

# Gestion de connaissances et de données dans l'aide à la conception de Tissue Microarrays

Julie Bourbeillon, Catherine Garbay, Françoise Giroud

Laboratoire TIMC-IMAG, IN3S, Faculté de Médecine, 38706 La Tronche cedex  
[julie.bourbeillon@imag.fr](mailto:julie.bourbeillon@imag.fr), [catherine.garbay@imag.fr](mailto:catherine.garbay@imag.fr), [francoise.giroud@imag.fr](mailto:francoise.giroud@imag.fr)

## 1 Introduction

La technique des « Tissue MicroArrays » (TMA) est une technologie récente, déjà très utilisée en oncologie (Kallioniemi *et al.*, 2001). Selon cette technique: 1) on sélectionne des patients, en fonction de l'étude à réaliser ; 2) un pathologiste analyse une lame histologique des biopsies des patients et détermine des régions d'intérêt, puis 3) on réalise le TMA. Pour ce faire, des carottes de tissu sont prélevées dans les zones prédéfinies du bloc de paraffine de la biopsie (bloc donneur). Ces carottes sont insérées dans un bloc receveur vierge (bloc TMA) à partir duquel des lames sont réalisées et traitées selon les méthodes conventionnelles. Des images de ces lames TMA sont acquises puis ensuite partitionnées et font l'objet d'une annotation et d'une analyse pour quantification de marquage...

Par rapport à des études menées avec des techniques classiques, celles utilisant cette technologie permettent des économies de réactifs et de matériel biologique et apportent une dimension statistique au travail du pathologiste. Ces avantages peuvent être complétés par le recours au concept de lame TMA virtuelle : des images de spots existantes peuvent être sélectionnées et réagencées pour une nouvelle étude sans construction d'un nouveau bloc.

Même si cette technologie semble prometteuse elle souffre d'un manque de connaissances formalisées et d'automatisation. Les outils développés actuellement autour de la technique se consacrent surtout à de la gestion de données (Berman *et al.*, 2003, Henshall, 2003). Il paraît donc nécessaire de proposer un outil d'assistance à deux étapes du cycle présenté ci-dessus :

- aide à la conception de blocs TMA réels, par génération de représentations virtuelles de blocs TMA en fonction de l'étude à réaliser,
- accompagnement de la fouille de données par génération de lames TMA virtuelles associées à des informations pertinentes pour l'étude en cours.

Le système devra générer ces deux types de représentation à la volée selon une requête utilisateur exprimant le but de l'étude. Afin d'être utiles, ces représentations devront consister en une composition complexe de données hétérogènes, ce qui a conduit à les considérer comme des documents multimédia adaptés en fonction des besoins utilisateur. En conséquence, nos travaux s'inscrivent dans la lignée des systèmes d'information adaptatifs (Wu *et al.*, 2002, Brusilovsky, 2002).

Dans cet article nous présentons une approche préliminaire pour formaliser le processus d'adaptation. La section 2 donne un aperçu de notre approche du problème. La section 3 présente les besoins de représentation des connaissances pour la mise en place du système. La section 4 introduit un début de réflexion sur l'architecture du futur système.

## 2 La conception de TMA : analyse du problème

Étant donné une requête, il s'agit de proposer une représentation de lame ou bloc TMA pertinente. Ces représentations peuvent être considérées comme des collections de documents

multimédia comportant :

- la requête utilisateur, qui permet la définition de l'étude à réaliser en utilisant la technologie (par exemple la comparaison entre deux groupes de patients...),

- une grille TMA, constituée d'un assemblage d'images de spots sélectionnés et organisés spatialement en fonction de la requête utilisateur ; à chaque spot peuvent être associées des informations concernant le patient associé (dossier clinique...) ou l'analyse et l'annotation d'image (quantification de marquage, description de structures tissulaires...).

À l'avenir, on peut envisager d'intégrer des références à des études similaires, de la littérature pertinente (PubMed...), ou des informations concernant les molécules étudiées.

