administered to randomly selected 150 Finnish SMEs employing more than 10 and less than 250 people in February 2000. The goal of the interview was to form topics of interest to the scanning wizard discussed above. Interviewees were asked how important it is to find information from different areas of the environment and how often they scanned the areas. Frequency of scanning is connected to the perceived importance of the issues scanned ([14], [15], [16]). The question format of frequency of scanning is derived from Hambrick's [13] study validated by Farh et al [17].

The environmental areas were named according to Daft et al [16] as competition, customer, technological, regulatory, economic and socio-cultural sectors. Natural environment was added as one sector (see e.g. [18]). New topics included were potential competitors [19], changes in customer's business area, suppliers [19], new

developments, changes or circumstances in the business area, environmental activists [20], natural environment [18]. Different environmental sectors were used as channels in the IMU-publication, and separate topics as search keywords in the scanning wizard.

The interviews showed that scanning was considered to be most important in competition, customer and technology sectors. The scanning wizard was based upon this.

## 4. The trial system

News articles are fetched daily from four on-line newspaper web-sites: Aamulehti, Helsingin Sanomat, Karjalainen and Turun Sanomat (Fig.9). Television news is digitized from Finnish Broadcasting Company (YLE) broadcasts. Up-to-the-minute news articles as well as teletext news are acquired from the YLE website. Teletext news provide shorter text-only versions of YLE news articles and are therefore suitable for WAP phones. Event information covers culture offerings in the cities of Helsinki and Tampere.

The content is viewed over fast networks on PC and TV platforms. In addition, WAP phones and voice synthesis devices are available terminals.
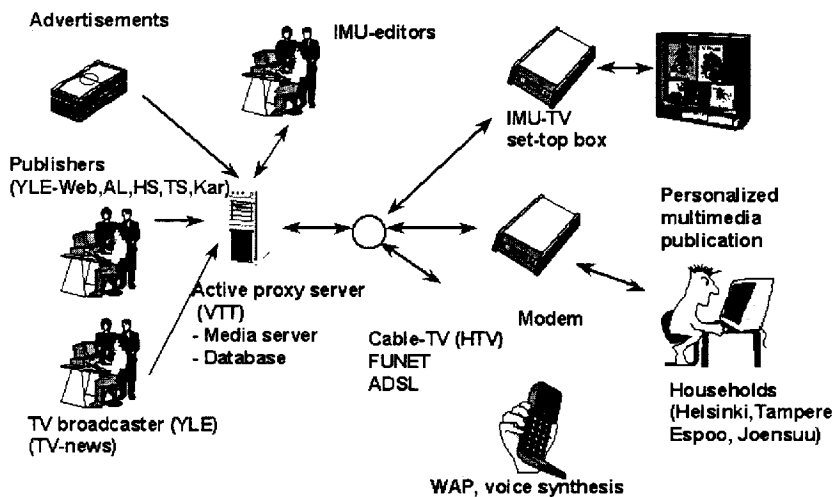


*Figure 9. The trial set-up, where the active proxy server processes the content from the media houses. The IMU editors control the machine-made choices and produce own news packages. The publication is used on different terminals.*

