

Née le 3 avril 1981 (26 ans)
à Dax (Landes)
Nationalité Française
chevalie@labri.fr

Université Paris-Sud XI
91405 Orsay Cedex
01 69 15 64 99
<http://www.labri.fr/perso/chevalie>

Expérience professionnelle

- 2008 ▷ Poste de recherche post-doctorale dans l'équipe AVIZ, INRIA Saclay, Université Paris-Sud XI.
- 2007-2008 ▷ Attachée Temporaire d'Enseignement et de Recherche au LaBRI, Université Bordeaux I.
- 2004-2007 ▷ Allocataire du Ministère de l'éducation nationale, de la recherche et de la technologie (MENRT) au LaBRI, Université Bordeaux I.
- 2004-2007 ▷ Monitrice au Département Informatique, IUT Bordeaux I.
- 2004 ▷ Ingénieur en Informatique à la clinique du sommeil, CHU de Bordeaux – CDD de 1 semaine.
- 2003 ▷ Stage d'initiation à la recherche au LaBRI, Université Bordeaux I – CDD de 2 mois.

Cursus

- 2004-2007 ▷ **Thèse de doctorat en informatique** au LaBRI, Université Bordeaux I, intitulée "Reconnaissance de motifs dans des graphes : heuristique et applications.", sous la direction de M. DELEST et J.-P. DOMENGER, (soutenue le 12 décembre 2007). *Très honorable*.
- 2004-2007 ▷ **Monitorat d'enseignement**, Université Bordeaux I
CIES d'Aquitaine et d'Outremer, sous la responsabilité d'I. DUTOUR.
Ateliers : Voix et gestes 1 et 3; Gestion du temps; Gestion des conflits
- 2003-2004 ▷ **Master recherche d'informatique**, spécialité Multimédia, Université Bordeaux I. *Bien*
Mémoire intitulé "Reconnaissance d'objets dans la vidéo"
sous la direction de J. BENOIS-PINEAU et M. DELEST.
- 2002-2003 ▷ **Maîtrise d'Informatique**, l'Université Bordeaux I. *Très bien*.
- 2001-2002 ▷ **Licence d'Informatique**, Université Bordeaux I. *Bien*.
- 1999-2001 ▷ **DEUG de Mathématiques, Informatique et Applications aux Sciences**, Université Bordeaux I. *Assez-bien*
- 1998-1999 ▷ **Baccalauréat Série S**, Option Mathématiques, Lycée de Borda, Dax (40). *Assez-bien*.

Formation

- 2006 ▷ Ecole Jeunes Chercheurs en Algorithmique et Calcul Formel
▷ Ecole Jeunes Chercheurs : Grands Réseaux d'Interactions
- 2004 ▷ Ecole Jeunes Chercheurs : Interaction et Visualisation de l'Information
- 2005-2006 ▷ Formation à l'anglais, niveau intermédiaire

Activités de recherche

- ▷ Publications (2 journaux – 2 conférences internationales – 1 conférence francophone)
- ▷ Communications orales (1 référence)
- ▷ Ecoles Jeunes chercheurs et Workshops (4 références auditeur - 1 référence intervenant)
- ▷ Relecture d'articles scientifiques (2 journaux - 3 articles)

Enseignements (296h au total)

<i>Intitulé de l'Enseignement</i>	<i>Type</i>	<i>Niveau</i>	<i>Durée</i>	<i>Effectif</i>	<i>Langage</i>
Initiation à l'algorithmique et à la programmation	TD et TP	Semestre 1	64h/64h/68h	26/28/27	Exalgo, C++
Base de données avancées	TD	Semestre 3	32h	30	SQL
Image	Cours et TD	Semestre 4	24h	24	C++, OpenGL, Java
Traitement d'images	Cours et TD	Licence Pro.	24h	13	Java
Graphes et algorithmes	Cours et TD	Semestre 4	8h/16h	28/24	-

Responsabilités administratives

- ▷ Participation à l'organisation de rencontres scientifiques (1 workshop, PaRISTIC 2005)
- ▷ Vie associative scientifique (1 journée thématique doctorants/entreprises, 1 concert – 1 association)

J'ai effectué mes travaux de recherche d'octobre 2004 à décembre 2007 au sein du LaBRI (Université Bordeaux I), sous la responsabilité de **Maylis DELEST** et **Jean-Philippe DOMENGER**. Depuis janvier 2008, j'ai intégré l'équipe **AVIZ** (<http://www.aviz.fr>), dirigée par **Jean-Daniel FEKETE**, à l'INRIA Saclay (Université Paris-Sud XI) où j'occupe un poste de recherche post-doctorale d'une durée de 8 mois.

Mon activité de recherche s'articule autour du thème de la comparaison de graphes. Mon travail de thèse s'est axé sur l'étude d'algorithmes pour la mise en correspondance de graphes. Plus particulièrement, j'ai travaillé sur l'analyse d'une heuristique de recherche de motifs quasi-similaires dans des arborescences introduite par Auber *et al.* [Evat]. La contribution majeure de mon travail concerne la généralisation aux graphes de la méthode de reconnaissance de motifs, avec la possibilité de choisir entre plusieurs stratégies pour l'identification des motifs, chacune d'entre elles correspondant à une différente définition de la notion de quasi-similarité.

