

Démarche centrée utilisateur pour la conception de SIAD basés sur un processus d'ECD, application dans le domaine de la santé

HELA LTIFI (1,2,3,4)

MOUNIR BEN AYED (1)

CHRISTOPHE KOLSKI (2,3,4)

ADEL M. ALIMI (1)

(1) REGIM : REsearch Group on Intelligent Machines, Ecole Nationale d'Ingénieurs de Sfax, Route Sokra Km 3.5 - BP W - 3038 Sfax, Tunisie

(2) Univ Lille Nord de France, F-59000 Lille, France

(3) UVHC, LAMIH, F-59313 Valenciennes, France

(4) CNRS, FRE 3304, F-59313 Valenciennes, France

Résumé : Cet article s'inscrit dans le cadre de la conception de systèmes décisionnels centrés utilisateur basés sur un processus d'Extraction de Connaissances à Partir des Données (ECD). Ce processus doit déboucher sur un ensemble de modules, il est itératif et interactif. De ce fait il nous semble indispensable de prendre en compte des principes et des méthodes de l'Interaction Homme-Machine dans le développement de tels systèmes. A ce sujet, le développement de systèmes décisionnels interactifs est actuellement abordé suivant deux approches antagonistes : la première est "technocentrée", dans laquelle la technologie est fondamentale ; la deuxième est "centrée utilisateur" plaçant les acteurs humains en position centrale. Bien que la première approche soit encore très présente en entreprise, la tendance actuelle est résolument "centrée utilisateur". Dans ce cadre, nous proposons une approche qui vise à intégrer les étapes du processus ECD dans un modèle de développement enrichi sous l'angle des interactions homme-machine appelé le modèle en U. Notre contexte applicatif est la lutte contre les infections nosocomiales en milieu hospitalier.

Mots clés : Système Interactif d'Aide à la Décision (SIAD), Extraction de connaissances à partir de données (ECD), Interaction Homme-Machine (IHM), Modèle en U.

Abstract: This article concerns the design of Decision Support Systems (DSS) based on a Knowledge Discovery from Data (KDD) process. This process aims at generating a set of software modules; it is iterative and interactive. For this reason, it is essential to us to take Human-Computer Interaction principles and models into account in the development of such systems. The interactive decisional system development is currently approached according to two antagonistic approaches. For the first approach, technology is fundamental; the second is "user-centered", placing the human actors in a central position. Although the first approach is still very present in companies, the current tendency "is user-centered". In this context, we propose an approach which aims at integrating the stages of KDD process in a development model enriched under the human-computer interaction point of view, which is the U-model. Our application context is the fight against the nosocomial infections in the healthcare domain.

Key words: Decision support system (DSS), Knowledge Discovery from Data (KDD), Human-Computer Interaction (HCI), U-model.

1. INTRODUCTION

Parmi les domaines les plus riches en matière de données stockées on peut citer le milieu hospitalier ; en particulier, les services de réanimation représentent une mine de données encore trop peu exploitée actuellement. Les infections nosocomiales représentent un des problèmes majeurs dans un service de réanimation. Ce sont des infections contractées dans un établissement de soins. Pour lutter contre la survenue de ces infections, le service de réanimation a besoin d'un système lui permettant de l'aider à prendre des décisions tout au long du séjour d'hospitalisation des patients [Kawamoto et al. 2005]. Les systèmes d'aide à la décision seraient appropriés au problème des infections nosocomiales [Guigue et Donadey 1999]. On parle aujourd'hui des systèmes **interactifs** d'aide à la décision facilitant l'accès aux données en ouvrant la possibilité à des analyses plus complètes [Pelayo et al. 2009]. Et au vu du nombre croissant de grandes bases de données médicales, organiser les données, les interpréter et en extraire des informations utiles pour l'aide à la décision est en fait un problème ouvert.

De ce fait, nos recherches s'inscrivent dans le cadre de la conception centrée utilisateur de Systèmes Interactifs d'Aide à la Décision (SIAD) basés sur un processus d'Extraction de Connaissances à partir des Données (ECD). La prise de décision permet d'opérer un choix entre diverses alternatives possibles pour résoudre un problème en choisissant le meilleur compromis parmi elles. Le SIAD traite le problème en fonction de connaissances décisionnelles. Certaines de ces connaissances peuvent être extraites à l'aide d'un outil décisionnel qu'est la fouille de données (appelée aussi Data Mining) [Fayyad et al. 1996] [Jambu 1999] [Lefébure et Ventari 2001] qui peut être vue comme une étape d'un processus plus global qui est celui de l'ECD. Il faudra donc interpréter cette information nouvelle, l'évaluer pour enfin la proposer comme élément valide d'aide à la décision.

Les SIAD consistent à assister les décideurs tout au long de ce processus, en particulier au moment du choix des études à effectuer ou des outils à utiliser. La coopération Homme-Machine est primordiale dans l'aide à la décision ; elle concerne l'assistance du décideur humain et donc la possibilité de partager les tâches entre humain et machine [Pacaux-lemoine 1998] [Tricot 2005]. Sous cet angle, l'interaction homme-machine est un aspect crucial au niveau des systèmes interactifs d'aide à la décision, leur conception devant nécessairement s'appuyer sur une démarche centrée utilisateur. L'évaluation des IHM vient afin de valider la qualité de cette modélisation ou d'énoncer les problèmes rencontrés sous l'angle particulièrement de l'utilité et de l'utilisabilité des systèmes [Nielsen 1993].

Dans le cadre de la conception des SIAD, on se rend compte qu'il s'agit de concilier et d'intégrer des méthodes et modèles provenant d'une part de ce domaine et d'autre part de celui de l'IHM [Lepreux 2005] [Lepreux 2007]. Il existe à ce sujet plusieurs modèles enrichis sous l'angle des IHM tels que le modèle en étoile [Hartson et Hix 1989], le modèle nabla [Kolski 1997] [Kolski 1998] ou le modèle en U [Abed 2001]. Dans cet article étendant [Ltifi et al. 2008], l'idée est d'adopter le modèle en U comme démarche de conception et de l'adapter dans les systèmes exploitant un processus d'ECD.

Notre article débute par un bref état de l'art présentant deux concepts fondamentaux que sont les SIAD et l'ECD, et ceci afin de préciser la vision que nous en avons. Par la suite nous présentons le modèle en U. La troisième partie vise à proposer notre approche supportant la conception et le développement des SIAD basés sur un processus ECD. Avant de conclure, nous présentons la réalisation de notre SIAD où nous donnons des

exemples concrets et des extraits d'interfaces Homme-Machine (relatifs à des modules interactifs sous-jacents au processus d'ECD) provenant du déroulement de l'application de notre approche proposée.

2. SYSTEMES INTERACTIFS D'AIDE A LA DECISION BASES SUR UN PROCESSUS D'ECD

Les systèmes décisionnels centrés utilisateurs se situent à la confluence de trois domaines que sont l'aide à la décision, l'Extraction de Connaissances à partir des Données (ECD) et l'Interaction Homme Machine (IHM). Nous pouvons citer dans ce contexte le travail de [Chevrin et al. 2007] qui met l'accent sur la place de l'IHM dans les processus d'ECD.