À partir des données et des connaissances, l'outil doit générer ce document à la volée, d'une manière proche des systèmes de suivi d'actualité (McKeown *et al.*, 2002). Ce problème d'**adaptation** complexe est décomposable en trois sous-problèmes :

- problème de **sélection** : il s'agit de rechercher, au sein d'une liste d'éléments (spots ou carottes), une collection (la liste des éléments pertinents) qui corresponde à la demande (la requête utilisateur) et respecte certaines règles générales,

- problème d'**organisation spatiale** : il s'agit ensuite de placer des objets (carottes ou spots) sur une grille,

- problème de **présentation** : l'ensemble précédemment généré doit être présenté à l'utilisateur sous la forme la plus conviviale possible, en respectant d'éventuelles préférences.

La complexité du problème est liée à plusieurs facteurs :

- les données à manipuler : cette génération implique l'agrégation d'un ensemble d'objets de nature complexe. Ces données peuvent être hétérogènes dans leur type : textes, images, données brutes... Certaines peuvent être pérennes et d'autres à validité limitée dans le temps ou l'espace. Certaines peuvent être quantitatives et d'autres qualitatives. De plus, les points de vue sur ces données dépendent de l'utilisateur ou de l'objectif qu'il veut atteindre. Enfin, la difficulté peut être accrue par la combinatoire importante associée.

- le but de la requête : la requête n'est pas limitée à des critères d'inclusion / exclusion comme en Recherche d'Information et sa finalité est différente, puisqu'elle vise à proposer une collection qui servira de support à une analyse statistique.

### 3 Ontologies pour la conception de TMA

Construire une ontologie (Gruber, 1995, Guardino, 1998) de la conception de TMA implique tout d'abord la modélisation d'une **ontologie de domaine**, comprenant le champ de la pathologie étudiée, ici le cancer du côlon. Une ontologie du cancer du côlon a été réalisée par le Dr. Simony-Lafontaine, du Centre Régional de Lutte Contre le Cancer de Montpellier.

Il est aussi nécessaire de représenter les autres objets et concepts concernant la technologie. Toutes ces notions ont été intégrées dans une ontologie réalisée avec l'outil Protégé2000 de l'université de Stanford. Cette représentation a guidé la construction d'une base de données relationnelle contenant actuellement une centaine de patients.

Il faut aussi représenter comment manipuler les objets et concepts inclus dans l'ontologie de domaine afin de permettre l'adaptation du document TMA à la question de l'utilisateur. Il s'agit d'une **ontologie de tâche** qui comprend en premier lieu une ontologie de la requête. Un exemple de question posée par l'anatomopathologiste, exprimée en langage naturel, pourrait être : « Étude de l'évolution du cancer du côlon chez les hommes par une lame TMA virtuelle en utilisant le marqueur Ki67 ». Cette requête se formalise sous la forme présentée Tableau 1.

Comme exposé précédemment, la génération d'un document TMA est un processus de

Élément de la requête	Élément du problème	Formalisation ( [Élément père]... ) [Élément] (= [Value])
Objectif	Organisation	[Objectif][Patient][Diagnostic] [Stade pTNM]
Critères d'inclusion	Sélection	[Critère d'inclusion][Patient][État Civil] [Sexe] =[homme] [Critère d'inclusion][Région] = [côlon][tout]
Modèle logique de document	Sélection / Organisation / Présentation	[Modele de Document] = [virtuel]
Préférences	Sélection / Organisation / Présentation	[Matériel] [Taille Grille] = [défaut] [Matériel][Diamètre Aiguille] = [défaut] [Matériel] [Marqueur] = [Ki67]

TAB. 1 – Formalisation d'un exemple de requête pour la conception de TMA

sélection et d'organisation spatiale d'éléments sur une grille, puis de présentation de cette grille à l'utilisateur. Ce processus est un processus d'adaptation du document final à une requête qui est dirigé par un ensemble de critères. Une partie de ces critères sont propres au domaine d'application, d'autres sont spécifiés au sein de la requête. Des préférences utilisateur, exprimées avec la requête ou au sein d'un profil peuvent influencer sur ces ensembles de critères.