Une partie de mon travail de thèse a porté sur la génération aléatoire de structures arborescentes de fichiers, qui correspondent aux données pour lesquelles la méthode initiale de reconnaissance de motifs a été développée. À partir d'une étude statistique du jeu de données réelles utilisé dans le cadre de ma thèse, j'ai introduit deux algorithmes stochastiques permettant de construire des structures arborescentes de fichier artificielles en tenant compte de paramètres observés, tel que le type de l'utilisateur.

J'ai proposé dans ma thèse une adaptation de la méthode de reconnaissance des motifs à deux problématiques particulières. La première concerne la recherche d'objets dans un flux vidéo à très basse résolution, pour laquelle j'ai également développé une méthode basée sur des techniques de relaxation probabiliste. La deuxième est utilisée dans le cadre de la visualisation de l'évolution d'un projet logiciel au niveau de détail du bloc de code. Dans cette deuxième application, nous proposons un outil d'aide à la maintenance logicielle en introduisant une interface de visualisation permettant le suivi des blocs de code similaires identifiés au fil des versions du projet.

J'approfondis actuellement ma recherche sur la comparaison de très grands réseaux en abordant le problème de la visualisation de l'évolution du graphe des catégories de Wikipédia. Le réseau considéré comporte plusieurs dizaines de milliers de sommets et est très fortement connecté, impliquant ainsi des difficultés de lecture, voire une lecture impossible si l'on utilise des techniques de dessins de graphes noeuds-et-liens classiques pour représenter le graphe dans son intégralité. Ainsi, je travaille à l'heure actuelle sur le développement d'une interface de visualisation du réseau basé sur la représentation de très grands graphes par les matrices d'adjacence agrégées [ZAME]. Ces travaux s'inscrivent dans le cadre du projet RNRT Autograph.

Etude et analyse de la méthode heuristique de comparaison de graphes

L'essentiel de mon travail de thèse a été consacré à l'étude de la méthode heuristique de comparaison de structures arborescentes, proposée par D. Auber *et al.* pour le concours Infovis 2003 [Evat]. L'objectif de cette méthode était de visualiser des sous-arbres similaires dans une arborescence de fichiers pour détecter par exemple, les changements entre deux versions d'une même arborescence (ajout/suppression de fichiers/répertoires) ou les déplacements de sommets ou de sous-arbres. La méthode d'évaluation de sous-arbres similaires (*i.e.* proches en terme de structure), est basée sur une heuristique utilisant des paramètres structurels calculés sur les sommets. La procédure se décompose en deux phases : la première étape consiste à regrouper les sommets en différentes familles distinctes en fonction de leurs caractéristiques structurelles (*e.g.* le degré du sommet, la taille du sous-arbre issu du sommet ou encore le nombre de Strahler qui permet d'évaluer le taux de ramification du sous-arbre issu du sommet). Dans la deuxième phase, il est question de raffiner la classification des sommets par familles en identifiant des paires de motifs similaires. On procède par construction descendante progressive : si pour deux sommets appartenant à la même famille, le voisinage est proche (en terme d'étiquetage par famille de sommets similaires), alors on a identifié un début de motif quasi-similaire. On cherche alors à continuer la reconnaissance par propagation du motif récursivement sur les voisins des sommets.

Mon travail de recherche a porté sur la généralisation aux graphes de la méthode. J'ai ainsi proposé une déclinaison de l'algorithme en différentes variantes qui se distinguent par la stratégie de comparaison du voisinage. En effet, la notion de motifs quasi-similaires peut varier selon l'application car le phénomène que l'on souhaite capturer dépend de la nature des données et de la tâche à accomplir.

J'ai proposé dans ma thèse une application directe de la méthode à la comparaison de réseaux aériens (graphes non orientés) et à la comparaison de réseaux de migrations alternantes (graphes orientés) dans le cadre du projet SPANGEO. Les résultats obtenus ont été jugés pertinents par les géographes collaborateurs du projet. Ces travaux ont donné lieu à une publication [5], et un article long est en cours de rédaction.

J'ai en outre introduit deux adaptations de la recherche de motifs à l'indexation vidéo basse résolution d'une part et à la visualisation de l'évolution d'un projet logiciel d'autre part.

[Evat] D. Auber, M. Delest, J.P. Domenger, P. Ferraro and R. Strandh. EVAT : Environment for Visualization and Analysis of Trees *IEEE Symposium on Information Visualisation Contest*, pages 124-126, 2003.

Modèle de génération aléatoire de structures arborescentes d'un système de fichiers utilisateurs

L'étude de la méthode heuristique de comparaison de graphes a motivé le développement d'un modèle de génération aléatoire permettant de produire des structures arborescentes proches des systèmes de fichiers utilisateurs. En effet, l'heuristique ayant été à l'origine développée pour de tels jeux de données, il était intéressant d'étudier les propriétés de ces structures arborescentes permettant une meilleure connaissance des raisons de l'efficacité de la méthode, et ainsi offrir des pistes de recherches pour son adaptation à des domaines différents. D'autre part, l'obtention de données réelles concernant ce domaine peut s'avérer difficile, voire impossible pour des raisons de sécurité et de confidentialité. Ainsi, le fait de disposer d'un modèle de génération aléatoire permettant de la création de jeux de données de ce type présente un intérêt certain pour la validation d'algorithmes.