2.1. Les Systèmes Interactifs d'Aide à la Décision

Il existe dans la littérature, différents systèmes d'aide à la décision dans le domaine médical [Kaplan 2001] [Matheny et Ohno-Machado 2007] [Evans 2007]. Très peu de ces systèmes accompagnent le décideur durant la totalité du processus de décision. Ils sont pour la plupart des systèmes, "intelligents" ou non, cherchant à optimiser des solutions [Kawamoto et al. 2005] sans être réellement interactifs tout en offrant peu de possibilité d'intervention à l'utilisateur. Ces systèmes informatiques pour l'aide à la décision doivent fournir de façon synthétique et simple les éléments nécessaires aux décideurs pour évaluer une situation ; il nous semble indispensable dans ce contexte de mettre en relief l'aspect interactif et de viser à construire des IHM adaptées aux besoins de chaque type d'utilisateur. C'est le **I** de SIAD, objet de cette section.

De ce fait, nous voyons les SIAD comme des systèmes qui, à travers un dialogue **homme-machine**, permettent l'identification, l'exploration et la résolution de problèmes par un utilisateur (décideur) et/ou groupe d'utilisateurs (décideurs). La communication entre les différents acteurs humains d'un système d'aide à la décision (décideurs, consultants, experts, etc.) s'avère primordiale [Turban 1993]. La notion d'interactivité dans un SIAD renvoie au rôle indispensable de l'homme dans son fonctionnement, rôle non passif qui sous-tend le terme "Aide à la Décision", mais aussi à la qualité de l'intégration des différents composants du système et à la nature de l'interface homme-machine jouant "le rôle de collaborateur" avec le décideur [Garlatti 1996] [Lévine et Pomerol 1989]. Un SIAD se déroule selon un processus composé de trois phases [Klein et Methlie 1990] [Lévine et Pomerol 1989] [Simon 1977] [Turban 1993] : (1) la recherche d'information ou renseignement : l'identification du problème à résoudre nécessite de rechercher les informations pertinentes en fonction des besoins du décideur ; cette phase se termine par un énoncé du problème à traiter ; (2) la conception comprenant la génération, le développement et l'analyse des différentes suites possibles de solutions ; (3) lors de la phase de choix, le décideur choisit entre les différentes solutions identifiées pendant la phase précédente. Cette phase inclut la recherche, l'évaluation et la recommandation d'une solution appropriée.

Dans la littérature sur les systèmes d'aide à la décision, le but d'un système est avant tout de nature pratique. Il s'agit en fait d'assister un décideur en mettant à sa disposition les connaissances dont il a besoin pour résoudre certains cas difficiles. Pour la résolution de certains problèmes le processus de décision est lié donc à la connaissance [Lepreux et al. 2003]. Celle-ci peut être extraite par un outil décisionnel, on parle de l'extraction de connaissances à partir des données (ECD).

2.2. L'extraction de connaissances à partir de données

Dans le domaine des SIAD on considère que l'utilisation d'outils pour extraire des connaissances améliore la pertinence des solutions proposées pour une bonne prise de décision [Fayyad et al. 1996] [Jambu 1999] [Sandoval 1997]. On s'intéresse donc à la technologie la plus répandue de découverte de connaissances, qu'est l'Extraction de Connaissances à partir des Données [Fayyad et al. 1996] [Jambu 1999] (ECD = "KDD", Knowledge Discovery in Databases, ou "data mining" en anglais, traduit par fouille de données, FD, en Français).

Le but de l'ECD est de pouvoir extraire un élément : la connaissance. Pour le définir, on peut dire aussi que l'on vise l'"Extraction des connaissances nouvelles, utiles, valides à partir d'une masse de données" [Fayyad et al. 1996]. On considère que l'ECD est un processus interactif et itératif. L'interactivité est liée aux différents choix que l'utilisateur est amené à effectuer. L'itérativité est liée au fait que l'ECD se déroule suivant une suite d'étapes et que l'utilisateur peut décider de revenir en arrière à tout moment si les résultats ne lui conviennent pas (figure 1).

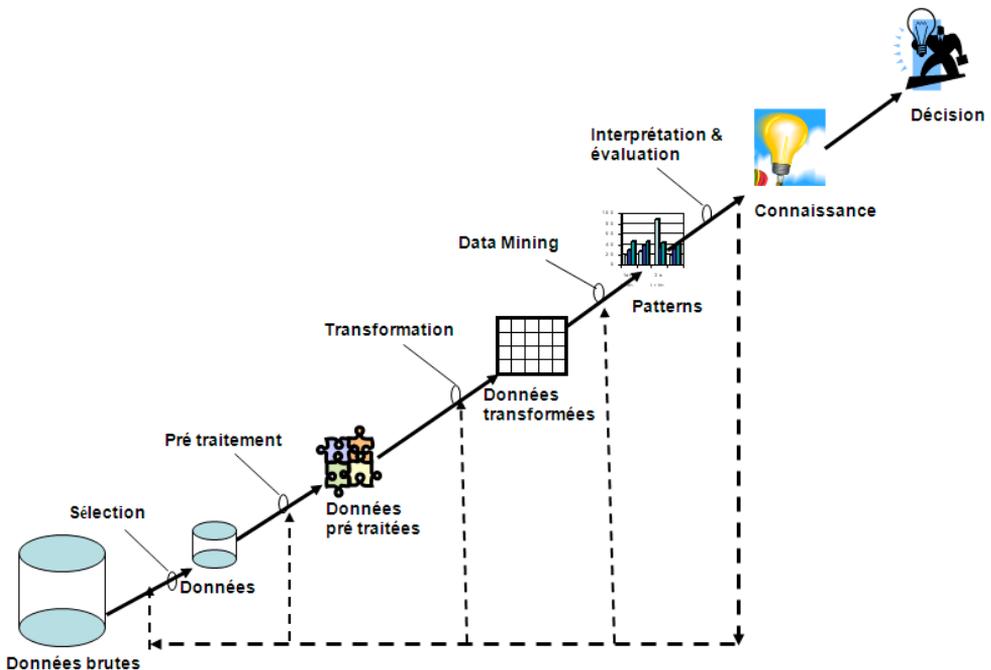


Figure 1 : Le processus d'ECD

Les différentes étapes du processus d'ECD sont les suivantes [Lefebvre et Ventari 2001] : (1) positionnement du problème en cernant les objectifs, définissant les cibles et vérifiant les besoins ; (2) recherche des données (identifier les informations et les sources, vérifier leur qualité ainsi que leur facilité d'accès) ; (3) sélection des données ayant un rapport avec l'analyse demandée dans la base ; (4) nettoyage des données en vue de corriger les inexactitudes ou erreurs de données ; (5) transformation des données dans un format qui les prépare au Mining (transformer dates en durées, ratios, etc.) ; (6) fouille de données, application d'une ou plusieurs techniques (réseaux de neurones, réseaux bayésiens, arbres de décision, etc.) pour extraire des patterns

intéressants ; (7) évaluation du résultat permettant d'estimer la qualité du modèle découvert ; (8) intégration de la connaissance en implantant le modèle ou ses résultats dans le système informatique de l'entreprise.

