

PERSONALISED NEWS ON DEMAND: THE “HyNoDe” SERVICE

*The HyNoDe consortium**

Abstract

This paper presents the approach of the ESPRIT project HyNoDe (EP 22160 - Hypermedia News on Demand) towards a novel Personalised News-On-Demand (PNoD) Service. The HyNoDe Service Personalised feature refers to customisation of the service to suit individual needs and preferences of the clients based on the information filtering concept. The service addresses all the involved phases in the news publishing industry, from News Capturing up to end-user News Presentation. The paper compares the HyNoDe service with existing ones, and presents the architecture of the service, and the HyNoDe modules and tools related with each of the news publishing phases.

1. Introduction

First we introduce the notion of personalised news-on-demand and give a short overview on existing PNoD systems. The rest of the paper is structured as follows: Section 2 presents the architecture of the HyNoDe system. In Section 3, the News-oriented procedures of the HyNoDe system are described. Section 4 covers aspects related to Quality of Service (QoS) and 5 presents the system/service management and user-oriented procedures. Section 6 contains implementation details and section 7 concludes the paper.

1.1 Personalised News on Demand

The advanced capabilities of the emerging information and telecommunication technologies that trigger the decrease in cost of software, hardware and telecommunication charges, as well as the increasing importance of quality information in modern economies are fundamental drivers of multimedia services [Anal94]. The combination of real time communication with sufficient bandwidth, guaranteed by the evolving network architectures [Ovum95], creates opportunities for a wide range of residential as well as business services and promises a new era in information acquisition and delivery through the provision of on-demand and personalised services. The *on-demand* mode enables clients to schedule the viewing time of programs and manage the information presentation by pausing, resuming and in general controlling the navigation in the information space. The *personalised* feature refers to customisation of the service in order to suit the individual needs and preferences of the clients. Personalising the service to meet the clients interests is

* The HyNoDe Consortium partners are: INTRACOM SA, Hellenic Telecommunications and Electronics Industry - IBM Deutschland Informationssysteme GMBH - COSI - ETNOTEAM, Multimedia Systems Institute of Crete (MUSIC) - University of Ottawa - Il Sole 24 Ore RADIOCOR Agenzia Giornalistica Economico Finanziaria - IMERISSIA S.E.A - TH Darmstadt.

based on the information filtering concept, where intelligent mechanisms are used to suppress automatically the delivery of information that does not directly interest the user.

1.2 Related Approaches

A number of news publishers have already provided electronic access to their news using WWW browsers over the Internet. Nevertheless, the majority of the existing applications lack the personalised and on demand features that give added value to the service, the latter mainly due to the limited bandwidth of the current packet WAN networks.

In the area of personalised systems, MIT's FishWrap (running at <http://www.sfgate.com>) provides the user with textual news stories that are personalised through categories and a fixed keyword list. The selection of stories is made by full text search and the user's personal newspaper is created at logon time. IBM's infoSage (recently cancelled) extended this system by a free keyword search at any node of the category tree and an optional e-mail delivery instead of Web access. This service adds the news notification process to news-on-demand. In the business news domain, the pushing kind of delivery has been established as the rule. PointCast (found at <http://www.pointcast.com>) concentrates on the delivery approach. It delivers hypertext news, sometimes containing images, and runs on Windows only. at PointCast restricts personalization to channel selection. InfoComm (<http://www.infocomm.ca>) provides only short textual ticker news, but supports various media from computer through fax to pager. Grayfire (<http://www.grayfire.com>) delivers categorised articles through e-mail or fax. Farcast (found at <http://www.farcast.com>) offers e-mail delivery based on categories, but also full-text search operations both in an online and an off-line mode (the agents doing the latter are called „droids“). It is noteworthy that all of the dedicated business news services concentrate on the individual news story and do not offer multimedia features in their news systems. This also holds for the agent based search and filtering approaches of the artificial intelligence community, e.g. [Mitic94], [Kara94]. Brown et al. describe in [Brow95] the integration of information retrieval with multimedia by allowing free text searches on subtitles. They do not use a thesaurus in this case, but apply character stripping to compute relevance. They apply only a retrieval model, just like Hyon et al. in [Hyon96], who support attribute-based searching on documents that can be prepared in various formats. They consider also editing of these documents, but they assume slow off-line processing using a storyboard-based approach. For continuous media support in NoD systems, Georganas proposes a dedicated continuous media server that supports QoS negotiation and synchronization in [Geor97].