J'ai réalisé dans ce cadre une étude statistique du système de fichiers utilisateurs du laboratoire que j'occupais alors, comprenant les arborescences de fichiers de plusieurs membres du laboratoire sans distinction du système d'exploitation utilisé. J'ai introduit, à partir des observations sur le jeu de données réel, deux algorithmes stochastiques de génération aléatoire d'arborescence de fichiers. L'algorithme de base permet la construction d'une série d'arbres indépendants aux propriétés proches des arborescences de fichiers réelles. Dans la version étendue de l'algorithme, nous avons cherché à modéliser les différents "comportements utilisateurs" pour la génération d'un système de fichiers. Dans ce cas, le profil de l'utilisateur influe sur la structure arborescente obtenue. Le modèle proposé permet ainsi de générer des systèmes de fichiers complets, comprenant des utilisateurs aux profils variés. La validation du modèle stochastique se base sur l'étude statistique des structures obtenues ainsi que la comparaison des résultats obtenus avec d'autres modèles de génération aléatoires classiques. Enfin, nous avons appliqué avec succès l'algorithme de reconnaissance de motifs quasi-similaires entre une arborescence de fichiers réelle (n'ayant pas été utilisée comme jeu de données pour l'étude statistique) et un système de fichiers artificiel généré à partir de l'algorithme étendu introduit dans ma thèse.

Un article portant sur ces travaux de recherche est en cours de rédaction.

Reconnaissance d'objets dans la vidéo à basse résolution

J'ai au cours de ma thèse traité le problème de la reconnaissance d'objets extraits de flux vidéo à très faible résolution. Dans le contexte du *rough indexing paradigm* (paradigme d'indexation grossière), les données vidéos que l'on manipule sont de très faible résolution spatiale et temporelle. Alors que l'étude des données en pleine résolution permet une analyse efficace du contenu vidéo, la manipulation du flux vidéo partiellement décompressé implique une perte d'information spatiale et temporelle considérable au profit de la rapidité du traitement. Les méthodes de travail sur de telles données doivent par conséquent pallier à la perte d'information, compenser les lacunes dues au moyennage ou au sous-échantillonnage des données originales en pleine résolution.

J'ai développé deux méthodes au cours de mes recherches pour répondre à cette problématique. Les deux algorithmes proposés sont basés sur l'utilisation des graphes d'adjacence des régions pour la représentation des objets et la comparaison de ces graphes pour évaluer leur similarité. L'évaluation de la similarité entre deux objets repose sur la mise en correspondance entre les différentes régions des objets (un région étant définie comme un ensemble connexe de pixels homogènes du point de vue colorimétrique), et par conséquent, entre les sommets des graphes les représentants. Le premier algorithme utilise les techniques de relaxation probabiliste. Dans cette approche, après avoir calculé une mesure de similarité initiale entre les régions qui constituent les objets à comparer, l'algorithme "corrige" la mesure de similarité entre les paires de régions en fonction de la cohérence du voisinage local. Pour cette approche, nous avons proposé d'une part une nouvelle mesure de similarité entre les régions basée sur des descripteurs de régions (*i.e.* couleur, forme) adaptés à la faible résolution des données. D'autre part, nous avons défini une nouvelle fonction de mise à jour de la similarité en fonction de la cohérence du voisinage. Enfin, nous avons introduit une stratégie de mise en correspondance progressive des régions selon leur surface : en considérant les régions les plus grandes dans un premier temps, nous limitons les problèmes liés à la faible résolution des données.

La seconde méthode est une adaptation de l'algorithme de reconnaissance de motifs quasi-similaires présentée deux sections plus tôt. Le principe est de modéliser les objets vidéo sous forme de graphes orientés et de comparer ces derniers pour identifier les parties quasi-similaires. Dans cette approche, les propriétés extrinsèques aux données telles que la couleur et la surface des régions sont utilisées dans la comparaison de structures de graphes.

Les deux approches ont été testées avec succès sur de longues séquences vidéo. On note cependant de meilleures performances en terme de précision globale et de temps d'exécution pour la deuxième approche utilisant l'heuristique de construction de motifs quasi-similaires.

Chacune des deux méthodes ont fait l'objet de publications : la méthode basée sur les techniques de relaxation probabiliste a été publiée dans un journal international avec comité de sélection [1] ; l'approche par l'identification de motifs quasi-similaires a été présentée en conférence internationale [3]. L'article a été primé (*Honorable*

mention for student paper) et sélectionné pour la publication d’une version longue dans un numéro spécial de journal international [2].

Comparaison de structures arborescentes pour la visualisation de l’évolution de projets logiciels

Dans le cadre d’une collaboration avec Alexandru Telea (Professeur à l’Université d’Eindhoven) et David Auber (Maître de conférences au LaBRI) au cours de ma thèse, j’ai proposé une adaptation de la méthode heuristique de comparaison de graphes pour visualiser l’évolution de projets de développement logiciels. L’algorithme développé pour la détection de code commun entre différentes versions du projet, inspiré de la méthode de construction de motifs quasi-similaires, est appliqué sur les arbres syntaxiques extraits du code source. Outre la procédure de détection de blocs de code similaires, nous avons proposé dans ces travaux, une interface permettant aux développeurs et aux responsables du projet de développement logiciel, de visualiser l’évolution des éléments du projets (déplacement de code, changements minimes de code, suppression de lignes de codes, ...) au cours des différentes versions du projets. La visualisation de l’historique du projet sans afficher le code source permet une vision globale de l’évolution du projet et doit aider à la compréhension des modifications apportées entre chaque version. La méthode de visualisation introduite ici est basée sur une représentation dense des arbres syntaxiques permettant l’affichage de nombreuses versions. Les blocs de code similaires entre les versions sont représentés par des tubes colorés permettant le suivi de ces derniers dans le temps. Nous avons proposé l’utilisation de techniques de rendu telles que la surimpression et l’effet “coussin” pour améliorer la lisibilité.