2.3. Système Interactif d'Aide à la Décision basé sur un processus d'ECD

Un SIAD basé sur l'ECD, sujet de cet article, est un système qui permet de détecter les stratégies de résolution d'un problème de décision par le biais d'un processus de fouille de données. Dans ce processus, l'analyse des besoins des décideurs, les différentes activités réalisées en rapport avec la préparation et la manipulation des données pertinentes, de même que la visualisation des résultats constituent des étapes très importantes. C'est sur de telles étapes que repose l'acceptation ou le refus par l'utilisateur final de l'outil d'aide à la décision visé [Lefébure et Ventari 2001]. Les interactions homme-machine au niveau de ce système devraient permettre de guider les utilisateurs tout au long des étapes d'ECD ; il est important aussi d'adapter au mieux l'IHM à chaque classe de décideur et/ou à chaque décideur [Lajnef et al. 2005].

Nous proposons le schéma visible en figure 2 présentant le déroulement du processus d'un SIAD basé sur un processus ECD, ou encore SIAD basé Data Mining.

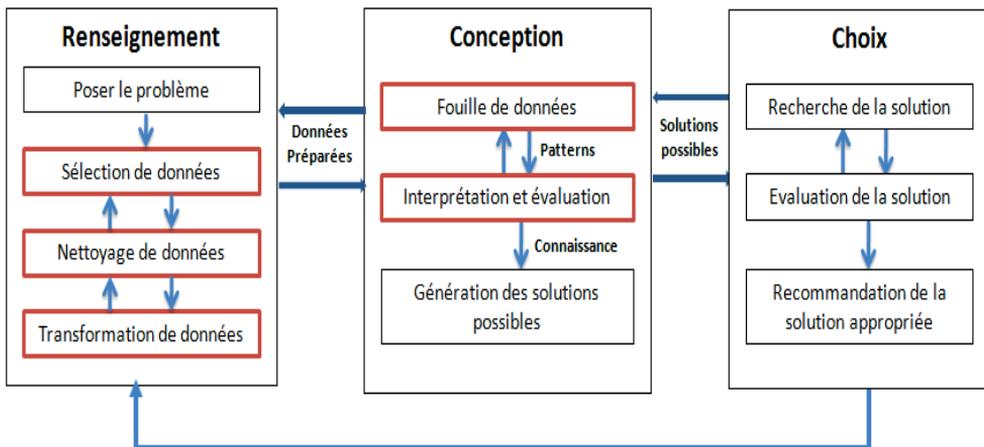


Figure 2 : processus d'un SIAD basé ECD

La figure 2 nous montre l'intégration des étapes de processus d'ECD dans le processus d'aide à la prise de décision. En effet, l'identification du problème, du processus d'ECD, permet de cerner les objectifs et définir les différents objectifs principaux du futur système. Les étapes de prétraitements consistent à construire des corpus de données spécifiques ainsi qu'à faire le nettoyage des données, le traitement des données manquantes, la sélection d'attributs ou la sélection d'instances puis la transformation de ces données. Ces étapes sont cruciales pour la recherche des informations pertinentes du processus de prise de décision [Turban 1993].

La fouille de données peut alors être opérée pour aboutir à des connaissances mises sous la forme de modèles qui doivent être validés. Des post-traitements sont nécessaires pour rendre ces modèles intelligibles soit par un humain soit par une machine [Turban 1993]. D'où la génération, l'analyse et le développement des solutions possibles au problème posé, basées sur les connaissances découvertes par le processus d'ECD.

Les premiers travaux de recherche dans le domaine ECD ont principalement porté sur le développement d'algorithmes automatiques performants. Actuellement, le rôle de l'utilisateur est devenu peu à peu un sujet de préoccupation majeur [Chevrin et al. 2007]. Ce besoin d'intégration a conduit à l'émergence de nombreux outils de visualisation et d'interaction [Fayyad et al. 2002]. Le rôle de l'utilisateur dans un processus d'ECD avait été mis en avant dès le début des années 90 dans l'article fondateur de W.J. Frawley et al. [Frawley et al. 1992]. Ces derniers insistent sur la nécessité d'intégrer l'utilisateur dans la "boucle de découverte" afin de combiner les potentialités humaines de jugement avec les capacités de calcul de la machine.

Dans ce fait, pour la modélisation et la réalisation d'un SIAD basé sur un processus interactif d'ECD, il faut fortement intégrer le facteur humain ; c'est-à-dire suivre une démarche méthodologique centrée utilisateur. Nous nous intéressons aux modèles issus du domaine des IHM qui a pour objectif de mieux prendre en compte l'utilisateur. La question qui se pose donc est la suivante : quel modèle du domaine de l'IHM faut-il choisir ?

3. LE MODELE EN U EN VUE DE SON ADAPTATION POUR LA CONCEPTION DE SIAD BASES SUR UN PROCESSUS ECD

Le développement de systèmes décisionnels interactifs, tels que les SIAD basés sur un processus d'ECD, nous amène naturellement à poser la question de la limite du système et de son interaction avec l'utilisateur. Un premier souci majeur de l'approche anthropocentrée est de travailler "sur le terrain": on n'étudie pas seulement le "travail prescrit" mais aussi le "travail réel" (au sens de [Leplat 1985]). Un autre souci est de se placer dans une logique d'utilisation et de donner à l'évaluation un rôle majeur ; l'évaluation doit pouvoir amener les concepteurs à intégrer, par itérations successives, le nouveau système décisionnel.

Du domaine de Génie Logiciel sont issus des cycles de développements classiques, de même que d'autres plus récents tels que les modèles en cascade [Royce 1970], en V [McDermid et al. 1984], spirale [Boehm 1988], en Y [André 1993] ou encore le processus unifié [Jacobson et al. 1999]. Mais ces modèles demeurent très généraux et les facteurs humains n'y sont pas explicités alors que les équipes de développement devraient être incitées à les prendre en compte. De ce constat, viennent les modèles enrichis sous l'angle de l'IHM positionnant des notions inexistantes (ou peu visibles) dans les modèles de développement qualifiés de classiques, issus du Génie Logiciel et rendant nécessaire la reconsidération de leur structure et organisation [Kolski et al. 2001].

Il existe plusieurs modèles de développement enrichis sous l'angle de l'IHM, synthétisés dans [Kolski et Loslever 1998] [Kolski et al. 2001]. Parmi ces modèles, on peut citer le modèle de [Long et al. 1990], celui de [Curtis et al. 94], en étoile [Hix et al. 1993], celui de [Collins 1995], Nabla [Kolski 1997] [Kolski 1998] ou encore en U [Abed 1990]. De tels modèles ne prétendent pas nécessairement fournir toute l'assurance qu'un projet visant la conception et le développement d'un système interactif soit réalisé avec un succès total. Le souci principal de ces modèles est surtout de souligner, du point de vue méthodologique, des aspects fondamentaux tels que la modélisation des tâches humaines, le développement itératif des prototypes et l'évaluation du système interactif [Kolski et al. 2001].