## 4.1   IMU2 Content Acquisition

The content acquisition not only fetches content from on-line publications, but it also extracts keywords and metadata from source content, generates news items from digitized TV news broadcasts and finally stores content and metadata into the database to allow a deep level of integration of different content items ( Fig. 10).
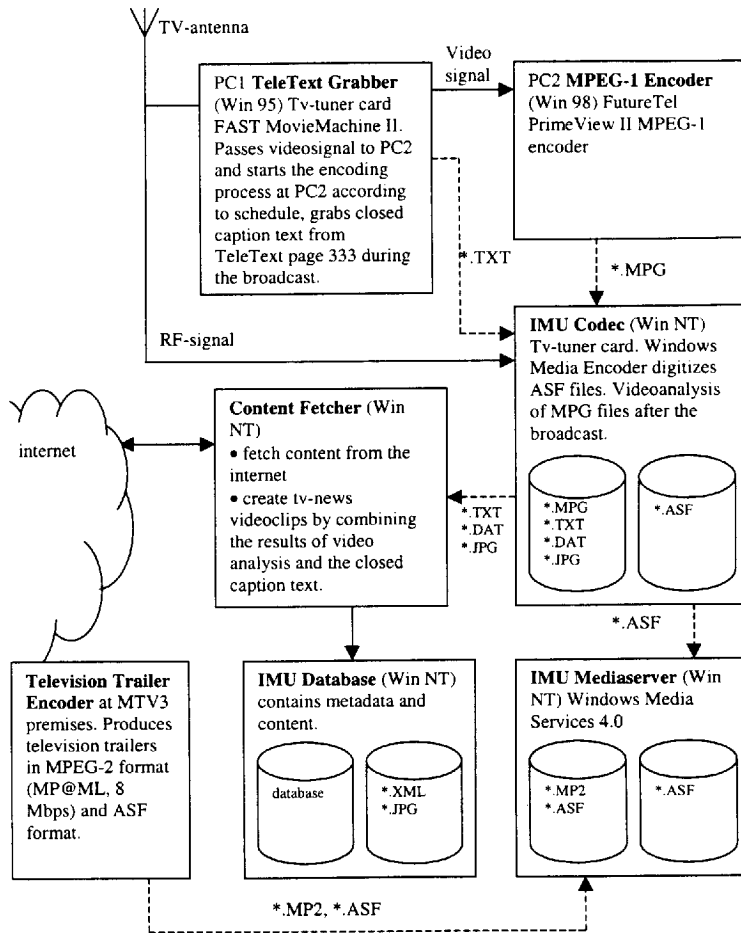
*Figure 10. The IMU Content Acquisition*

The fetched HTML-pages are parsed according to predefined rules coded into each different fetcher. Content data keeps the article ingress, bodytext, subtitles and caption, a JPEG- or GIF-file, which contains picture data and television news broadcast in Advanced Streaming Format[7]. Closed caption text with timestamps are grabbed from teletext page in real-time. This transcript of the speech in the broadcast is combined with video analysis to produce a video news article.

The fetchers are written in Java (JDK version 1.2.2)[8] and the HTML-parsing is done using Java Swing-classes. The Java Secure Socket Extension (JSSE 1.0.1)[9] and HTTPClient Java package written by Ronald Tschalär (HTTPClient Version

[7] http://www.microsoft.com/asf/
[8] http://www.javasoft.com/products/jdk/1.2/
[9] http://java.sun.com/products/jsse/

0.3-2)[10] is used. Other parts of the content acquisition system (Teletext Grabber, MPEG-1 Encoder and the Video Codec) have several "glue" applications written in C.

## 4.2 The TV-settopbox

### 4.2.1 Hardware and software

The user interface was made for 640x480 display. An infrared remote control reader was made and installed into STB case. On top of Windows operating system, there is a Java application for browsing the IMU pages (Fig. 11). Behind the browser window there is Visual C++ application, which takes care of remote controller control, video card controls, DVD and streaming video playback and programme trailer handling. The application receives commands from IMU server and the browser in a XML tagged file using a subset Louth VDCP commands [22].
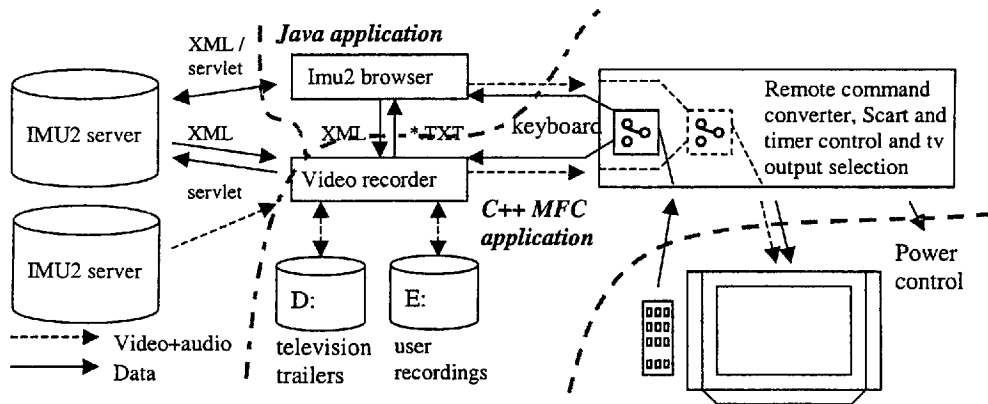