Ces travaux de recherche ont donné lieu à une publication dans une conférence internationale avec comité de sélection [4].

Visualisation de l’évolution de très grands réseaux : le graphe des catégories Wikipédia

Depuis le début de l’année 2008, je travaille dans le cadre du projet Autograph sur la visualisation du réseau Wikipédia. Wikipedia est devenu tant l’exemple que la preuve qu’un projet ouvert et coopératif peut se constituer, construire et diffuser des contenus multimédias, et principalement du texte. Pourtant, la communauté scientifique (notamment en sciences sociales) n’a que très peu étudié ce phénomène des grands collectifs sur internet. Dans le cadre de ce projet de recherche, nous envisageons de combler quelques lacunes en proposant des outils de visualisation destinés à voir, sous un autre angle de vue que la pratique, l’évolution et la structuration de ces collectifs. En effet, très peu d’outils sont aujourd’hui disponibles pour rendre compte de questionnements scientifiques (sociologie, sémantique, théorie des graphes, sciences cognitives, ...), il nous faut donc les construire.

Mes travaux actuels consistent à proposer un outil de visualisation de l’évolution du réseau des catégories Wikipédia. La visualisation d’un tel réseau pose des problèmes : d’une part le réseau est très grand, les techniques d’aggrégations de sommets s’avèrent donc être une approche intéressante pour la simplification du graphe ; d’autre part le réseau est très fortement connecté, ce qui implique des problèmes de lisibilité dans le cadre d’une représentation noeuds et liens classique dû au nombre important de chevauchements des arêtes. Pour pallier à ce problème, nous proposons de représenter les graphes sous forme de matrices d’adjacences. Elmqvist *et al.* ont déjà proposé ZAME (*Zoomable Adjacency Matrix Explorer*), un outil de visualisation de très grands graphes basé sur la représentation des réseau sous formes de matrices d’adjacence aggrégées [ZAME]. La méthode consiste à appliquer un algorithme de ré-ordonnancement de la matrice dans le but de grouper les éléments en relation. On construit ensuite une pyramide hiérarchique correspondant aux différents niveaux d’aggrégation des sommets. Les sommets sont aggrégés en groupes de taille fixe, et la structure de pyramide hiérarchique permet une navigation interactive entre les différents niveaux de détail.

Dans une telle approche, le principal objectif est de trouver un algorithme d’aggrégation qui est à la fois rapide : les algorithmes classiques reconnus comme étant performants sont quadratiques, ce qui ne convient pas dans notre cas ; et qui produisent une hiérarchie qui présente un intérêt sémantique pour l’utilisateur : en l’absence d’une telle pertinence dans la hiérarchie, l’utilisateur navigue dans des clusters contenant des entrées non corrélées et ne peut extraire du sens à partir de la visualisation. Je travaille actuellement sur une généralisation de l’outil ZAME aux graphes dynamiques orientés. Cette généralisation nécessite le développement d’un nouvel algorithme de ré-ordonnancement et d’aggrégation de la matrice prenant en compte l’orientation des arêtes. Pour traiter l’aspect dynamique, nous avons opté pour une visualisation animée de la matrice, telle que l’utilisateur puisse naviguer entre les différents instants au moyen d’un curseur sur l’axe du temps. Dans une telle approche, il est très important de conserver la carte mentale d’une version du graphe à une autre, d’autant que le réseau considéré contient un grand nombre d’éléments. Dans le réseau des catégories de Wikipédia, le nombre de catégories qui disparaissent d’une version à l’autre est extrêmement faible comparé au nombre de catégories naissantes. Nous avons donc choisi une approche qui consiste à calculer l’ordonnancement et l’aggrégation de la matrice sur le graphe le plus récent (*i.e.* le plus grand) et de réutiliser cet ordre pour toutes les versions antérieures. Les lignes et les colonnes correspondant à des catégories absentes d’une version sont quand même affichées, mais en grisé, de manière à retrouver toujours une même position pour un même élément au fil du

temps, qu'il ait disparu ou qu'il n'ait pas encore été inséré. Enfin, pour que la visualisation du réseau à un niveau de détail supérieur au niveau initial soit utile, les attributs des sommets originaux (tels que le nom des catégories, la date de création ou de suppression) et des arcs (tel que le poids) doivent être combinés pour permettre un accès rapide au détail et une vue d'ensemble pertinente. Nous proposons de remplir les cellules avec de la couleur traduisant la proportion d'arcs de chacun des types parmi : arc nouvellement créé par rapport à l'instant précédent, arc qui va disparaître à l'instant suivant, arc stable par rapport aux instants précédents et suivants, et arc absent. Une telle représentation synthétique du contenu agrégé permet de localiser rapidement les zones du réseau qui ont fortement évolué, ou au contraire, qui sont restées stables.

Ces travaux sont encore en cours de développement.

[ZAME] N. Elmqvist, T-N. Do, H. Goodell, N. Henry and J-D. Fekete. ZAME : Interactive Large-Scale Graph Visualization. In *proc. of the IEEE Pacific Visualization Symposium 2008*, IEEE Press, Japan, 2008.

