A System for Task-Oriented Content Analysis and Search in Media Production

W. Bailer[1], P. Altendorf[2], A. Messina[3], F. Negro[3], G. Thallinger[1]

[1] JOANNEUM RESEARCH Forschungsgesellschaft mbH, DIGITAL – Institute for Information and Communication Technologies, Austria
[2] Institut für Rundfunktechnik GmbH, Germany
[3] RAI - Radiotelevisione Italiana - Centre for Research and Technological Innovation, Italy

E-mail: [1] {werner.bailer,georg.thallinger}@joanneum.at, [2] altendorf@irt.de,
[3] {a.messina,f.negro}@rai.it

Abstract: **The project TOSCA-MP aims at providing more efficient ways of annotating and searching audiovisual content in professional media production processes. To achieve this goal, automatic content analysis tools and novel search methods are developed. In this paper, we first present an overview of the components in a system addressing this challenge. We then analyse the scenarios, use cases and user tasks that such a system needs to support. Based on this, a logical and a technical view of the system design are presented, strongly relying on service oriented architectures and the recent FIMS standard.**

Keywords: media search, content analysis, content annotation, usage scenarios, system design, SOA, FIMS

1 INTRODUCTION

The work presented in this paper focuses on professional audiovisual media production and archiving workflows, in particular workflows deployed at broadcasters and media production houses. Many of the media professionals' tasks in such a workflow include searching for content in different modalities, such as finding appropriate clips from recent material to be included in the production, locating relevant archive material or reviewing relevant sections of other media's coverage on the same event or topic. Media production and archiving workflows are changing rapidly due to the appearance of a wide range of new ways of media production, distribution and consumption. These changes pose several challenges for the technology used in media production and archiving.

For reasons of flexibility and efficiency, the workflows need to be increasingly networked and distributed. Journalists on location are taking over more and more tasks that were previously done by other dedicated staff at a local studio or central facility. Due to increased collaboration and content syndication, media search is no longer a problem of searching an in-house repository, but one of searching distributed large-scale repositories, which are multilingual and heterogeneous in terms of their structure and data models. As content owners want to keep control over their repositories, centralised indexing is not an option. Novel networked media search technologies are required. The main challenge for this approach is the successful development of an infrastructure capable of handling multiple heterogeneous repositories distributed across the network, encompassing a sophisticated user interface for the seamless integration of different distributed repository types and services.

The transition from traditional tape-based production to an entirely file-based workflow is now a reality that needs to be managed efficiently. This change, however, requires the adoption of information technology (IT) as a fundamental technical enabler for production and archiving of content. Broadcasters are now in the position to embrace a new era, where barriers between dedicated technologies like VTRs (Video Tape Recorders) or SDI (Serial Digital Interface) networks and IT components are fading to nought. However, using IT means interfacing with a highly heterogeneous market made up of a broad panel of different standards and industries. A solution is to apply for the workflow principles of open service-oriented architecture (SOA), that allow customising systems on the basis of standardised component interfaces. The availability of off-the-shelf distributed, IT-based services for search and retrieval, seamlessly integrated with the production, publication, and archival processes is considered crucial for the near future. Distributed and networked search technologies represent a viable solution for implementing systems supporting media production processes in a sustainable and future-proof way.

The paradigm shift we see in media production will not result in a new set of fixed workflows, but the new workflows will be dynamic and constantly changing, as novel IT-based production and distribution technologies emerge. The tools for tomorrow's media production and archiving workflows must thus be able to adapt to the user's tasks, content types and production context. This adaptation can be enabled by task models, means of implicit and explicit user feedback and integration of benchmarking capabilities. Despite the complexity of search systems distributed throughout the network, tools must be user-centric, and provide a single, integrated user

Corresponding author: Georg Thallinger, JOANNEUM RESEARCH, DIGITAL – Institute for Information and Communication Technologies, Steyrergasse 17, 8010 Graz, Austria, phone: +43 316 876 1240

interface, in order to support professionals in their daily tasks.

The rest of this paper is organised as follows. Section 2 presents an overview of a system as being developed by the TOSCA-MP project[1], addressing these challenges. In Section 3 we discuss the use cases and tasks considered, as well as the derived requirements. Section 4 presents the system design and the standardised service interfaces being implemented. Finally, Section 5 provides conclusions and outlook.