To relate the distributed multimedia issues of HyNoDe to other approaches, systems of other domains like video on demand or multimedia mail were considered. The GLASS project implemented the authoring and distribution of multimedia contents [Glas96]. This work heavily used the multimedia standard MHEG and helped to prove and improve MHEG's appropriateness for storing and communicating multimedia data in distributed systems. In [Glas96] a distributed multimedia authoring

environment is shown. It defines a proprietary authoring data representation and exports a textual MHEG notation. Within the BERKOM Multimedia Mail project (MMM), [Thim94] describes the dynamic composition of multimedia documents, controlled by the user's preferences in e.g. QoS or topic. Like HyNoDe, MMM has the notion of different user roles, but is oriented towards workgroups and computer experts in contrast to HyNoDe relying only on common users' expertise and end systems. Caloini et al. describe a news distribution system in [Calo96], which provides personalised views on documents by application of alternative styles that can be used with a document. Their approach is also used to handle errors. Martinez et al. use the concept of views into the hypermedia system [Mart96], too, with a support for non-unique addresses of referenced contents and follow-up documents. The CHIMP system [Selc96] introduces personalisation by using the concepts of filters. Filtering considers user access rights and end-system capabilities and provides user-dependent projections of each document.

2. The HyNoDe System

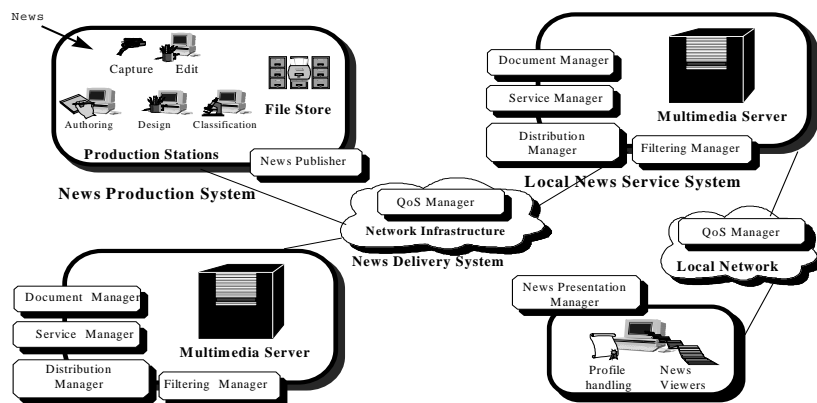


Figure 1. The Architecture of the HyNoDe System

The design goals of HyNoDe were mainly defined according the needs of professional news providers and 'professional news users'. This resulted in various constraints concerning delivery speed, target environments, user interfaces and system partitioning. The wide range of end user network connectivity options, ranging from dial-up modems to ATM leased lines, demands individually scaled news contents and system's knowledge about the users' connections. Despite just having PC type end systems, there is a wide variation in the power of the systems, which also influences the multimedia delivery to individual end users. The domain of news provisioning has time constraints, therefore a NoD system requires near real-time behaviour. Two content providers are among the partners of the HyNoDe consortium who have extensive experience both in the real-time and non-real-time news business. Their experience led to a system design for fast composition of news stories and delivery to the customers, while other parts of the system, which are not directly involved in the delivery of news, are not considered time-critical. The chapter 3 illustrates the HyNoDe system architecture following the path of a single news story Figure 1

depicts the architecture of the HyNoDe system. The basic components of it are closely related to the phases of production, storage, distribution, delivery and presentation of News Stories.