Le problème de comparaison de graphes occupe une place prépondérante dans le domaine de la visualisation et de la recherche d'information dans des données représentées par des graphes. Une telle problématique d'étude, comparaison et visualisation de graphes, concerne un grand nombre d'applications dans des domaines aussi variés que le multimédia, la biologie, la chimie, les réseaux sociaux ou encore les réseaux informatiques. L'éventail de perspectives de recherches liées à l'étude des graphes et à leur visualisation est large.

Je souhaite vivement approfondir les travaux de recherche initiés dans le cadre de ma thèse concernant l'étude des graphes et leur comparaison. Je projette de poursuivre ma recherche sur la visualisation de très grands graphes dynamiques. En effet, il n'existe que très peu d'outils permettant de comprendre l'évolution et la structuration de grands réseaux tels que les graphes sociaux (réseaux de collaboration, réseau Wikipédia), les réseaux géographiques (réseaux aériens, réseaux de migrations alternantes, trafic marchand), les réseaux biologiques, *etc.* De plus en plus de questions (sociologiques, politiques, sémantiques, ...) se posent autour de ces masses de données toujours plus nombreuses. Il est ainsi nécessaire de proposer de nouveaux outils, qu'ils soient d'ordre général ou spécifiques à une problématique bien précise, permettant l'extraction de connaissances à partir de ces grands réseaux. Je souhaite très vivement m'investir dans un projet me permettant de poursuivre et approfondir mes recherches dans ce domaine. L'amélioration de la méthode de visualisation de blocs de code similaires dans un projet logiciel s'inscrit dans cette perspective de recherche.

L'indexation et la recherche de scènes et d'images dans les documents multimédia est un domaine de recherche très actif. La quantité d'information que l'on retrouve dans les vidéo est telle qu'il est nécessaire d'introduire des outils performants pour effectuer des requêtes portant sur le contenu. L'efficacité de l'utilisation des graphes comme modèle de représentation des données n'est plus à prouver et j'envisage de poursuivre mes travaux de recherche sur la problématique de l'indexation multimédia basée sur la comparaison de graphes. À court terme, il serait intéressant d'adapter les méthodes de reconnaissance d'objets dans la vidéo basse résolution à la recherche d'images pleine résolution dans une collection basée sur une requête par l'exemple. À plus long terme, et à titre exploratoire, je souhaiterais entamer un travail de recherche sur l'utilisation des matrices d'adjacence et d'auto-similarité pour l'indexation multimédia. Il existe en effet des techniques d'analyse d'image permettant d'extraire des motifs à partir d'une représentation de telles matrices sous forme d'image. Par exemple, A. Don a proposé dans sa thèse [Don] une technique d'identification de scènes de dialogues par la reconnaissance de motifs en croix dans la matrice d'auto-similarité des plans vidéo d'une séquence. Il faut entamer une réflexion, selon les données que l'on affiche et le type de matrice que l'on considère, sur les motifs géométriques pertinents à rechercher (motifs en étoile, motifs en bloc, motif en ligne, ...) et le phénomène qu'ils traduisent. On peut envisager d'appliquer des techniques de traitement d'images à la manière de Y. Chiricota *et al.* pour l'identification de communautés dans un réseau [Chiricota]. Dans leur article, les auteurs proposent d'utiliser un filtre gaussien sur un scatterplot 2D construit à partir du réseau pour faire ressortir les régions d'intérêt dans le graphique et ainsi identifier les éléments corrélés dans le graphe. Dans les semaines qui suivent, je devrais entamer une collaboration avec P. Ferraro (Maître de Conférences au LaBRI, équipe MaBioVis), P. Hanna (Maître de Conférences au LaBRI, équipe Image et Son) et M. Robine (ATER au LaBRI, équipe Image et Son) sur un projet visant à identifier des motifs similaires dans un document audio basée sur l'analyse de la matrice d'auto-similarité.

En ce qui concerne le domaine de la génération aléatoire, les perspectives de travail sont là aussi nombreuses. Si l'on se restreint à l'étude des systèmes de fichiers utilisateurs, un travail d'analyse approfondi de l'évolution des arborescences permettrait de comprendre plus finement les mécanismes de structuration des données si toutefois il existe de tels mécanismes. Ce problème soulève la question de l'exploration et l'analyse d'une structure arborescente dynamique : on rejoint alors la thématique principale autour de laquelle s'articule tout mon travail de recherche, à savoir la comparaison de graphes et leur visualisation. Ainsi, pour permettre d'extraire de nouvelles connaissances sur les propriétés des systèmes de fichiers utilisateur, il est nécessaire de réfléchir à un outil permettant de comprendre l'évolution des arborescences, et ce, pour plusieurs utilisateurs. Une telle étude permettrait de proposer des modèles de génération aléatoire plus fidèles à la réalité ; mais présente aussi d'autres intérêts, comme par exemple, dans le domaine de la recherche portant sur les systèmes d'exploitation lorsqu'il est question de valider des systèmes de fichiers, ou encore dans la perspective de développer des outils d'organisation automatique des données numériques.

En conclusion, l'axe principal de recherche autour duquel je me suis investie jusqu'à présent offre d'innombrables perspectives de recherche, et ce, dans des domaines d'applications variés. Aussi, bien que ma thématique prédominante concerne le domaine de la visualisation et de la théorie des graphes, je suis très enthousiaste à l'idée d'appliquer mes connaissances à de nouvelles problématiques.

[Don] A. Don. Indexation et navigation dans les contenus visuels : approches basées sur les graphes *PhD thesis*, Université Bordeaux 1, 2006.