2 OVERVIEW

The system described in this paper is developed in the context of the TOSCA-MP project, which aims at developing user-centric content annotation and search tools for professionals in networked media production and archiving (television, radio, online), addressing their specific use cases and workflow requirements.

The project performs research and development in the following technology areas. For advanced multimodal information extraction and semantic enrichment, scalable and distributed content processing methods are developed, focusing on visual content and speech. Existing approaches are adapted in order to be applied in task and genre adaptive ways. Task adaptation is achieved by formalised models of user tasks in the media production workflow. These models are used for the orchestration and configuration of automatic services as well as for benchmarking the tools and services.

A second key area concerns methods for searching across heterogeneous networked content repositories. In order to enable professionals in media production and archiving to seamlessly access content and indexes across distributed heterogeneous repositories in the network, a distributed repository framework is being developed. This repository framework will allow instant access to a large network of distributed multimedia databases and including beyond state-of-the-art metadata linking and alignment. The distributed repositories can be accessed through a single user interface that provides novel methods for result presentation, semi-automatic annotation and means of providing implicit user feedback. All these components are integrated in an open service-oriented architecture.

3 SCENARIOS, TASKS & REQUIREMENTS

3.1 Usage Scenarios and Requirements

The TOSCA-MP consortium chose to use the S-Cube methodology [3] to identify a two-level hierarchical description of usage scenarios. This methodology distinguishes between high-level goals (named business goals), i.e. target conditions which are to be met by the system from mainly a business process-oriented perspective, and more detailed scenarios, which describe more practical settings in which actors and systems interact to achieve a specific result.

Overall, the consortium identified 10 different business goals and a total number of 15 scenarios, i.e. at least one scenario for each identified business goal. Each business goal comes with a textual description which illustrates the rationale underlying the goal and its main objectives. Business goals and scenarios span a considerable range of real-world media production processes, and capture an important portion of the media production value chain, all of which could actually benefit from the employment of TOSCA-MP results [4].

The identified business goals and scenarios fall into broadly four categories: content access & retrieval, news service distribution, assisted production, and infrastructure. Table 1 summarises the description of the main identified business goals and scenarios.

Business goal	**Scenario for business goal**	**Category**
Fast retrieval of very recent material	Fast content discovery for news production	Content access & retrieval
Efficient retrieval of historical archive material	Searching archived material, including deep archive search	Content access & retrieval
Access to international feeds and their use in news production	Distributed semantic search and retrieval of multilingual content, dynamic configuration of features for content enrichment, machine-supported subtitle generation	Content access & retrieval, assisted production
News daily report with event detection and impact analysis	Assisted production of news stories using distributed multilingual sources	News service distribution
Assisted production of sports events	Summary of downhill race, summary of downhill world cup season	Assisted production
Distributed repository for all steps in metadata production and usage chain	Distributed content metadata production and post-production, distributed search and recommendation	Infrastructure

Table 1. Summary of identified business goals and scenarios.

