

Multimedia Data Management To Assist Tissue Microarrays Design

Julie Bourbeillon, Catherine Garbay, Joëlle Simony-Lafontaine
and Françoise Giroud

Laboratoire TIMC-IMAG, IN3S, Faculté de Médecine,
38706 La Tronche cedex, France
 {Firstname.Name}@imag.fr

Abstract. In oncology research, Tissue Microarray (TMA) technology allows for the mass treatment of hundreds of tissue samples and rapid visualisation of molecular targets. Since this technique is relatively new, there are very few dedicated information systems and little formalised knowledge about the technique is available. We therefore intend to set up an integrated system around TMA technology that is accessible from the Internet. In particular we intend to set up a multimedia document generation system to assist with TMA design.

1 Introduction

Tissue Microarray (TMA) technology [1] is a new technique used in oncology research. It allows for quick *in situ* visualisation of molecular targets in thousands of tissue samples. In this technique patients are selected according to the study to be conducted. A pathologist analyses a histological slide for each patient biopsy block and marks areas of interest. Tissue cores are extracted from each biopsy paraffin block (donor block) and inserted in a new paraffin block (receiver or TMA block) in which slides are cut and treated as conventional slides would be. Images for TMA slides are acquired annotated and submitted to image analysis treatments.

Compared to classical studies, those using TMA technology economise on reactants and biological material. Moreover, mass treatment brings a statistical dimension to the pathologist's work. Both advantages can be increased by using the virtual TMA slide concept: spot images can be reorganised according to a new study without constructing a new real block. However TMA technology suffers from a lack of formalised knowledge and automation of TMA design and data analysis.

A major step in the computerisation of TMA technology has been the definition of the TMA Data Exchange Specification [2]. But most tools focus on data management [3], and an increasing need for assistance in data mining is being noticed [4]. We therefore intend to set up an integrated web-based system providing assistance in real TMA block design and TMA data mining through the generation of virtual representations of future TMA blocks or virtual TMA slides. The system will have to generate these representations "on the fly", according to a user query. These will have to be a complex mix of heterogeneous data adapted to users' needs.

In this paper we present preliminary work to formalise this adaptation process.

2 Related Works

User-adaptive software systems and especially adaptive hypermedia and adaptive web [5] compute the layout and contents of web pages according to a user model. Beyond this simple personalisation according to preferences, some systems such as e-Learning or on-line newspapers [6] aim at representing a narrative coherence through a spatial organisation. Compared to such approaches the envisioned system intends to reach a representative thematic organisation according to a query which is analytic by nature. This implies that we also focus on task adaptation.

Indeed, as with all information retrieval systems the goal is to provide some relevant information according to a user query. In our case however, as with web page summarising [7] or news tracking systems (such as News-map [8]), the returned data must be seen as an organised collection of documents. In the pictured system, the collected data is even meant to support data mining operations. As a consequence, the retrieved data has to be selected and organised to present an overview of the complete collection through a representative set which keeps the variety present in the original collection and allows for statistical analysis.

To achieve these adaptations, it is necessary to acquire and represent some knowledge about the adaptation task and the application domain. Using ontologies [9] will facilitate the modelling and knowledge sharing amongst human and software agents.

3 TMA Design: Problem Overview

Given a request the document to be designed may be thought of as an arrangement of objects, each being characterised by information extracted from a data repository. Therefore virtual representations of real TMA blocks and virtual TMA slides can be described as collections of multimedia documents that are generated according to a user request. They consist of a user query defining the study that the biologist or the physician wants to conduct (for instance comparing two patients groups...) and a TMA grid (the assembly of spot images selected and spatially arranged according to the user query). Each spot can be linked with data regarding the corresponding patient (clinical data...), image annotation and analysis (staining quantification...).

Starting from available data and knowledge, the tool we intend to build has to generate the TMA multimedia document on the fly. This is a complex problem which can be split into three sub-problems:

- **Selection:** Building a TMA can be thought of as looking at a list of elements (biopsy areas or spots) for a collection (the list for pertinent patients or spots) which fits the demand (the user query) and complies with some general rules.
- **Spatial organisation:** These objects (cores or spots) have to be placed on a grid.
- **Presentation:** The previously generated set has to be displayed to the user in the most user-friendly way while still allowing for varying user preferences.