## 4 Processus d'adaptation

### 4.1 Modèles d'adaptation

Étant donnée la collection complète des critères, générer le document TMA va consister à sélectionner et ordonner un groupe de critères pertinents puis de les spécialiser en fonction de la requête et des préférences utilisateur afin de proposer un **plan d'adaptation**. Mais la manipulation d'une telle collection de critères est un problème complexe, du fait de sa taille et de possibles contradictions. Une analyse de différents plans d'adaptation suggère qu'il existe des familles de requêtes et qu'il est possible de construire des **modèles d'adaptation** pour résoudre le problème. Un modèle d'adaptation consiste en un ensemble de critères plus spécialisé extrait de la collection complète.

Afin d'adapter le document TMA à la requête utilisateur, ces modèles d'adaptation doivent être spécialisés avec des données extraites de la requête. En conséquence les modèles d'adaptation constituent un niveau de représentation dépendant du but de l'étude. Ce niveau est intermédiaire entre les niveaux domaine et requête. Son utilisation facilite la génération du plan d'adaptation et sa spécialisation pour une requête spécifique.

### 4.2 Décomposition du processus d'adaptation

Partant de la section précédente, trois niveaux de spécialisation, dont la dépendance vis à vis de la requête est croissante, peuvent être définis:

- Niveau Domaine : ce niveau contient la collection complète des critères,
- Niveau Objectif : ce niveau contient un extrait de la collection précédente, sélectionné, organisé et paramétré en fonction de la famille de requêtes, appelé modèle d'adaptation,
- Niveau Requête : ce niveau contient un modèle d'adaptation spécialisé en fonction de la requête courante et d'éventuelles préférences utilisateur.

Parallèlement, trois niveaux d'adaptation, correspondant aux trois sous-problèmes à résoudre peuvent être proposés:

- Niveau Factuel : il correspond à l'étape de sélection,
- Niveau Logique : il correspond à l'étape d'organisation spatiale,
- Niveau de Présentation: il correspond à l'étape de présentation.

La génération d'un document de synthèse adapté à une requête utilisateur implique la manipulation de l'ensemble de ces niveaux par un moteur d'adaptation.

### 4.3 Architecture du moteur d'adaptation

La génération du document TMA est un processus en deux étapes, réalisé par le moteur d'adaptation présenté Figure 1. Il faut tout d'abord générer le plan d'adaptation : le parcours des trois niveaux de spécialisation permet de créer un plan d'adaptation spécifique à la requête en cours. Ce parcours passe par:

- Le choix de la Collection de Critères au Niveau Domaine, pour l'application considérée,
- Le choix d'un Modèle d'Adaptation pour la famille de requêtes au Niveau Objectif,
- La spécialisation du Modèle d'Adaptation en fonction de la requête, qui est construite en suivant un Modèle de Requête spécifique du domaine, au Niveau Requête,

Il s'agit ensuite d'adapter le document TMA à la requête: l'application du plan d'adaptation précédemment généré aux trois niveaux de composition ou étapes de l'adaptation permet de créer le document proposé à l'utilisateur :

- Étape de Sélection : au Niveau Factuel, un processus de composition factuel est utilisé pour sélectionner un ensemble de documents pertinents regroupés au sein d'un Document Virtuel Orienté Collection,
- Étape d'Organisation Spatiale: au Niveau Logique, un processus de composition logique permet d'agencer spatialement la liste précédente en un Document Virtuel Orienté Tâche,
- Étape de Présentation: au Niveau Logique, un processus de composition logique permet d'agencer spatialement la liste précédente en un Document Virtuel Orienté Tâche,