[1] http://www.tosca-mp.eu

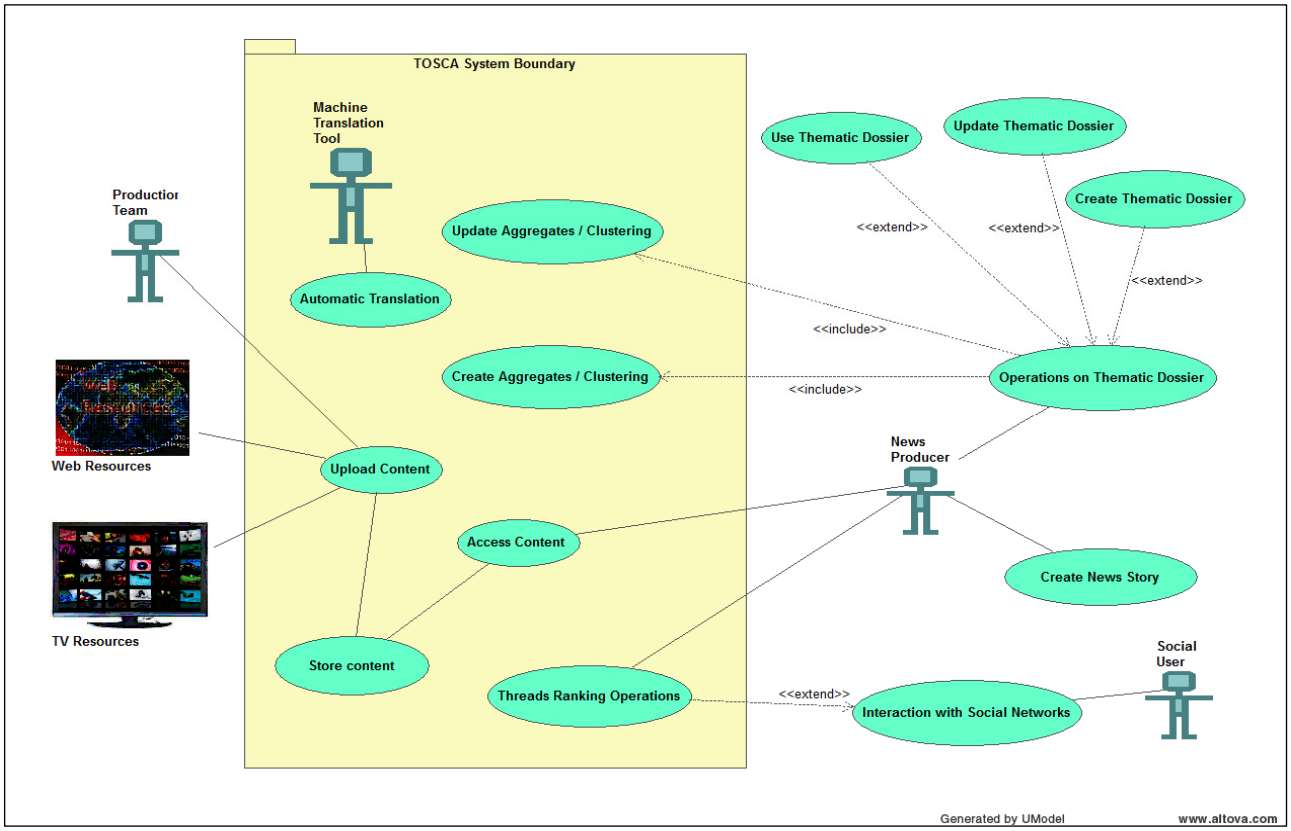

Figure 1. Example of usage scenario connected with business goal "Assisted production of news stories using distributed multilingual sources".

Scenario descriptions are complemented by UML use case diagrams, which go into more detail by breaking down scenarios into the actors, components and functionalities involved. Thus they provide insights about what functionalities will be part of the system and what actors will be involved in the practical test cases. An example of scenario description is depicted in Figure 1. At the headquarters of an Italian broadcaster, a news producer has to create a story for the evening edition of the newscast. He can choose the subject to talk about, so he would like to select the one which is, at the moment, considered the main thread on the net. The news producer queries a central system to give him the threads ranked according to the number of published items on the web and on television and to the behaviour of users on different social networks. These items are uploaded in the system by a dedicated production team. After having selected the subject, he asks the central system to get a constant update on the selected theme. The update comes in form of a multimedia report (or dossier) containing multilingual material properly aggregated/clustered coming from several distributed sources of information, both television and web. The Italian news producer can enrich his news story taking suggestions in his own language coming from these heterogeneous materials.

Based on the scenario descriptions and use case diagrams, also workflow and interaction diagrams had been developed. An analysis of these workflow diagrams has helped identifying similar workflow patterns and to isolate the related functionalities into workflow building blocks that can be plugged into higher level workflow descriptions. As an example of this abstraction process, Figure 2 depicts the generic sequence diagram for metadata generation, which involves some of the main components of the architecture (see Section 4). The requirements for the components involved in each of these workflows, and in particular workflow management and repository components which form the core of the system, have been derived by means of this approach.

3.2 User Tasks

The project has collected a set of real-world tasks in the media production workflow that are considered within the scope of TOSCA-MP. In order to help to characterise them and to formalise task models and success metrics based on them, properties of these tasks have been collected by performing a survey[2], and using information gathered by the EBU MIM/SCAIE working group[3]. A task is defined as a sequence of actions performed by one or more users to achieve a defined goal in the production process, possibly using a set of tools. The task has a defined set of input documents and produces a set of output documents. For example, a "Content Search Task" would be defined as "The action performed by a journalist to find an audiovisual content item with a specified title".