The complexity of the problem is first of all linked to the heterogeneity of the data to handle. Some elements are purely computer data (virtual objects such as the numeric representation of the slides) and others are references to real objects such as the slides themselves. Some, such as staining intensity, are quantitative, whereas oth-

ers such as patient gender are qualitative. The difficulty is increased due to the associated combinatorics.

Moreover the query is more than just inclusion/exclusion criteria since the aim is to obtain a spatially organised collection of documents which will be used for data mining. As a consequence, it appears necessary to acquire some knowledge about the TMA technology domain and the adaptation process at hand. This will be achieved using ontologies.

4 TMA Design: Ontologies

Building a TMA design ontology first of all requires modelling the pathology field. Currently available medical ontologies are either too general for our purpose or do not include abnormal tissue structures [10]. As a consequence, a colon carcinoma ontology has been put forward by Dr. Joëlle Simony-Lafontaine (a pathologist from the Centre Régional de Lutte Contre le Cancer in Montpellier, France). We also need to represent the objects and concepts associated with the technology. All these notions are integrated in a TMA domain ontology.

Along this domain ontology a task ontology needs to be modelled. An example of a query made by a pathologist expressed in natural language might be: 'Study of colon cancer evolution among men with a virtual TMA slide and using the Ki67 marker.' Such a query can be decomposed into several elements. The Goals of the Study guide the way to arrange the objects into the TMA grid. Inclusion Criteria define the list of patients or spots to take into account to build the grid. The Samples are a list of interesting areas in the biopsies where samples should be taken. The Document Logical Model is the TMA document type to build. The Lab protocols list specific protocols to build the TMA. This decomposition led us to construct a query ontology which guided the formalisation presented in Table 1.

The generation of a TMA document is a process of selection, spatial organisation on a grid, and presentation of this grid to the user. This is a process of adaptation of the final document to a query which is directed by a set of criteria. Some of these criteria are specific to the application domain, whereas others are specific to the query, and they can be influenced by user preferences expressed in the query or in a profile.

Table 1. Example for a formalised query where the elements are referenced according to the query ontology.

Query Element	Problem Element	Formalism ([Father Element]...) [Element] (= [Value])
Goal of the Study	Organisation	[Goal] [Patient] [Diagnostic] [pTNM grade]
Inclusion Criteria	Selection	[Inclusion Criteria] [Patient] [Sex] = [Male]
Samples	Selection	[Sample] [Area] = [Colon] [All]
Document Logical Model	Selection / Organisation	[Document Logical Model] = [Virtual]
Lab Protocols	Selection / Organisation	[Lab Protocol] [Grid Size] = [Default] [Lab Protocol] [Nail diameter] = [Default] [Lab Protocol] [Marker] = [Ki67]

5 Adaptation Process

Given the collection of criteria introduced in the previous section the TMA design issue consists of selecting and ordering a group of relevant criteria, and categorising them according to the query and user preferences in order to propose an **adaptation plan**. The manipulation of such a criteria collection is a complex problem because of its size and possible contradictions. However the studies that can be conducted can be classified into families of similar studies. We can also build an **adaptation model** for each of these families. An adaptation model consists of a more specialised subset of constraints and rules extracted from the whole collection according to the specifics of the corresponding study family.

To achieve the adaptation of the final TMA document to the task those adaptation models have to be instantiated with data extracted from the query. As a consequence adaptation models constitute a representation layer depending on the goal of the study which plays an intermediate role between the domain and query layers.

Three specialisation layers have then been defined. The **Domain Layer** includes the complete collection of criteria for the considered application domain. The **Goal Layer** includes a subset of the previous collection, selected, ordered and parametrised according to the goal of the corresponding query family. The **Query Layer** includes an adaptation model specialisation according to the current query and user preferences.