[Chiricota] Y. Chiricota, F. Jourdan and G. Mélançon. Metric-based network exploration and multiscale scatterplot. In *IEEE Symposium on Information Visualization. INFOVIS 2004*, pp 135-142, 2004.

Revues d'audience internationale avec comité de rédaction

- [1] Fanny Chevalier, Jean-Philippe Domenger, Jenny Benois-Pineau and Maylis Delest. Retrieval of objects in video by similarity based on graph matching. *Pattern Recognition Letters*, pages 939-959, 28(8), Juin 2007, Elsevier.
- [2] Fanny Chevalier, Maylis Delest and Jean-Philippe Domenger. A heuristic for the retrieval of objects in low resolution video. *Signal Processing : Image Communication on Content-Based Multimedia Indexing (SPIC)*, Elsevier Science, Volume 22, pages 622-634, 2007, article [3] long.

Manifestations d'audience internationale avec comité de sélection

- [3] Fanny Chevalier, Maylis Delest and Jean-Philippe Domenger. A heuristic for the retrieval of objects in low resolution video. In *5th International Workshop on Content-Based Multimedia Indexing (CBMI'07)*, pages 144-151, Juin 2007. *Honourable mention for student paper*.
- [4] Fanny Chevalier, David Auber and Alexandru Telea. Structural Analysis and Visualization of C++ Code Evolution using Syntax Trees. In *Ninth international workshop on Principles of software evolution (WPSE'07)*, pages 90-97, Septembre 2007.

Manifestations d'audience nationale avec comité de sélection

- [5] Fanny Chevalier, Maylis Delest and Jean-Philippe Domenger. Recherche de motifs quasi-similaires dans des graphes. In *Acte du 4ème Atelier Qualité des données et des connaissances, 8èmes Journées francophones Extraction et Gestion des Connaissances 2008 (EGC'08)*, Sophia Antipolis, January 2008.

Mémoires de recherche

Thèse de doctorat (soutenue le 12 décembre 2007 - Mention *très honorable*), sous la direction de Maylis DELEST et Jean-Philippe DOMENGER. LaBRI, Université Bordeaux I, 2007.

- [6] Fanny Chevalier. Reconnaissance de motifs dans des graphes : heuristique et applications.

Mémoire de DEA, sous la direction de Jenny BENOIS-PINEAU et Maylis DELEST. LaBRI, Université Bordeaux I, 2004.

- [7] Fanny Chevalier. Reconnaissance d'objets dans la vidéo.

Communications à des journées thématiques

▷ Similarité basée objet des segments vidéo pour la navigation dans une vidéo. *Journées NavGraphe*, LINA, Nantes, Avril 2005.

Relecture d'articles scientifiques

- ▷ Numéro spécial du journal *Signal Processing : Image Communication on Content-Based Multimedia Indexing (SPIC)*, Elsevier Science, 2007. (2 articles)
- ▷ *Information Sciences*, Elsevier Science, 2008. (1 article)

Participation à des projets de recherche

J'ai participé à plusieurs projets de recherche au sein de différentes équipes de recherche thématiques.

Équipe GRAVITE

INRIA Futurs - Bordeaux

J'ai été membre de l'équipe GRAVITE, dirigée par Guy Mélançon (professeur au LaBRI, Université Bordeaux 1). Le thème de recherche de l'équipe concerne la visualisation et l'exploration interactives de graphes. L'ensemble de mes travaux de recherche s'inscrivent dans cette thématique.

Équipe AVIZ

INRIA Futurs - Paris-Sud XII

Depuis janvier 2008, je suis membre de l'équipe AVIZ, dirigée par Jean-Daniel Fekete (professeur à l'INRIA Saclay, Université Paris-Sud XI). L'activité de recherche de l'équipe se concentre essentiellement sur la visualisation de très gros réseaux (de l'ordre de millions de sommets et d'arêtes) et les séries temporelles (plusieurs milliards d'enregistrements capturés en continu et en temps réel). Les domaines d'application incluent l'analyse de grands réseaux sociaux (Wikipedia, les développeurs de logiciels libres), les réseaux biologiques, l'intelligence économique, les bibliothèques numériques et les séries temporelles issues des activités de chercheurs. Mes travaux au sein de l'équipe concernent l'analyse et la visualisation de l'évolution du réseau Wikipédia, dans le cadre du projet Autograph.

Projet NavGraphe (2004-2006)

ANR - Masses de Données

Le projet, dirigé par Nicolas Hanusse (MdC au LaBRI, Université Bordeaux 1), traitait du problème de la navigation et l'exploration dans les grands réseaux. L'idée fondamentale de ce projet était de combiner les méthodes automatiques de l'algorithmique de graphes, les requêtes de bases de données et l'oeil de l'expert pour construire une navigation performante.

Ma contribution à ce projet de recherche concerne l'indexation vidéo, avec la recherche d'objets dans une base de données vidéo à très basse résolution [1][2][3].

Projet SPANGEO

GDRE S4 CNRS

L'objectif de ce projet, dirigé par Céline Rozenblat (professeure en Géographie à l'Université de Lausanne) et Guy Mélançon (professeur en Informatique au LaBRI, Université Bordeaux 1), concerne l'analyse et la visualisation des réseaux multiniveaux en Géographie. Je participe à ce projet sur le thème de la comparaison de réseaux géographiques tels que les réseaux aériens et les réseaux de migrations alternantes.