The collected/described tasks cover many aspects of the audiovisual media production workflows, such as

[2] The survey is still open at http://www.tosca-mp.eu/tasksurvey
[3] http://tech.ebu.ch/groups/pscaie

annotation and documentation of incoming news and sports material as well as archive content, search for multilingual news content, personalised news production and live subtitling of news. Other tasks deal with gathering material for a documentary, performing editing in a distributed environment and creating highlight summaries for news and sports content. A complete list and more detailed information about the tasks can be found in [4].

We are currently working on a formalisation of the collected tasks descriptions, in order to obtain machine readable descriptions of tasks. The ConcurTaskTrees (CTT) formalism [6], a graphical model for tasks, is used. These will be used to orchestrate the services in the metadata production management framework described in the next sections, and to derive benchmarks from the expected tasks results for evaluating individual tools or sets of tools developed in the project. The latter enables the integration of benchmarking into workflows and to dynamically adapt the processing tools to new types of input.

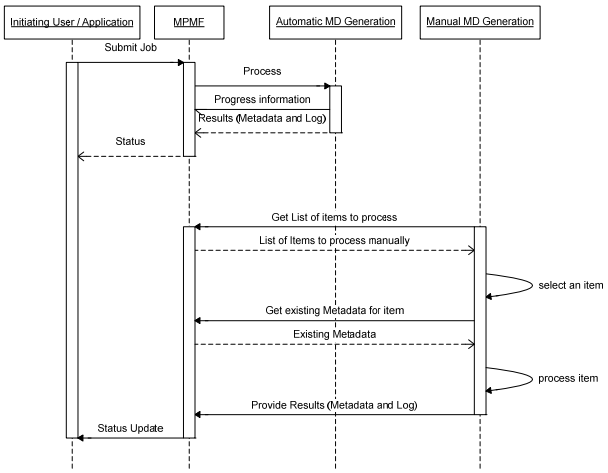

Figure 2. Interaction diagram of metadata generation process.

4 SYSTEM DESIGN

4.1 Logical View

The analysis of the above mentioned tasks and requirements first resulted in a logical view of the system: the Logical System Design (LSD). The LSD as depicted in Figure 3 is composed of four subsystems: the Metadata Production Management Framework (MPMF), the Distributed Repository Framework (DRF) and sets of services and graphical user interface components.

The MPMF consists of an Enterprise Service Bus (ESB) and a process engine and is the central system component which integrates all other components. Tasks and workflows are modelled in this integration platform and orchestrated by means of message mediation between the components.

The DRF is the data management centre of the system where internal databases, triple stores, file systems and other types of storage are administrated. It is a scalable integrated repository for data used at different stages of a media production workflow.

All components providing dedicated functionalities like the processing of data in the production workflow e.g. for automatic metadata generation, semantic enrichment and linking etc. are exposed as services. This allows loosely coupling of components and thus a highly flexible architecture.

Several GUIs are needed to allow users to interact with the system, i.e. to administer the system, to manage and monitor processes, to launch queries, inspect results etc. In the LSD all GUIs are packaged together as one logical component which is connected to the MPMF and the DRF, but there will be several GUIs providing specific functionalities.

Figure 3 also shows the relations and the different connections between the four main subsystems. We differentiate between connections used to exchange essence and metadata between the system components; and those for control purposes only and thus only exchange metadata.