Following these three specialisation layers, three composition levels corresponding to the three sub-problems to solve can be defined. The **Factual Level** deals with the selection stage where data or facts are analysed in order to present a pertinent element list. The **Logical Level** deals with the spatial organisation step where a thematic ordering of the previous list is achieved. The **Lay-out Level** deals with the

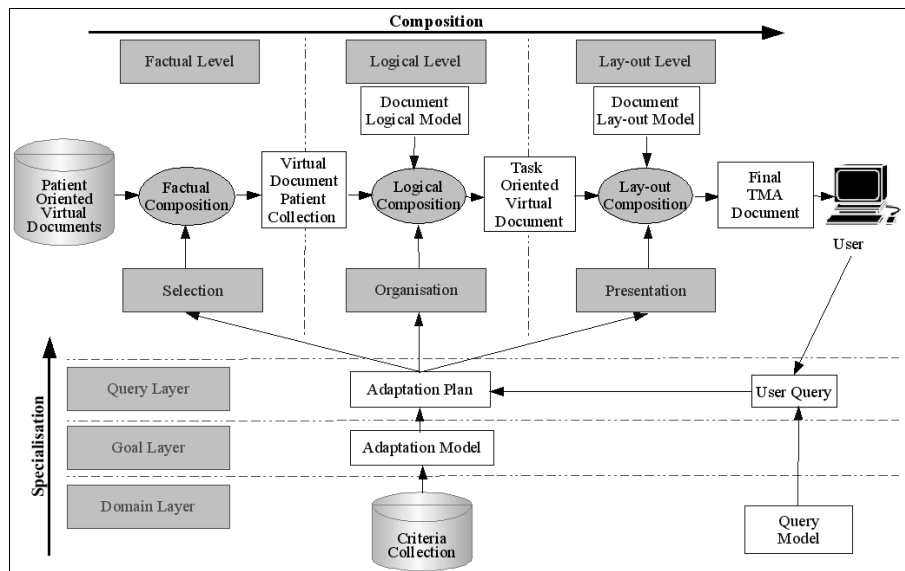


Fig. 1. Adaptation Engine Architecture

presentation stage where the the previous document is prepared to be displayed while taking into account some user preferences.

A TMA document generation is a two-stage process which is achieved by the adaptation engine presented Figure 1. The adaptation engine goes through the three specialisation layers to generate the Adaptation Plan. The application of the adaptation plan to the three composition levels allows to create the final TMA document.

6 Conclusion

In this paper we have proposed an architecture for an adaptation engine which for now, is dedicated to TMA design assistance. The proposed architecture relies on an adaptation engine operating along two axes. The multi-layer Specialisation consists of a progressive refinement of the adaptation procedure towards a particular case. It facilitates knowledge representation and acquisition and allows for flexible query formulation. The multi-level Composition consists of a progressive refinement of the adaptation process through successive steps. It allows for task decomposition and it eases the formalisation of expertise.

The next step would be to go further into the conception and build a prototype.

References

1. Kallioniemi, OP., Wagner, U., Kononen, J., Sauter, G.: Tissue microarray technology for high-throughput molecular profiling of cancer. *Hum Mol Genet.* 10(7) (2001) 657-662
2. Berman, JJ., Edgerton, ME., Friedman, BA.: The tissue microarray data exchange specification: a community-based, open source tool for sharing tissue microarray data. *BMC Med Inform Decis Mak.* 3(1) (2003) 5
3. Henshall, S.: Tissue microarrays. *J Mammary Gland Biol Neoplasia.* 8(3) (2003) 347-358
4. Shergill, IS., Shergill, NK., Arya, M., Patel, HR.: Tissue microarrays: a current medical research tool. *Curr Med Res Opin.* 20(5) (2004) 707-712
5. Brusilovsky P.: From Adaptive Hypermedia to the Adaptive Web. *Communications of the ACM.* 45(2) (2002) 31-33
6. Iksal, S., Garlatti, S.: Adaptive Special Reports for On-line Newspapers. In: Workshop Electronic Publishing, Adaptive Hypermedia (AH) 2002. Malaga, Espagne. (2002)
7. McKeown, K., Barzilay, R., Evan, D., Hatzivassiloglou, V., Klavans, J., Sable, C., Schiffman, B., Sigelman, S.: Tracking and Summarizing News on a Daily Basis with Columbia's Newsblaster. In: Proceedings of HLT 2002: Human Language Technology Conference. San Diego. USA (2002)
8. <http://www.marumushi.com/apps/newsmap/newsmap.cfm>
9. Gruber, TR.: Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies.* Special issue: the role of formal ontology in the information technology 43(5-6) (1995) 907-928
10. Rosse, C., Mejino, JL. Jr.: A reference ontology for biomedical informatics: the Foundational Model of Anatomy. *J Biomed Inform.* 36(6) (2003) 478-500