Le modèle en V fait partie des modèles les plus connus du génie logiciel. Il structure les étapes en deux phases : une descendante pour la conception et la réalisation du

système, et une ascendante pour l'intégration et l'évaluation. Les moyens d'évaluation sont définis (prévus) dans la phase descendante. Ce modèle n'est pas suffisamment spécifique pour permettre un développement efficace d'un SIAD et ne considère pas explicitement les facteurs humains. Ce modèle est un pilier du génie logiciel ; il a souvent été réutilisé et adapté dans d'autres domaines [Lepreux 2006]. Dans le domaine de l'IHM, le cycle en V a été adapté, étendu, revu pour conduire au **modèle en U** [Abed 1990] [Milot 1990] [Abed et al. 1991] [Abed 2001] [Lepreux et al. 2003] [Cathelain 2005]. Ce dernier modèle intègre des étapes centrées sur les facteurs humains et a par la suite été enrichi d'étapes permettant de mieux intégrer les experts du domaine dans la phase descendante.

De ce fait, le modèle enrichi sous l'angle de l'IHM qui nous semble le plus adéquat et qui répond aux besoins décrits dans cette section (facteurs humains, évaluation, etc.) est le modèle en U. Il a été décrit en détail dans ces différents documents ; nous nous contentons donc de le résumer ici. Comme précisé dans [Lajnef et al. 2005], nous avons choisi le modèle en U dans la mesure où il a déjà montré son applicabilité et son adaptabilité dans différents domaines complexes (contrôle aérien, applications chimiques, ferroviaires...).

Ce modèle en U situe les étapes qui n'existent pas dans les modèles classiques du génie logiciel, lesquels demeurent très généraux, tout en partant sur l'hypothèse que les facteurs humains doivent y être considérés par l'équipe de développement. Le modèle en U est structuré en deux phases (figure 3) :

- (i) **une phase descendante** qui commence par une description structurelle et fonctionnelle du système en vue de fournir un cadre structurant pour les futures activités aussi bien que pour les solutions techniques. Le résultat de cette étape est une liste des tâches devant être réparties en tâches automatiques et en tâches interactives (impliquant différents degrés de collaboration entre l'utilisateur et le système), puis analysées et modélisées. Concernant les tâches interactives, il est nécessaire de pouvoir définir l'interface homme-machine pour chacune d'entre elles en se référant au comportement probable de l'utilisateur. Il s'agit de modéliser l'aspect comportemental de l'IHM tout en prenant en compte le modèle des divers utilisateurs en termes de limites et de ressources physiques et cognitives. Le modèle utilisateur correspond à un domaine de recherche vaste et à part entière visant en particulier à comprendre le processus humain de raisonnement. Dans ce cadre, la modélisation des tâches interactives permet de révéler les besoins informationnels des utilisateurs, aussi bien que leurs besoins dans d'éventuels outils d'aide. La modélisation des tâches sert de base pour la spécification des interfaces homme-machine. Une évaluation préliminaire peut être effectuée à ce niveau pour vérifier si le modèle de tâches prescrites et le modèle de tâches réelles sont compatibles. Une fois que les besoins informationnels et les besoins d'aide ont été identifiés, il devient alors possible de définir une architecture pour le système interactif complet. Ses spécifications visent à analyser et définir, en particulier, le comportement de l'interface homme-machine. Il est très important pendant cette étape de prendre en compte un ensemble de critères ergonomiques, bien connus en dans le domaine de l'IHM (cf. à ce sujet [Bastien et Scapin 2001]). La spécification des interfaces homme-machine conduit ensuite à l'implémentation et à l'intégration du système complet.

- (ii) **une phase ascendante** focalisée sur l'évaluation du système global, selon des critères d'efficacité du système et également des critères centrés sur l'être humain. La validation consiste à confronter le modèle des tâches théoriques (prescrites) de la phase descendante avec le modèle des activités réelles mis en évidence en phase ascendante (selon les principes originaux proposés par [Abed et al. 1991]). Le résultat de cette confrontation permet de valider le système Homme-Machine ou bien de mettre en évidence ses carences ainsi qu'affiner progressivement celui-ci, particulièrement au niveau des IHM et des outils d'aide. Le modèle final résultant de la confrontation permet ainsi de généraliser des comportements spécifiques des utilisateurs dans des conditions particulières de travail. Ils sont réutilisables dans des situations relatives à des systèmes similaires. Ce modèle peut être adapté selon les caractéristiques de l'application. Par exemple, dans [Lepreux et al. 2003], il a été adapté pour des types particuliers de SIAD dans des projets dont le point central consiste en l'extraction de l'expertise des experts ; et ceci dans un but de conception de composants logiciels correspondant chacun à des modules d'aide aux activités expertes.

Le modèle en U est axé sur l'IHM, néanmoins, nous pensons qu'il est améliorable. Par exemple, il ne présente pas clairement le développement itératif et incrémental du système interactif à réaliser. Une adaptation au domaine qui nous concerne, associée à quelques améliorations, fait l'objet de la partie suivante.

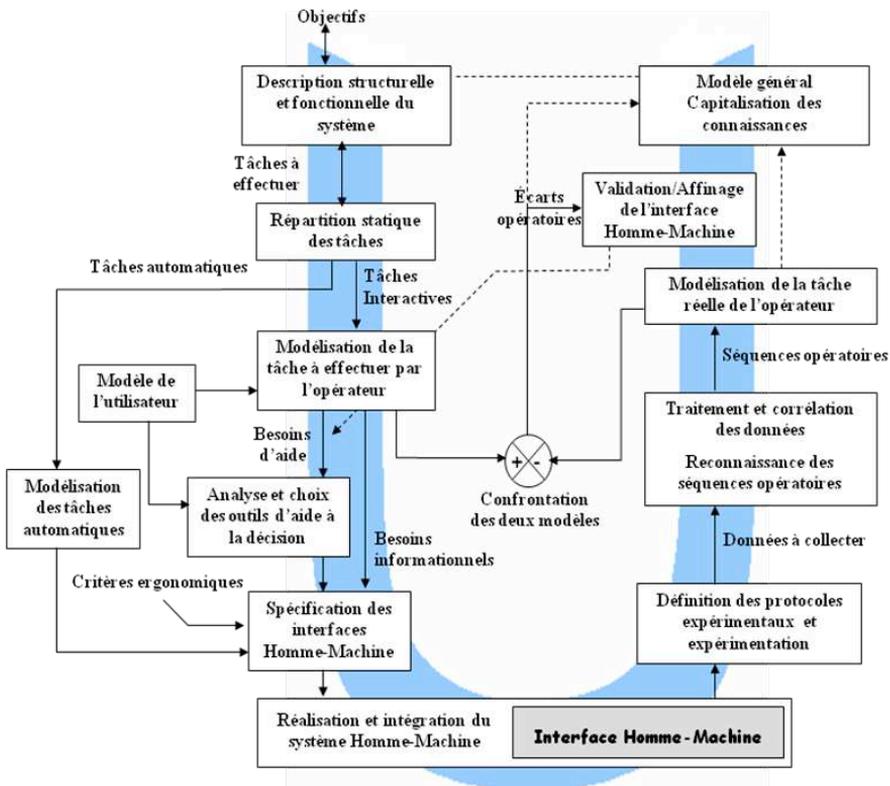


Figure 3 : Le modèle en U (version décrite dans [Abed 2001])