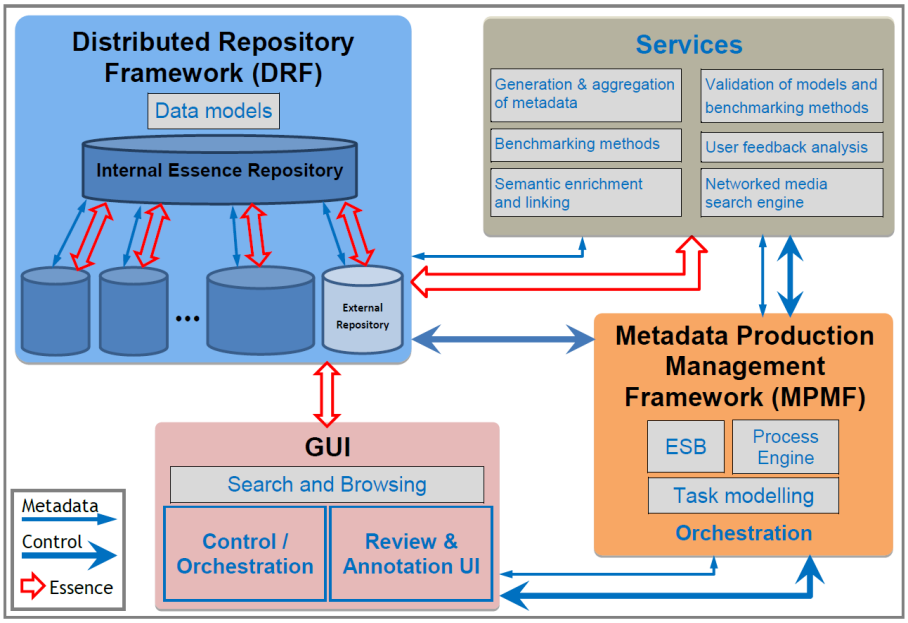

Figure 3: Logical System Design (simplified version).

4.2 Technical View

Systems for file-based media production became increasingly complex in the past years. Standard IT-based hardware and software components are typically tightly coupled in spider web-like system environments. To reduce the complexity in such heterogeneous systems the paradigm of Service Oriented Architectures (SOA) is increasingly adopted. SOA-based systems consist of individual services that are loosely connected with each other by a service bus (see Figure 4) [1].

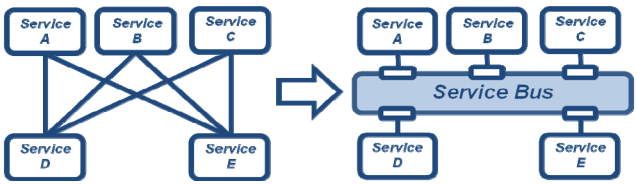

Figure 4: From spider web to service bus [5].

The SOA-based Technical System Design (TSD) consists of five layers and provides a more technical view on the system (see Figure 5). The Application Layer forms the top layer which provides GUIs allowing users to access the system. Applications implementing the GUIs interact with the services in the Orchestration Layer.

The Orchestration Layer is a wrapping layer for the Service Layer and the MPMF. The Service Layer provides access to and control of the different components in the system through web service interfaces. The MPMF contains the definitions of service orchestrations, called processes. An ESB as open integration platform allows in combination with a process engine the execution of processes started by the Application Layer to invoke business logics.

The Component Layer below the Orchestration Layer contains the actual components for metadata extraction, search etc. as well as the DRF. The functionalities of these components are exposed as services to the Orchestration Layer using web service interfaces in the Service Layer.

The Data Layer handles the exchange of content (essence and metadata) between the DRF and the components or the GUI.

4.3 A SOA Framework for Media Services

In addition to a reduced system complexity, the SOA approach enables easier integration of new components and better scalability. Many vendors in the broadcast industry already offer service-oriented interfaces for their systems [5]. However, even if the systems generally serve the same purpose, their interfaces typically differ in their functionality, complexity and data model. To reduce the integration efforts a joint task force initiated by EBU and AMWA called FIMS [4] targets the standardisation of interfaces and formats.

The FIMS specification of a common SOA framework for media services [2] describes a high-level architecture and framework. In its initial version it also defines service interfaces for three basic media services: capture,