3. News Provision

This section presents the news life-cycle from collection through the News Production System up to delivery and presentation. The intermediate steps include content preparation, news authoring, filtering, news storage and retrieval, and news propagation through a hierarchy of servers.

3.1 Delivery Path

Content Preparation: A news story in the HyNoDe system is written using one of the tools that are common to the infrastructure of the news agency. The News Story is a multimedia document; it contains information in various formats such as text, hypertext, graphics, audio and video and also encapsulates any associations among them (e.g. synchronised play out of audio and video components). The author is required to move the news body (text, video, etc.) to a specific directory to make it available for further processing.

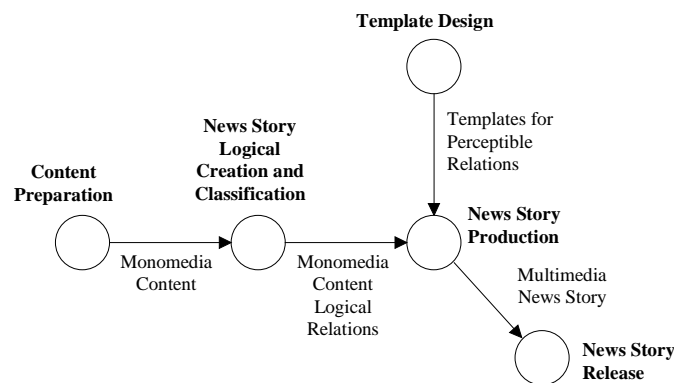


Figure 2: Data Flow Scheme in the News Authoring Process

Classification: Through this, the news story becomes available to the news editor. The editor is responsible for the classification of the story. The main issue of the editor's work, deciding the correctness and relevance of a story, is the same in the HyNoDe system. The editor is also responsible for making the logical connection between the central news story and additional information. In the HyNoDe framework, news stories are independent entities that do not require layout consideration or a positioning of the story with respect to other available stories. Rather, categories and keywords must be assigned to a news story, which are then used to decide whether the story fits to the profile of a consumer or not. To do this, the Media Relations Tool is developed within HyNoDe. At the end of this step, the news story, which is now attributed with category, keywords and additional content data, is passed on to the tool that assigns the layout.

Production: This layout tool is called the News Production (NP) Tool. The NP Tool allows the editor or the presentation designer to decide quickly for one layout from those templates that can present the news story and all the additional contents that have been assigned to the story by the editor. It is important for the content providers that no full-fledged multimedia editor is necessary for this step but that the connection of informative data and multimedia enhancement is made quickly. Although such a connection could also be made automatically on the fly, it was decided for the textual news stories of the HyNoDe system to involve human interaction. This is mainly because enforcing consistency of available templates and the selection made from these templates may not be accepted by the content providers at this stage. After the decision about the connection of template and contents, the NP Tool generates a multipart document for the transmission to the storage component.

Storage: The storage component is the Regional Multimedia Server, which stores all attributes of the news story, now aggregated in a news story descriptor, in a database and forwards this descriptor to the next component, the Filtering Manager. If that component determines that the story is relevant for at least one client, all of the files of the news story are copied into dedicated filesystems. While the Regional Multimedia Server is the only server involved in the current installations, the later installations of the system will be organized as a hierarchy of servers for more efficient distribution of the news stories, because in a real world situation the number of end customers of the HyNoDe system may become prohibitively large, and their geographical location is assumed to be arbitrary. In the hierarchical system, each server at the next lower level in the server hierarchy will appear to the Regional Multimedia Server as a client. Not all uploaded information is forwarded to the next server, but only the story descriptors. If the server at the next level decides from the set of known client profiles (a client may be a customer or another server) that the content is relevant, it requests the transmission from the Regional Multimedia Server.