4. APPROCHE PROPOSEE

L'approche proposée consiste en une adaptation du modèle en U [Ltifi et al. 2008] [Ltifi et al. 2009] [Ben Ayed et al. sous presse]. Cette adaptation part du constat suivant : le modèle en U tel qu'il est présenté précédemment (figure 3), se décompose en plusieurs étapes qui montrent clairement l'intégration de l'IHM dans le processus de développement du système. Néanmoins, ce modèle possède des points possibles d'adaptation pour le contexte qui nous intéresse. Ainsi :

- (1) lors de l'étude de l'existant, il est souvent pertinent que le décideur, utilisateur potentiel du système, puisse décrire aussi bien ses caractéristiques que ses besoins fonctionnels, et évaluer et valider des premières maquettes d'interfaces en vue de montrer la manière dont il veut voir les interfaces Homme-Machines de la future application. Certaines de ces informations peuvent être utilisées pour "modéliser" le décideur (caractéristiques, préférences, stratégies...) [Robert 2003] [Seffah et al. 2009].
- (2) De plus, pour les SIAD à base d'ECD, la définition et la répartition des tâches est très importante et doit être bien mise en avant. En outre, le modèle initial ne présente pas clairement l'ordre des activités de conception (capture des besoins, analyse, conception, implémentation et test). Les étapes du modèle en U s'intéressent à la conception et à l'évaluation de la partie interactive du système tout en s'intéressant peu à l'aspect applicatif du système en question, alors que cet aspect est important pour les SIAD à base d'ECD qui concernent cette recherche.
- (3) Nous pensons aussi que l'étape "Modèle général, capitalisation des connaissances" visant à généraliser le comportement humain spécifique de l'utilisateur peut être simplifiée par un bouclage sur l'étape de modélisation de l'utilisateur et de sa tâche réelle. De ce fait nous suggérons d'intégrer ensemble les deux étapes pour une modélisation des tâches réellement effectuées.

En conséquence, le modèle en U adapté que nous proposons pour la conception et la réalisation de SIAD découlant d'un processus d'ECD est visible en figure 4. Sa description globale est la suivante :

Phase descendante du modèle en U adapté : le début de la phase descendante commence par deux étapes essentielles qui ont lieu simultanément et qui marquent le commencement du projet : (1) l'analyse du domaine d'application permet le plus souvent une première description fonctionnelle et structurelle de celui-ci ; (2) le plus précocement possible dans le projet, l'élaboration de premiers prototypes d'interfaces du SIAD visé permettent d'impliquer le plus rapidement possible les futurs utilisateurs en leur donnant un aperçu de solutions possibles (ou alternatives de solutions). Ces deux étapes sont prévues pour fournir un cadre structurant, en ce qui concerne les futures activités aussi bien que les solutions techniques. A l'issue de ces deux étapes, on obtient une définition d'un modèle de processus de travail (rejoignant ce que nous avons précisé en figure 2). Ce qui permet de cerner la liste des tâches à effectuer pour le fonctionnement du futur SIAD à réaliser, que ces tâches soient manuelles, automatiques ou interactives (cf. plus loin). Elles peuvent correspondre à des besoins fonctionnels et non fonctionnels : par exemple la nécessité de facilité d'utilisation du SIAD visé correspond à un besoin non-fonctionnel incontournable ; de même la confidentialité des données relatives aux patients est très importante dans le domaine de la santé. Le travail précédent peut être cyclique, comme le suggèrent les flèches sur le modèle en U adapté.

Les tâches prescrites suite à l'exécution des étapes précédentes, doivent être modélisées [Abed et al. 1991].

Dans cet article, nous adoptons UML (Unified Modeling Language) [Booch et al. 2000] comme langage de modélisation/conception ; ainsi pour définir les tâches du système à l'aide d'UML, nous utilisons d'abord les diagrammes de cas d'utilisation avec leur description détaillée. Après une définition des cas d'utilisation, il faut procéder à l'analyse et la conception centrée tâche du SIAD aussi bien pour la partie IHM que pour les outils d'aide basés sur le Data Mining. Chaque tâche dans le système interactif a un degré d'interactivité. Trois catégories principales de tâches peuvent être identifiées (ces catégories sont classiques en Génie Logiciel et IHM depuis les années 80) [Barthet 1988] [Breedvelt et al. 1997] [Paternò 1999] : (1) les tâches dans lesquelles seul l'utilisateur est impliqué, appelées tâches manuelles, (2) les tâches dans lesquelles seul l'aspect applicatif est représenté, appelées tâches système ou tâches automatiques, (3) les tâches impliquant des changements de degré de collaboration entre l'utilisateur et le système, appelées tâches interactives.

Les diagrammes de collaboration et de séquence d'UML permettent de spécifier et concevoir l'interface pour chaque tâche (tout en associant ces modèles à des représentations de plus en plus précises). Pour ce faire, il s'agit de se référer aux comportements probables des différents types d'utilisateur comme base : il prend la forme d'un modèle de l'utilisateur au sens large (ou « modèle utilisateur ») en termes de limites et ressources physiques et cognitives pour les différents types d'utilisateurs concernés [Yoshikawa 2003] [Robert 2003] ; certaines connaissances sur les utilisateurs sont générales et proviennent de la littérature, d'autres sont spécifiques et proviennent du terrain (interviews, observations...).

En se référant à la figure 2 et selon les spécificités du domaine, il s'agit d'analyser les différents outils d'aide à la décision en vue de définir ceux les plus adaptés au contexte du système à réaliser. La dernière étape de la phase descendante du modèle en U conduit à l'implémentation du SIAD complet ou de son prototype. Cette étape d'implémentation transforme les caractéristiques d'interfaces concrètes (définies dans les diagrammes de séquences associés à d'éventuelles maquettes plus ou moins évoluées) en représentations qui peuvent être implémentée directement par une boîte à outils graphique ou par un générateur d'interface homme-machine. Côté UML, des diagrammes de déploiement et de composants sont élaborés en vue de montrer l'évolution du système de la conception vers sa réalisation.

Phase ascendante du modèle en U adapté : l'évaluation d'un système homme-machine consiste à tester si l'utilisateur (de manière générale rappelons-le, car on peut avoir un ensemble de types d'utilisateur) peut accomplir sa tâche en utilisant l'interface fournie. Deux propriétés sont habituellement explorées pour de telles évaluations : l'utilité et l'utilisabilité, elles-mêmes décomposées en un ensemble de critères bien connus en IHM [Senach 1990] [Nielsen 1993]. Il existe un grand nombre de méthodes utilisables pour la mise en œuvre des évaluations [Senach 1990] [Wilson et Corlett 1996] [Bastien et Scapin 2001] [Ivory et Hearst 2001] [Ezzedine et Kolski 2004] [Huart et al. 2008] : observations, interviews, mouchard électronique, analyse des traces, oculométrie, questionnaires, etc. Dans cette phase d'évaluation, nous nous concentrons généralement sur l'exécution des tâches en lien avec les éléments visibles en figure 2 : (1) d'une part selon le comportement de l'utilisateur pendant l'interaction avec le système (difficultés rencontrées, temps requis pour accomplir une tâche, exactitude du

résultat, nombre et type d'erreurs, opinion, etc.), (2) et d'autre part, selon le système en termes de différences entre les objectifs et résultats obtenus.