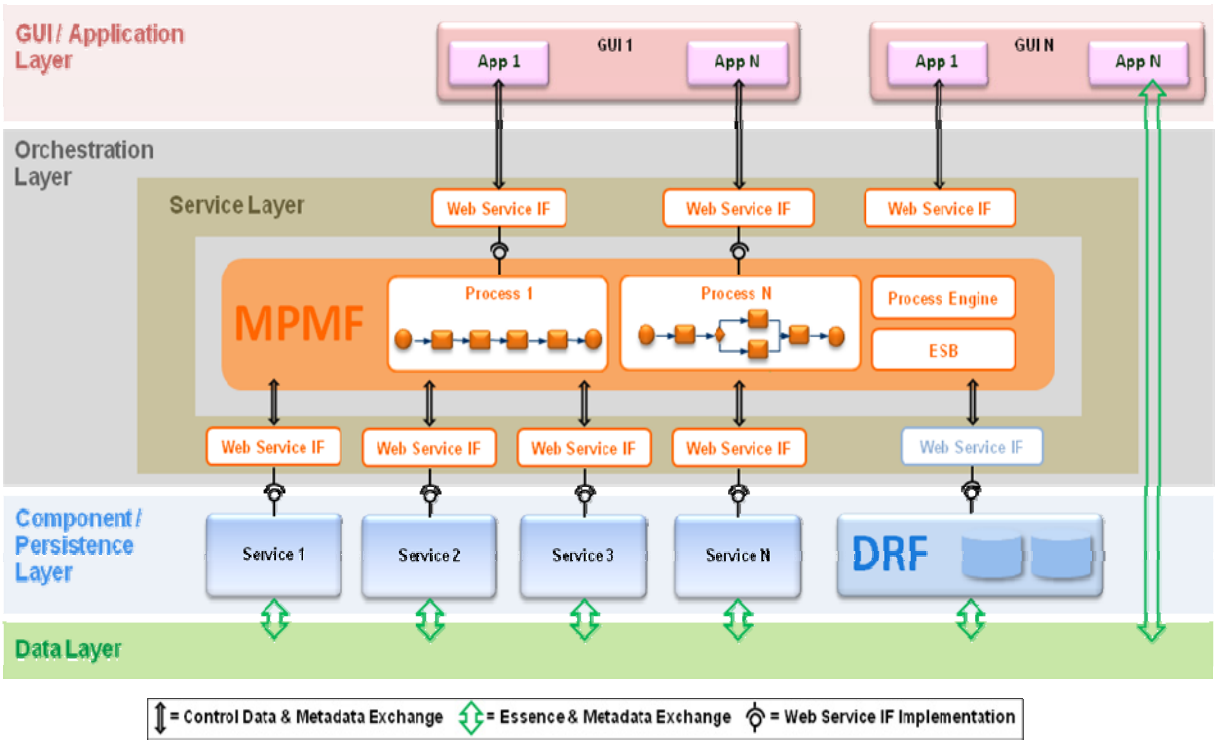

Figure 5: Technical System Design.

[4] Framework for Interoperable Media Service, http://wiki.amwa.tv/ebu/

transform and transfer.

The services used by the system described in this paper will be implemented following the baseline of the FIMS specification as far as possible.

5 CONCLUSION & OUTLOOK

The TOSCA-MP project aims at providing more efficient ways of annotating and searching audiovisual content in professional media production use cases. To achieve this goal, automatic content analysis tools and novel search methods are developed.

In this paper, we have analysed the scenarios, use cases and user tasks that a system for this purpose needs to support. Based on this, a logical and a technical view of the system design was presented. The system design is based on service-oriented architectures and will make use of the recent FIMS standard.

At the time of writing this paper, the specification of the systems has been completed. Many of the related documents are public, available at http://tosca-mp.eu/publications/public-deliverables. The implementation of initial components is in progress. A first integrated version of the system is expected to be ready by mid 2013.

Acknowledgement

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 287532, TOSCA-MP – Task-oriented search and content annotation for media production (http://www.tosca-mp.eu). The project brings together 10 partners from 5 European countries, including industry partners providing solutions for the media industry, public service broadcasters as well as their European association, a university and research centres.

References

[1] J. Footen, “The Service-Oriented Media Enterprise”, Focal Press, Burlington 2008.

[2] FIMS, “FIMS Media SOA Framework 1.0”, September 2011, http://wiki.amwa.tv/ebu/index.php/Framework_Specification (retrieved May 9th, 2012).

[3] E. Di Nitto, P. Plebani, “Describing Case Studies: the S-Cube approach”, Technical Report, Dipartimento di Elettronica ed Informazione Politecnico di Milano, 2010. Available at http://www.s-cube-network.eu/pdfs/EC-document-20100713.pdf

[4] W. Bailer, A. Messina, “Relevant Tasks in A/V media Production Workflow”, TOSCA-MP Public deliverable D4.1.

[5] Elser, Matthias et al., “Integration von Produktionssystemen (Proof of Concept) – ein anderer Ansatz”, FKT 6/2012, 2012.

[6] F. Paterno, C. Mancini, S. Meniconi. "ConcurTaskTrees: A Diagrammatic Notation for Specifying Task Models," Proceedings of the IFIP TC13 International Conference on Human-Computer Interaction Pages, pp.362-369, Syndey, 1997.