Distribution: The Filtering Manager determines that a requested upload operation handled by the Multimedia Server was successful. In case that an active end-system is directly attached to the server, the Filtering Manager checks whether one or more customers interested in the news story are logged onto the system. Although the system is basically an unconnected service based on HTTP, so a connection monitoring application is running on each end-system indicating that a customer is currently using the service. This application maintains an open connection to the nearest server, through which notifications of newly arrived stories are forwarded.

Delivery: A notification indicates that the new story has been made available, and it provides a simple means (a hyperlink) to retrieve it. There is no fixed format for the notification, but an installation-defined template is available to construct the notification from the attributes of the incoming news story. This allows the decision about the amount of data to be made at installation time. In contrast to pure push systems, the customer is not only a consumer of the incoming news. News stories that are not immediately consumed are stored and not lost. A HyNoDe user can specify notifications for a subset of all interesting news by specifying multiple profiles.

3.2 Filtering and Searching

News Filtering

The HyNoDe News Filtering Manager automatically selects information based on a user's pre-recorded profile such as preferences of interests, with the main objective to eliminate news stories that do not directly interest the user. The filtering scenario of the HyNoDe system can be described as a set of independent operations which are depicted graphically in Figure 3.

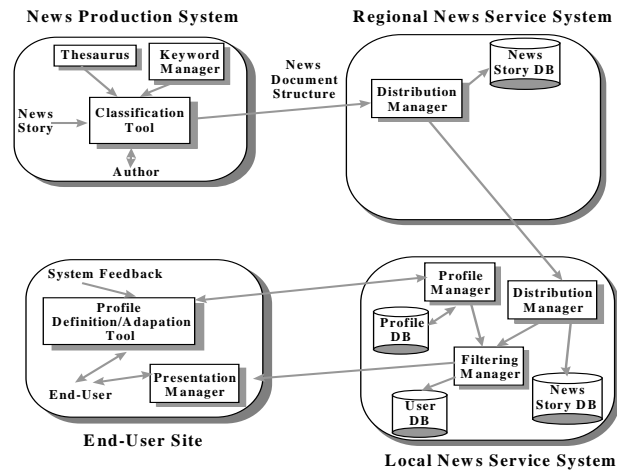


Figure 3. Filtering Process

At the end-user's site, the user formulates interactively a preference vector that reflects his current interests. He specifies his preferences about news authors, categories, language, content provider and a set of weighted keywords may be specified. This preference vector is then transferred to the Profile Manager at the Local NSS that the end-user is connected to and is stored in the Profile Database in order to be used during the filtering process.

At the news production phase, the Classification Tool generates the essential filtering information, which is included in the News Story Descriptor and in general is considered as the identification vector of the news story. Specifically, documents are characterised by a set of attributes provided by the news author which include category, language and content provider. In addition, a document is described with a set of keywords which is also provided by the news story author. For each keyword, the number of times the keyword appears in the document, as well as the number of times the keyword appears in all the documents of its collection, are used to determine the importance (weight) of the keyword in the collection (category or sub-category).

The above suggest that a dictionary for keywords (handled by a Keyword Manager), exist for every information provider, language and sub-category. The system maintains statistics for the number of occurrences of each keyword in the documents of every collection. These statistics are updated as new documents are coming into the

system in order to be able to compute similarities. Finally, a thesaurus mechanism is associated with the dictionary. It allows the news author to find synonyms for infrequently appearing keywords, and to find specialisation for very frequently appearing keywords.

This information is then communicated through the Distribution Manager to the corresponding Regional NSS and on a similar way to the Filtering Manager of the Local NSSs.

As soon as a new story arrives at the Local NSS, the filtering mechanism is triggered. For every news story published, the Filtering Manager matches the identification vector associated with it with the preference vectors of the clients and assigns a similarity measure to each piece of information. If the similarity measure estimated is within the correlation bounds defined in the specific profile, then the document is considered to be interesting for the client. In such a case the list of unread news stories of interest of the user is updated to include the new news story. If the user is connected to the system and a preference exists in his profile stating his interest in real time notification of arrival of important news stories, a message is sent to the Presentation Manager of the end-user, triggering a notification to the end-user about this new news story.