Cette phase ascendante commence par la définition rigoureuse des protocoles expérimentaux (sujets, déroulement, situations et tâches concernées, IHM et aides impliquées, données à collecter...) [Abed 2001]. Une fois collectées, les données sont traitées en fonction des principes opérationnels dégagés de l'étape précédente. Il s'agit de mettre en corrélation les données avec l'activité humaine qui a été observée, et ceci pour l'ensemble du processus décisionnel mis en évidence en figure 2. Des séquences opératoires sont ainsi mises en évidence. Ce travail permet de reconstituer (modéliser) progressivement les activités réelles (appelées aussi tâches réelles). L'expérience montre que ces tâches réelles peuvent être sensiblement éloignées des tâches prescrites, prévues initialement dans la phase descendante.

Un principe fondamental du modèle en U est justement la confrontation entre les tâches réelles et les tâches prescrites [Abed 1990] [Millot 1990]. Le résultat de la confrontation conduit, soit à valider le système complet (IHM et aides), soit à révéler ses imperfections pour l'améliorer progressivement. Des retours vers différentes étapes de la démarche descendante sont donc nécessaires, selon l'ampleur des améliorations à apporter au système.

Dans la figure 4, le sens des flèches indique le sens d'envoi d'informations (modèles, prototypes, tâches, etc.) entre les différentes étapes du modèle en U adapté.

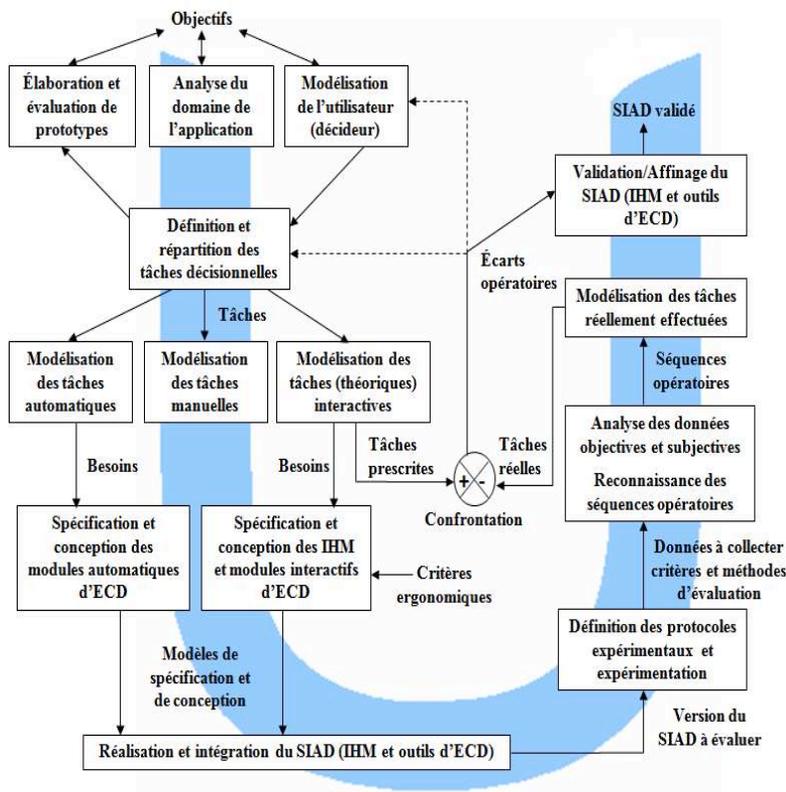


Figure 4 : modèle en U adapté [Ltifi et al. 2008] [Ltifi et al. 2009] [Ben Ayed et al. sous presse]

L'approche de conception du processus d'ECD doit être adaptée dès le début d'un projet afin de satisfaire les besoins des décideurs. En s'inspirant du travail de [Lajnef et al. 2005], nous proposons des similitudes méthodologiques entre le domaine de l'interaction homme-machine et celui de l'Extraction de connaissances (figure 5) :

- (a) Les objectifs principaux du processus ECD sont ceux du futur système décisionnel interactif à concevoir. En effet, il s'agit d'identifier l'ensemble des cas auxquels l'outil de fouille de données vise à répondre : il s'agit de la spécification des exigences du système interactif à concevoir. Une fois les objectifs en termes d'ECD identifiés, il s'agit de les classer puis de décomposer chaque problème en sous-problèmes – en fait, on considère que beaucoup de problèmes complexes peuvent être divisés en sous-problèmes plus simples à résoudre [Garlati 1996].

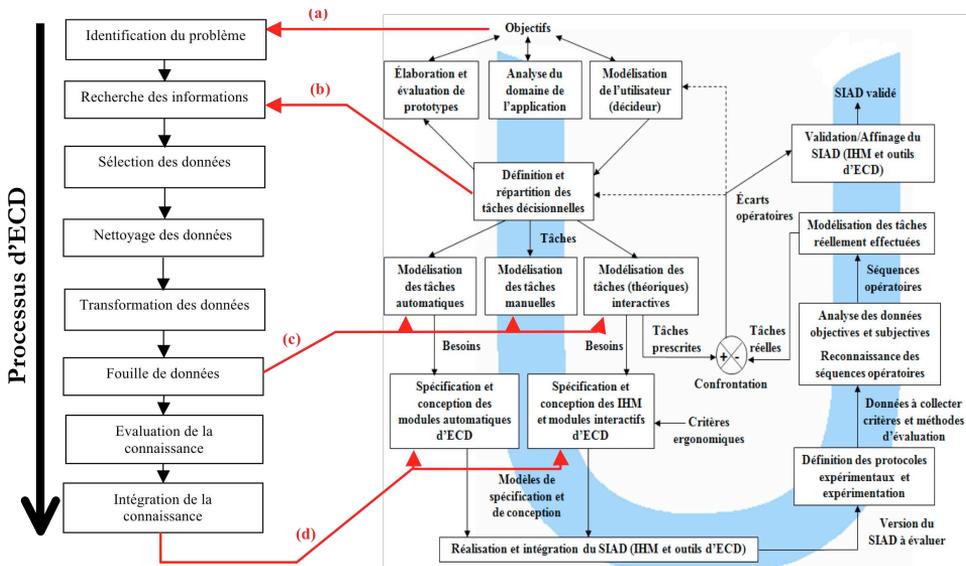


Figure 5 : Le modèle en U adapté au processus d'ECD [Ltifi et al. 2008]

- (b) La définition des tâches décisionnelles permet de donner une idée sur les données pertinentes qu'il faut rechercher, sélectionner (trier, regrouper ou répartir ces données selon les tâches définies), nettoyer et transformer (en réalisant divers calculs tels que : totaux, moyennes, écarts types, comparatif d'une période à l'autre, etc. ; en présentant les résultats d'une manière synthétique ou détaillée, le plus souvent graphique selon leurs besoins ou les attentes des différents types d'acteurs concernés de l'entreprise, etc.).
- (c) Une fois les données bien préparées, elles peuvent être fouillées pour extraire des connaissances. Une ou plusieurs techniques peuvent être utilisées permettant d'extraire les acteurs pertinents et de construire le modèle initial (un réseau de neurones, un arbre de décision, etc.). L'étude de cette tâche est très importante pour le concepteur du système décisionnel. Une telle activité peut se réaliser au niveau de l'étape de modélisation de la tâche afin de délivrer un modèle de la tâche prescrite. Les résultats extraits peuvent aider à spécifier les modules du système à concevoir.