*Figure 11. Block diagram of IMU Set Top Box application software*

### 4.2.2 Using IMU Set Top Box

The STB has four different operating modes. In TV-IMU and IMU modes the browser is visible and at user's command. In TV-IMU there is also a small TV window on top of the browser showing the real time television programme. In the video mode the user can view his own recordings, television trailers, IMU news clips or DVD's. In television mode the STB serves as a television.

Commands, menus and links are adressed in a sequence by pushing the arrow buttons on the remote controller. The activation path starts at the advertisement in the upper right corner and goes through the news banner titles to the right to the opened article and the article list.

---

[10] http://www.innovation.ch/java/HTTPClient/

## 5.  User trials

### 5.1  Description of users and the usability evaluation

The trial system was tested in households in the Helsinki, Tampere, and Joensuu area
(Table 1). Of these 412 potential users, 360 has actively used the system.

*Table 1 . The user groups in the trial.*

| User group (potential users) | Size |
|---|---|
| Cabel modem users | 233 |
| Students using the Helsinki University of Technology network | 44 |
| Students using the University of Tampere network | 33 |
| ADSL users | 27 |
| Iidesranta community | 21 |
| Merihaka community | 12 |
| WAP users | 16 |
| Environment watching users | 10 |
| Journalist trial | 9 |
| TV-IMU-users | 12 |
| Total (IMU community) | 412 |

Most users represented single (n=147) or two (n=148) person households 278 of all
respondents reported not to have children.The users are a very "computerised": over 98
percent have a computer in their household and 99 percent uses Internet daily or almost
every day.

### 5.2  Usage patterns

During the research period (18.9.-3.12.2000) the PC users logged in IMU on average 11
times. The most active user logged in IMU 152 times. The sessions lasted on average 11
½ minutes. During each session the PC users chose on average 3,2 articles. During the
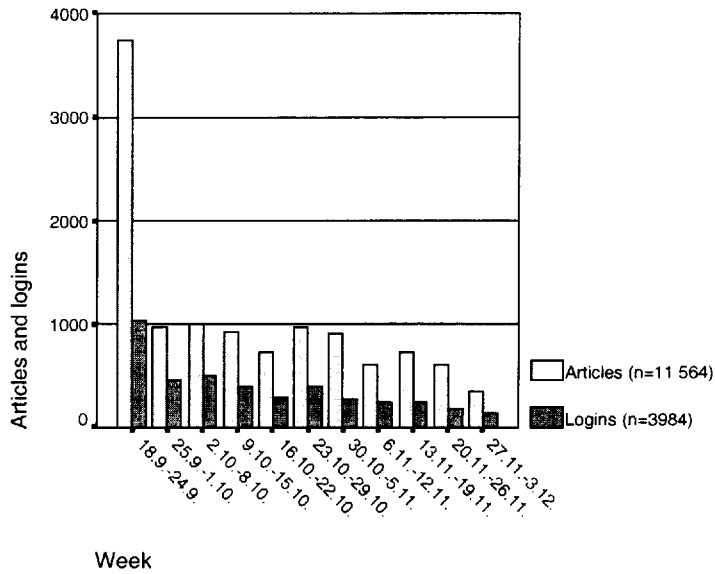first week IMU was in most active use (Fig 12).