News Searching

All information needed to execute the search is stored in the descriptors. Due to the fact that descriptors are copied into the descriptor database of each local server, searching can be done locally. As a search result the user gets a list of matching news, sorted by a ranking value. The matching news are not necessarily available on the local server. This fact is visualised by using a traffic light paradigm. A flag with the colour red, yellow or green at each matching news describes the quality of service which may be achieved when accessing the news. Green means that the nearest copy of the news is located either on the local server or on another server in the upper server hierarchy which supports guaranteed QoS for the connection to the user. A yellow flag indicates that the news may be located on a system where the connection to the user may provide an acceptable throughput to transfer the news but there is no guaranteed QoS for the connection to guarantee a smooth replay of continuous media. A red flag indicates that the news has first to be downloaded onto the local server before it is accessible for the user. This happens when the news is only stored in the archive.

HyNoDe provides two search interfaces. One very simple interface uses HTML forms. The user can fill in some simple search parameters to create a standard query. The receiving CGI script transforms the form parameters into a search query conforming to the HyNoDe search language. More elaborated queries can be created using a search language. The language consists of any number of triples combined by the boolean operators AND and OR. A triple consists of a matching condition, a flag, and a value. The matching condition names one of the attributes, a relation (is equal, is unequal, is smaller, is greater, ...) and a value. If a triple is marked „mandatory“, the condition

has to be valid for a news to be selected. If the flag is „optional“ this triple is only used for the ranking part of the search query which follows the selection part.

4. Management

4.1 HyNoDe QoS Manager

The HyNoDe project invests special effort to serve the quality of service requirements of the NoD service. The QoS Manager gives emphasis to the *Application Level QoS* since there are constraints to the realisation of a complete *Network Level QoS Manager*. Most constraints are imposed by the capabilities of the underlying communication protocols and the existing network infrastructure.

The HyNoDe *Application Level QoS* is based on the concept of the *alternative media*. A HyNoDe News Story may have for each monomedia component more than one monomedia alternatives at different quality levels. A News Story is delivered to the end-user with the best possible quality based on the available media alternatives in the news story, the end system capabilities, the user preferences, and the network connection capabilities.

For each of the monomedia components, a list of alternatives is kept in the *Descriptor* of the HyNoDe News Story. In addition, the QoS Manager has access to information about the hardware and software capabilities of the client's terminal as well as his/her preferences for the media types he wishes to be presented with. In this way, the QoS Manager decides which of the available alternatives match the user's needs and preferences best and delivers an appropriate news story instance, the *on demand* view of the story for the particular client. In this way, the application can flexibly serve requests for limited delivery delay (interactivity level), duration of the overall presentation and preferred presentation format.

Regarding the Network Level QoS, for each of the media components, the requirements for real-time transfer and sufficient network bandwidth are examined and serviced appropriately. Based on the media types supported, the QoS Manager decides the network interfaces and resources it should allocate to deliver the new story to the client. In this way, continuous data can be downloaded to the client using RTP/UDP whereas discrete data can be delivered using TCP/IP over low-speed reliable channels.

4.2 System/Service Management

Some additional issues are considered to be important for the successful commercialization of the HyNoDe News on Demand application. These issues are the management of the HyNoDe system and the management of the NoD Service offered.

The HyNoDe system management monitors and controls the access and the status of the system. This includes status/performance monitoring, system configuration, network management (connectivity control, and network configuration), fault detection, and error/crash recovery.

On the other hand, the HyNoDe service management monitors and controls all the services offered by the system. This includes: access control, server security, copyright protection, subscription processing, accounting and billing. Electronic payment methods had been considered but postponed until a standard in this area has been settled.

5. Conclusions & Future Plans

Innovative information services with strong need of multimedia content, as the Personalised News-on-Demand Service, prove to be a driving force for the deployment of Advanced Communication Infrastructures. Key components of the Service are the NPS, the NSS and the NDS. Within the NSS, the Filtering manager adds an important new asset, compared to other existing News Services.