Projet NOSSI

ANR - Calcul Intensif et Simulation

Ce projet, dirigé par Isabelle Baraille (professeure en Chimie théorique à l'Université de Pau), a débuté en mars 2008. Le sigle NOSSI signifie Nouveaux Outils pour la Simulation des Solides et Interfaces. Ma contribution dans ce projet consiste à proposer des outils permettant la visualisation, l'interaction et la comparaison de structures moléculaires représentées sous forme de graphes.

Projet AUTOGRAPH

RNRT Simulation

Le projet Autograph vise à étudier, comprendre et développer des outils avec les acteurs de grands collectifs sur Internet. Ce projet interdisciplinaire réunit informaticiens, sociologues, linguistes et spécialistes de théorie des graphes. L'objectif est de comprendre les formes de gouvernance et les principes d'auto-organisation observables dans de grands collectifs en ligne à base coopérative (Wikipédia, militants des forums sociaux, liste de discussion Debian, joueurs en réseaux...). Il s'agit dans ce cadre de développer avec les acteurs de ces collectifs des outils de visualisation des activités, pratiques et enjeux des participants. Mon travail de recherche lié à ce projet concerne l'analyse et la visualisation de l'évolution du graphe des catégories du réseau Wikipédia.

AquiDoc : Innovation, Monitorat, Insertion

Juin 2007

Journée sur le thème de l'insertion professionnelle des doctorants en informatique dans l'industrie. Cette journée a été organisée à l'initiative de plusieurs associations de doctorants en sciences et fait suite à la journée IDÉE organisée par l'Afodib en 2006. J'ai participé à l'organisation de la programmation des tables rondes et à la préparation de l'évènement.

PaRISTIC : Panorama des Recherches Incitatives en STIC

Novembre 2005

Workshop organisé sur 3 journées et regroupant les membres de 4 ACI du Ministère (Masses de données, Grid5000, IMPBio, Sécurité et Informatique). J'ai participé à l'organisation de l'évènement et à l'accueil des participants. Je me suis également occupée de la gestion de l'affichage pour les sessions de posters.

Afodib : Association de la Formation Doctorale en Informatique de Bordeaux

2006-2008

L'Afodib organise des manifestations destinées à l'information des doctorants du LaBRI, ainsi que des évènements de divertissement qui favorisent les échanges et contribuent ainsi au développement de la cohésion entre les membres du laboratoire. J'ai été membre du Conseil d'Administration pour l'année scolaire 2006-2007. J'ai participé à l'organisation des soirées de divertissement des doctorants. En particulier, j'ai organisé un concert en partenariat avec l'association musicale *Ici Ou Ailleurs* à la Maison des Activités Culturelles à Talence. J'ai également aidé à la préparation de la journée AquiDoc en 2007.

Enseignements

J'ai effectué des enseignements pour l'Université Bordeaux I en tant que monitrice durant les trois années de mon doctorat, puis en tant qu'ATER. J'ai été affectée au Département informatique de l'IUT de Bordeaux I, au sein duquel j'ai effectué l'intégralité de mon service. À noter qu'à l'IUT, la formation continue nous impose d'attribuer une note de TD aux étudiants, impliquant un suivi rapproché de chacun d'entre eux. Les sujets d'examens et corrections des copies sont à la charge de l'enseignant de TD.

Imagerie numérique (Semestre 4)

2006-2007

- Module de 24h de cours et TD du semestre 4 de la formation au DUT.
- Synthèse d'image (12h) et analyse d'image (12h).
- Séances de cours intégré au TD (4h consécutives hebdomadaires).
- Outils pour la modélisation 3D : ac3d, OpenGL.
- Outils pour la retouche d'image : The Gimp.
- Outils de programmation pour la réalisation des TDs : C++, OpenGL, java, eclipse, Visual Studio.
- Rédaction des sujets d'examens et de projets de programmation pour mon groupe d'étudiants.

Ce module traite la programmation pour l'imagerie numérique. Le module est décomposé en deux parties : les trois premières séances de 4h sont consacrées à la synthèse d'image, les trois autres traitent de l'analyse d'image.

J'ai effectué cet enseignement lors de ma dernière année de monitorat. Issue d'une formation en multimédia, le contenu du cours ne m'a posé aucun problème. Dans ce cas, il m'a fallu établir intégralement le cours et les sujets de TP et TD, les supports existants étant très sommaires. J'ai énormément pris de plaisir avec ce module, car j'ai construit mon enseignement, j'ai moi-même rédigé mes sujets d'examens et de projets et j'ai pu orienter le cours dans les directions qui me semblaient les plus appropriées. Il est très rassurant de constater que l'on est capable de mener à bien tout un module, sans que tout les supports ne soient déjà fixés auparavant.

Traitement d'images (Licence professionnelle - parcours Image et Son)

2007-2008

- Module de 24h de cours et TD de la Licence Professionnelle, parcours Image et Son.
- Séances de cours intégré au TD (4h consécutives hebdomadaires).
- Outils de programmation pour la réalisation des TDs : java, eclipse.
- Production des supports de cours, TPs, TDs et rédaction du sujet d'examen.

Ce module est destiné à présenter les techniques de traitement d'images numériques.

Comme pour le module d'image décrit ci-dessus, j'ai eu la charge intégrale de ce module, avec la composition des supports de cours et des sujets TDs et TPs associés, ainsi que la rédaction des sujets d'examen.