(d) La connaissance extraite et évaluée doit être intégrée dans les modules du SIAD à spécifier et concevoir. La présentation de cette connaissance doit répondre aux critères ergonomiques bien connus dans le domaine de l'IHM afin d'assurer la meilleure interaction possible entre le décideur et la machine [Lepreux et al. 2003].

Le déroulement des différentes étapes du modèle en U et du processus d'ECD se fait d'une manière parallèle. Les flèches de la figure 5 présentent les relations entre les différentes étapes du modèle en U adapté basé sur les étapes du processus d'ECD.

Dans la partie suivante, il est expliqué comment cette démarche est mise en application dans le domaine médical en vue du développement d'un SIAD pour la lutte contre les infections nosocomiales.

5. CONTEXTE APPLICATIF

Dans la partie précédente nous avons présenté notre proposition de démarche centrée utilisateur adaptée au processus d'ECD. Elle a été appliquée à un cas concret dans le domaine médical. Le SIAD en question, basé sur un processus d'ECD s'appuie pour l'étape de fouille de données sur le raisonnement à partir de cas [Riesbeck et Shank 1989] ; ce SIAD est en cours d'utilisation et d'évaluation dans le service de réanimation du Centre Hospitalo-universitaire Habib Bourguiba à Sfax, Tunisie. Il a pour but d'aider les médecins du service, utilisateurs du système, à comprendre, prédire et prévenir les infections nosocomiales.

Les infections nosocomiales (IN) représentent un des problèmes majeurs de la santé publique. Ce sont des infections contractées dans un établissement de soins. Une infection est considérée comme telle lorsqu'elle était absente au moment de l'admission du patient [Garner et al. 1988]. Lorsque l'état infectieux du patient à l'admission est inconnu, l'infection est classiquement considérée comme nosocomiale si elle apparaît après un délai de 48 heures d'hospitalisation. Une infection nosocomiale peut être localisée dans un organe. Les organes les plus touchés sont les reins et le cœur. Ces infections peuvent être directement liées aux soins ou simplement survenir lors de l'hospitalisation indépendamment de tout acte médical. Dans les services de réanimation le problème des infections nosocomiales est d'autant plus préoccupant que les patients qui y sont hospitalisés sont plus fragiles. La lutte contre les IN est un problème complexe. Plusieurs travaux ont proposé des systèmes informatiques de surveillance des IN basés sur les techniques de fouille de données [Brossette et al. 1998] [Brossette et al. 2000]. Les travaux montrent leur efficacité et leur capacité à produire des règles utiles. Mais, telles que ces règles sont décrites dans les articles, leur utilisation par des médecins paraît difficile.

Une enquête a été réalisée dans l'hôpital CHU Habib Bourguiba à Sfax, Tunisie pour mesurer un jour donné (entre le 17 avril 2002 (minuit) et le 18 avril 2002 (minuit)) la prévalence de la survenue d'infections nosocomiales [Kallel et al. 2005]. Cette étude a montré que 17,9 % des 280 patients hospitalisés le jour de la prévalence dans l'ensemble de l'hôpital ont été victimes d'une IN. Les patients hospitalisés dans le service de réanimation sont les plus touchés (77,2 %). Notre projet est en lien avec les médecins du service de réanimation de cet hôpital. Ces médecins sont les experts et futurs utilisateurs du SIAD sur lequel nous travaillons.

5.1. Modules visés couvrant le processus d'ECD

Pour développer le SIAD proposé basé sur l'ECD, nous avons suivi les différentes phases du processus d'ECD (figure 1) ; le système a été subdivisé en modules en se

référant aux étapes de processus d'ECD. Deux modules (applications) ont été conçus et réalisés : un pour le stockage et la préparation des données ; et un pour la fouille de données proprement dite. L'algorithme de fouille que nous avons choisi est une version améliorée de l'algorithme dit des k plus proches voisins (kPPV) utilisé dans la technique du raisonnement à partir de cas (RBC) [Ben Ayed et al. 2006].

Dans ce qui suit, nous allons décrire le déroulement de la démarche relativement à chacun de ces deux modules du SIAD en nous inspirant de [Ben Ayed et al. sous presse].

5.2. Déroulement de la démarche relativement au premier module

Les tâches effectuées dans le cadre du premier module sont décrites dans le tableau 1. Pour celui-ci, nous avons surtout étudié le contexte et défini les objectifs en organisant des réunions fréquentes avec le médecin qui s'occupe du problème des IN, et ayant le double rôle de décideur et d'utilisateur. Parfois il invitait ses collaborateurs à assister aux réunions. Cette collaboration s'est avérée très précieuse, et a permis la définition et la répartition des tâches.

L'implémentation a concerné la mise en place de la base de données ainsi que son prétraitement (nettoyage et transformation des données) et des interfaces homme-machine de saisie des données. Les premiers tests des versions préliminaires des interfaces de saisie ont été très utiles pour le reste du processus de développement. L'intérêt qu'a montré l'utilisateur relativement à notre méthode lui a permis de bien exprimer les évolutions de ses objectifs.

L'interface Homme-Machine de la Figure 6, donne l'accès à l'utilisateur pour saisir les informations relatives à l'hospitalisation du patient (durée de séjour, diagnostics, prise antibiotiques, etc.). Cette interface a été validée par l'utilisateur après plusieurs itérations.

5.3. Déroulement de la démarche relativement au deuxième module

Une fois la réalisation du module de "Stockage et préparation des données" terminée, nous avons entamé celle du module de "Fouille de données" (tableau 2). L'objectif initial de cette application est de fournir au médecin une prédiction sur la probabilité d'apparition d'une IN chez un patient entrant dans le service. Pour la fouille de données, en nous conformant aux besoins exprimés par les experts (utilisateurs), nous avons développé une application pour la prédiction de la survenue d'IN, se basant sur la technique du raisonnement à partir de cas [Kolodner 1996] [Ben Ayed et al. 2006].