ESPRIT Project HyNoDe (Hypermedia News on Demand) aims to design and develop an efficient and market-oriented PNoD service. This will be achieved through the adaptation and integration of existing IT components needed in the phases of news authoring, storage and delivery. Appropriate mechanisms will be developed enabling news filtering according to each end-user profile.

An Initial prototype implementation of the HyNoDe system has been completed and a trial phase with real world actors is expected to begin before the end of June 1997. Consequently, it is expected that the presentation of this work at IDMS97 will also include the evaluation results of the trial phase.

The programming languages used in the implementation are C++ and Java. The database for storing the Descriptors of the News Stories, the user data, the user profiles and the accounting info has been built using the *Relational Database Management System DB2 2.1.2* from IBM, while all the database accessing functionality has been implemented using the *ODBC* interface. The HyNoDe Server operates on Solaris 2.5.x and AIX 4.1.4 (or later) platforms while the clients need to be capable of running Java-enabled browsers.

The following components will be added in the future: A profile optimiser will support the user in the management of his profiles. This component will register and evaluate the users behaviour to propose modifications of the profiles with the goal to adapt the profiles to the changing interests of the user. Additionally to the simple and the enhanced interface to searching a third one will be implemented. It will provide a graphical means to formulate complex search queries.

References

- [Anal94] Multimedia in Telecoms, Simon Norris, Susan Ablett, Analysys Publications 1994.
- [Brow95] Automatic Content-based Retrieval of Broadcast News, M. G. Brown et al., ACM Multimedia 95, Nov. 5-9, 1995, San Francisco
- [Calo96] Extending Styles to Hypermedia, A. Caloini et al., Proceedings of ICMCS 96, pp. 417-424

- [Geor97] Multimedia Applications Development: Experiences, N. D. Georganas, Journal of Multimedia Tools and Applications, Vol. 4, No. 3, May 1997
- [Glas95] GLASS: A Distributed MHEG-Based Multimedia System, Helmut Cossmann et al., COST 237 Proceedings, 1995
- [Glas96] GLASS-Studio: An Open Authoring Environment for Distributed Multimedia Applications, Brian Heumann et al., IDMS'96 Proceedings, 1996
- [Html96] HTML 3.2 DTD
- [Hyon96] A Web-based Archive and Retrieval System for Multimedia Production, J. Hyon et al., Proceedings of SPIE Multimedia Storage and Archiving Systems Proceedings, 1996, Boston, pp. 23-32
- [Kara95] A Computational Market for Information Filtering in Multi-Dimensional Spaces, G.J. Karakoulas and I.A. Ferguson, in Working Notes of AAAI-95 Fall Symposium Series on AI Applications in Knowledge Navigation and Retrieval, 1195, pp. 78-84
- [Mart96] Generic Information Engine for Use in Information System Servers, J. M. Martinez et al., Proceedings of SPIE Multimedia Storage and Archiving Systems, 1996, Boston, pp. 278-289
- [Mheg96] ISO/IEC 13522-5, ISO/IEC JTC1/SC29/WG12: Multimedia and Hypermedia information coding Expert Group (MHEG), 1996
- [Mitt94] Experience with A Personal Learning Assistant, T. Mitchell et al., Communications of the ACM, 37(2):80-91, 1994
- [Ovum95] Applications for the Superhighway: Market drivers, John Moroney, John Matthews, Ovum Reports 1995
- [Selc96] CHIMP: A Framework for supporting Distributed Multimedia Document Authoring and Presentation, K. Selcuk et al., ACM Multimedia 96, Nov. 18-22, 1996, Boston
- [Thim94] A Mail-Based Teleservice Architecture for Archiving and Retrieving Dynamically Composable Multimedia Documents, Heiko Thimm et al., COST 237 Proceedings, 1994
- [Wolf96] Resource Management for Distributed Multimedia Systems, Lars C. Wolf, Kluwer, 1996