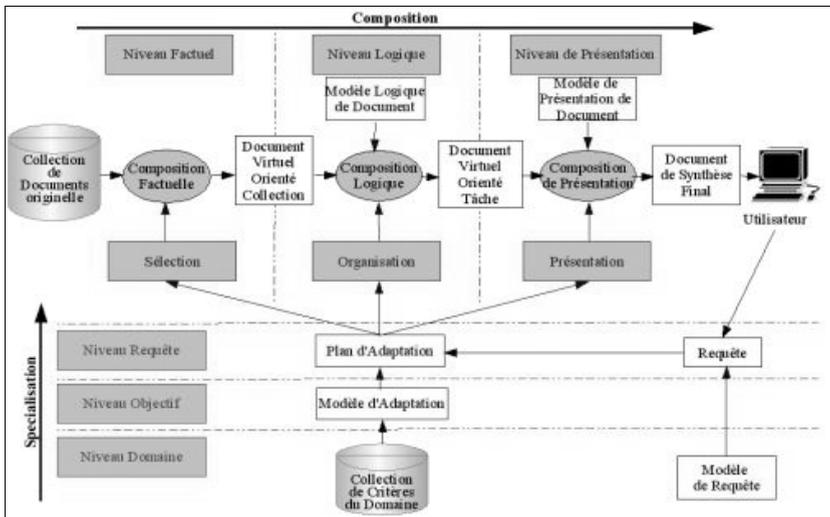


FIG. 1 – Architecture du moteur d'adaptation

— Étape de Présentation : au Niveau de Présentation, un processus de composition de présentation sert à préparer l'affichage du Document de Synthèse final, en suivant un Modèle de Présentation de Document.

## 5 Conclusion

Dans cet article, nous avons proposé une architecture pour un moteur d'adaptation dédié à l'assistance à la conception et à l'exploitation de TMA. Cette architecture repose sur une hiérarchie construite selon deux axes:

— Spécialisation : l'adaptation à plusieurs niveaux de spécialisation consiste en un raffinement progressif de la procédure d'adaptation vers un cas particulier. Du point de vue ingénieur, cela facilite la procédure d'acquisition et de représentation des connaissances. Du point de vue utilisateur, cela permet une formulation et reformulation flexibles de la requête,

— Composition : l'adaptation à plusieurs niveaux de composition consiste en une construction progressive du document de synthèse final en étapes successives. Du point de vue ingénieur, cela permet une décomposition de la tâche, ce qui facilite sa manipulation. Du point de vue utilisateur, cela apporte une possibilité de visualiser des résultats intermédiaires et facilite la formalisation de l'expertise.

Le recours à des modèles tels que les Modèles de Requêtes, Modèles d'Adaptation, ou Modèles Logiques de Documents permet une grande souplesse de l'ensemble du processus d'adaptation. L'étape suivante sera d'affiner la conception et de réaliser un prototype.

## Références

- Berman JJ., Edgerton ME., Friedman BA. (2003), The tissue microarray data exchange specification: a community-based, open source tool for sharing tissue microarray data, *BMC Med Inform Decis Mak*, 23(3), pp. 5.
- Brusilovsky P. (2002), From Adaptive Hypermedia to the Adaptive Web, *Communications of the ACM*, 45(2), pp. 31-33.
- Gruber T. (1995), 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), pp. 907-928.
- Guarino N. (1998), Formal Ontology and Information Systems, *Proceedings of the 1st International Conference on Formal Ontologies in Information Systems 1998*, pp 3-15.
- Henshall S. (2003), Tissue microarrays, *J Mammary Gland Biol Neoplasia*, 8(3), pp. 347-358.
- Kallioniemi OP., Wagner U., Kononen J., Sauter G. (2001), Tissue microarray technology for high-throughput molecular profiling of cancer, *Hum Mol Genet*, 10(7), pp. 657-662.
- McKeown K., Barzilay R., Evan D., Hatzivassiloglou V., Klavans J., Sable C., Schiffman B., Sigelman S. (2002), Tracking and Summarizing News on a Daily Basis with Columbia's Newsblaster, *Proceedings of Human Language Technology Conference 2002*.
- Wu H., De Bra P., Aerts A., Houben G-J. (2000), *Adaptation Control in Adaptive Hypermedia Systems*, Lecture Notes in Computer Science. Vol 1892, pp. 250, 2000.