*Figure 12. Amount of articles retrieved per week*

The IMU *prime time* for both PC and STB users was, like in the previous trial [2], between late afternoon and late evening hours (3 p.m. - midnight), which resembles the habits of television viewing. Morning use was more common to the STB users than the PC users.

By far, the most popular channel in IMU is television news (Fig. 13) with around 22 percent of all channels chosen. In the interviews, 65 % of the users wished for more television content.
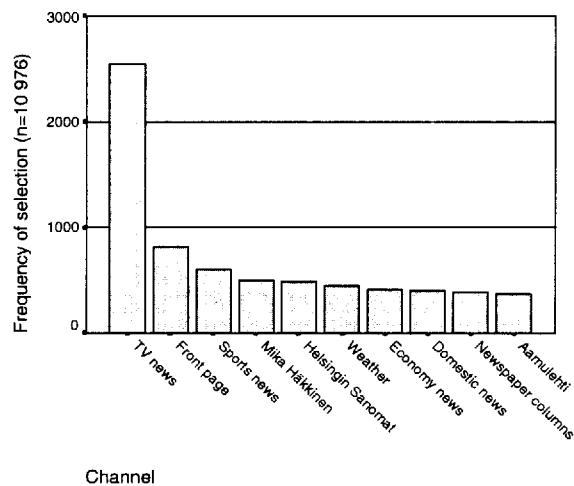
client architectures.

In spite of the merits of automatic procedures, journalistic skills and judgement cannot be overtaken by machines. Therefore, we have developed an applet based editor for journalists to control and override the automatically made operations. The journalistic team set up in the trial also created there own news packages.

The trial with PC and TV terminals showed a stable interest in the service. The typical user retrieved a few fresh articles at prime time in the evening. One fourth of the retrieved articles on PC was television news topics. The television set user selected two times more articles than the PC user, which supports our initial belief that the TV set is most suitable for this type of material. The most popular channels for the TV-user were TV programme schedules, TV clips coming in second. Fairly few used personalisation, but users were active. Searches were very scarce. News packages were used only marginally. The interviews showed that the system was well accepted, except for the navigating in the TV-IMU application with the remote controller.

The IMU media portal environment can be extended in various way. We are developing a system, where personalized news content through speech synthesis is converted into MP3 files that are downloaded on portable MP3 players. The main users of this system are visually impaired people.

## 7. Acknowledgements

## 8. References

R.G Picard., 2000. Changing Business Models of Online Content Services. The International Journal on Media Management, Vol. 2, pp. 60 – 68.

C. Södergård, M. Aaltonen, S. Hagman, M. Hiirsalmi, T.Järvinen, E.Kaasinen, T. Kinnunen, J. Kolari, J. Kunnas, A. Tammela, Integrated multimedia publishing; combining TV and newspaper content on personal channels. Proceedings of the 8th International World Wide Web Conference, Elsevier Science, Amsterdam, 1999, pp. 1111 – 1128.

S.Staab , J.Angele , S.Decker, M.Erdmann, A.Hotho, A.Maedche, H.-P.Schnurr, R.Studer , Y.Sure, Semantic Community Web Portals, Proceedings of the 9th International World Wide Web Conference, Elsevier Science, Amsterdam, 2000, pp. 473 – 491.

F. Aguilar, Scanning the Business Environment, Macmillan Co, New York, NY, 1967.

R. Beal, Competing Effectively: Environmental Scanning, Competitive Strategy, and Organizational Performance in Small Manufacturing Firms, Journal of Small Business Management, 2000, vol 38, no.1. pp. 27-57.

B. Boyd & E. Reuning-Elliot, A Measurement Model of Strategic Planning. Strategic Management Journal, 1998, vol 19, pp.181-192.

K. Moss, Java Servlets, McGraw-Hill, New York, 1998.