Base de données avancées (Semestre 3)

2006-2007

- Module de 32h de TD du semestre 3 de la formation au DUT.
- Manipulation, gestion et optimisation de bases de données.
- Outils pour le développement et la gestion de bases de données : sqlplus, mysql, Access.
- Participation à la rédaction du sujet d'examen.

Ce module traite des bases de données : manipulation, gestion et optimisation. Il est question d'apprendre à avoir une réflexion sur l'efficacité d'une base de données et les alternatives pour optimiser celle-ci. Le module permet également la manipulation des outils pour le développement et la gestion de bases de données.

Il faut noter que lors de ma formation, je n'ai pas étudié les bases de données. En informatique, les progrès constants impliquent que les enseignants mettent à jour leurs connaissances en permanence afin d'offrir aux étudiants la formation la plus adéquate possible au vu des outils informatique alors disponibles. J'ai choisi ce module pour ma dernière année de monitorat afin de m'exercer à l'enseignement d'une matière non maîtrisée. La préparation des cours m'a demandé beaucoup d'investissement. Il m'a fallu apprendre tout ce que j'ai enseigné par la suite.

Graphes et algorithmes (Semestre 4)

2005-2007

- Module de 16h de cours et TD du semestre 4 de la formation au DUT.
- Notions de base sur la théorie des graphes (partie cours : 8h).
- Apprentissage de l'algorithmique et la programmation basée sur la théorie des graphes et quelques-unes des structures de données que l'on peut utiliser pour les représenter. (partie TD : 8h)
- Unité d'enseignement dispensée deux années consécutives (la première année, uniquement la partie TD)

Ce module propose d'aborder quelques notions de base sur la théorie des graphes. Il est question d'apprendre et de pratiquer l'algorithmique et la programmation basée sur la théorie des graphes.

J'ai enseigné ce module lors de mes deuxième et troisième années de monitorat. La première année, le module était composé d'une partie cours regroupant l'ensemble des étudiants et une partie TD en groupes individuels. Je n'ai assuré que la partie TD la première fois. Nous avons choisi de fusionner la partie cours et la partie TD en une seule formule cours intégré au TD car le module ne propose vraiment qu'un aperçu de la théorie des graphes (16h au total) et le cours était souvent en décalage avec les applications directes que nous traitions en TD. Avec la nouvelle formule, il a été possible d'agencer les séances à la demande en faisant du cours ou des exercices à la demande en conservant une logique chronologique dans le module. Grâce à cette formule (que j'ai également pratiquée pour le module Image), j'ai pu exercer un mode différent d'enseignement : le cours magistral. Bien que différent que le cours enseigné en amphithéâtre car le cours ne concerne qu'un groupe d'étudiant, ce nouvel aspect de l'enseignement est différent de ce que j'avais pu pratiquer auparavant. Il n'est ici pas question de résoudre un exercice en rappelant les notions déjà vues lors du cours, il est question d'inculquer de toutes nouvelles connaissances aux étudiants et de les mettre en pratique par la suite.

Initiation à l'algorithmique et à la Programmation (Semestre 1)**2004-2008**

- Module de 64h du Semestre 1 de la formation au DUT.
- Notions de base d'algorithmique, de programmation et de complexité.
- 4 heures hebdomadaires de TD et TP d'algorithmique et de programmation C++.
- Outils pour le développement en C++ : make, g++, emacs, gdb et Visual C++.
- Unité d'enseignement dispensée deux années consécutives.
- Participation à la mise à jour des sujets de TD et TP, à la rédaction des sujets d'examens et de projets de programmation.

Dans ce module, il est question de se familiariser d'une part avec l'algorithmique qui constitue la base de la programmation, d'autre part d'apprendre à utiliser les outils pour la programmation. Les notions acquises en cours sont directement mises en application à travers des exercices pratiques en TD et en TP.

J'ai choisi d'enseigner ce module dès la première année de mon monitorat car il s'agit d'un domaine que je maîtrise complètement. Ainsi, afin de me sentir à l'aise pour mes premières séances, j'ai préféré me concentrer dans un premier temps sur ma pratique de l'enseignement et non sur le contenu du module.

Le programme du module a été remanié par rapport aux années précédentes et j'ai pu, à l'occasion des discussions pédagogiques à ce sujet, proposer mon point de vue et ainsi contribuer à la mise à jour des sujets de TD et TP. J'ai également participé à la rédaction des sujets d'examens et de projets de programmation.

J'ai dispensé ce cours en tant que monitrice et en tant qu'Attachée Temporaire d'Enseignement et de Recherche.

Projet d'enseignement

Durant les quatre dernières années, j'ai pu mettre à profit les possibilités offertes par le département Informatique de l'IUT de l'Université Bordeaux I pour diversifier le plus possible les enseignements qui m'étaient proposés. J'ai pu ainsi montrer mon aptitude à enseigner des matières diverses : des modules issus de ma spécialité aux modules nouveaux pour moi. D'autre part, à l'exception faite de la pratique du cours magistral en amphithéâtre, j'ai pu m'exercer à tous les aspects de l'enseignement : cours, travaux dirigés, travaux pratiques avec la production des supports et sujets. Je souhaiterai à l'avenir poursuivre mon activité d'enseignement dans cette optique d'auto-formation, à savoir régulièrement enseigner de nouveaux modules. J'aimerais d'autre part, proposer des enseignements spécifiques relatifs à ma spécialité, tels que la visualisation de l'information, ou l'indexation multimédia.