| | Étapes du Modèle en U adapté | Tâches effectuées |
|--------------------|--|---|
| Etape (a) (Fig. 5) | Analyse du domaine d'application | Etude des infections nosocomiales (définition, causes, risques...) |
| | Premiers prototypes d'interfaces. Description des besoins fonctionnels et structurels du système et définition des tâches | Description structurelle et fonctionnelle du domaine d'application médical déduite après consultation des futurs utilisateurs et élaboration de prototypes d'interfaces Homme-Machine en déterminant leur architecture générale. Les médecins proposent des fenêtres en onglets. |
| | Modélisation de l'utilisateur | Une modélisation de l'utilisateur (type unique pour le moment) est réalisée : il est expert dans son domaine ; il peut être considéré comme expert aussi concernant l'utilisation des outils informatiques (plus de 3 heures par jour devant son PC). |
| Etape (b) (Fig. 5) | Définition et répartition des tâches décisionnelles | Une première description des tâches manuelles (remplissage de fiches remplies par les médecins internes, contenant les données nécessaires à l'étude, au chevet des malades), automatiques (calcul de l'âge et d'un paramètre de risque de décès) et interactives (choix des libellés des zones de saisie, ...) est faite à l'aide des cas d'utilisation d'UML. |
| Etape (c) (Fig. 5) | | <i>Cette étape n'entre pas dans le cadre du premier module du SIAD en question</i> |
| Etape (d) (Fig. 5) | | <i>Cette étape n'entre pas dans le cadre du premier module du SIAD en question</i> |
| Réalisation | Réalisation et intégration du système Homme-Machine | <ol style="list-style-type: none"> 1. Implémentation de la base de données en utilisant le SGBD Oracle® et ses outils (SQLplus,...). 2. Nettoyage de la base de données (valeurs manquantes, répétitions, erreurs de données, etc.) 3. Transformation de la table contenant les informations d'hospitalisation des patients pour les préparer à la fouille (transformation de la date de naissance des patients en âge, etc.) 4. Implémentation des interfaces Homme-Machine de saisie pour entrer les données nécessaires pour appliquer la fouille de données (figure 6). 5. Correction des derniers bugs liés aux derniers tests utilisateurs |
| Validation | Branche ascendante du modèle en U adapté | <ol style="list-style-type: none"> 1. Mise en œuvre de tests utilisateurs : mise en évidence de problèmes d'utilisabilité au niveau des saisies, mise en évidence d'autres besoins pour les médecins, ajout de boutons, etc. 2. Mise en œuvre de tests utilisateurs selon une démarche de vérification des caractéristiques d'une IHM de qualité (cohérence, prévention des erreurs, ...) et validation finale. |

Tableau 1 : Déroulement du développement du module de stockage et préparation de données

Figure 6 : Extrait de l'interface Homme-Machine de saisie de données concernant le patient

Le raisonnement à partir de cas est une approche de résolution de problèmes basée sur la réutilisation par analogie d'expériences passées appelées *cas*, au cours d'un cycle de raisonnement [Kolodner 1996]. Un cas est composé de deux parties décrivant un problème et la solution qui lui a été appliquée. Un cas est une expérience passée permettant au système de résoudre des problèmes plus efficacement ou d'éviter les échecs passés : un cas doit donc être utile pour le raisonnement. Dans ce but, un ensemble de caractéristiques est extrait ou sélectionné dans la représentation de chaque cas. Ces caractéristiques déterminent dans quelle situation un cas est applicable et utile.

Ces caractéristiques doivent être prédictives, c'est-à-dire qu'elles doivent être responsables de l'élaboration de la solution mise en œuvre dans le cas et permettre ainsi de déterminer dans quelle mesure une solution est applicable pour un autre problème. Les caractéristiques doivent également être suffisamment abstraites pour permettre la réutilisation des enseignements du cas pour différents problèmes, tout en restant suffisamment concrètes pour être facilement identifiables dans un nouveau problème. Lorsqu'un nouveau problème est à résoudre, un cas cible est construit avec la partie solution inconnue, celle-ci devant être apportée par le raisonnement. Les cas sources représentent des expériences passées qui sont stockées dans une mémoire. Les cas sources et le cas cible partagent le même formalisme de représentation. L'objectif du raisonnement est de transférer des enseignements pertinents des cas sources pour élaborer la solution du cas cible.

| | Etapes du Modèle en U adapté | Tâches effectuées |
|-------------|--|---|
| Etape (a) | | <i>Cette étape n'entre pas dans le cadre du deuxième module du SIAD en question</i> |
| Etape (b) | | <i>Cette étape n'entre pas dans le cadre du deuxième module du SIAD en question</i> |
| Etape (c) | Modélisation des tâches automatiques | Utilisation des diagrammes de collaboration d'UML pour modéliser les tâches automatiques relatives à l'application de la technique KPPV de la fouille de données Modélisation de l'intégration de la connaissance extraite dans la prise de décision |
| | Modélisation des tâches manuelles | <i>Sans objet dans notre cas</i> |
| | Modélisation des tâches (théoriques) interactives | Modélisation des interactions de l'utilisateur avec la machine (donner la valeur de « K » pour l'application de l'algorithme de fouille, nombre de combinaisons possibles, etc.) |
| Etape (d) | Spécification et conception des modules auto d'ECD | Utilisation des diagrammes UML pour la spécification: - des modules de calcul (âges, durées de séjour, risques de décès, ...) - des transactions avec le SGBD. |
| | Spécification et conception des IHM et modules interactifs d'ECD | Spécification des modules interactifs en utilisant les diagrammes de séquence d'UML |
| Réalisation | Réalisation et intégration du système Homme-Machine | <ol style="list-style-type: none"> 1. Implémentation des interfaces de fouille de données (développement de code source relatif à l'algorithme de KPPV). La fouille se fait en deux étapes suivant deux IHM (figures 7 et 8) 2. Correction des défauts résiduels dégagés par l'utilisateur du système. |
| Validation | Branche ascendante du modèle en U adapté | <p>En vue de la validation et l'affinage du module de fouille de données et l'évaluation de la connaissance extraite, des tests d'évaluations ont été planifiés et réalisés selon trois critères [Nielsen 1993] :</p> <ol style="list-style-type: none"> 1. Utilité (tests fonctionnels) : il s'agit de vérifier si l'interface homme-machine est utile c'est-à-dire qu'elle fournit les fonctions nécessaires à l'utilisateur. Ce type de test nous permet de vérifier si l'algorithme de KPPV développé répond bien aux besoins de l'utilisateur (validité du résultat, temps d'exécution, etc.). 2. Utilisabilité (tests ergonomiques) : elle concerne la facilité de manipulation et rend compte de la qualité de l'interface en termes de facilité d'apprentissage et d'utilisation. Les médecins ont la possibilité de guider l'application de l'algorithme de KPPV en choisissant la valeur k (1, 3 ou 5) avec un simple clic. En plus ils ont été très intéressés par l'exactitude des résultats présentés (Fig. 7 et 8). 3. Acceptabilité : il s'agit de vérifier si l'interface est compatible avec les valeurs et la culture du service de réanimation de l'hôpital ; on parle dans ce contexte des sigles de la médecine et du service concerné ainsi que des couleurs qui doivent varier entre le bleu et le blanc, etc. <p>En fait, des premiers prototypes de l'application présentent des interfaces homme-machine avec un fond beige qui a été refusé par les médecins. Une réunion a été organisée avec ces derniers pour consulter la palette de couleurs en vue de choisir la couleur qui leur convient le plus ; c'est la couleur courante de l'interface.</p> <p>Des derniers tests ont été réalisés pour l'évaluation du système complet.</p> |

Tableau 2 : déroulement du développement du module de fouille de données

Les interfaces homme-machine visibles en figure 7 et figure 8 montrent l'application de l'algorithme de la fouille de données, qui est le classifieur supervisé KPPV (les k-plus-proches-voisins).

Caractéristiques des cas sources

Ce bouton permet d'appliquer l'algorithme de fouille et de calculer les K plus proches voisins

Ce bouton permet de passer à la deuxième interface pour la prédiction

Ce bouton permet d'appliquer l'algorithme de fouille à un autre patient

Figure 7 : Extrait de l'interface Homme-Machine de l'étape 1 de fouille de données (définitions des KPPV)

Le principe de l'algorithme KPPV consiste à identifier comme son nom l'indique les k plus proches voisins. Pour ce faire, il faut calculer les distances entre les caractéristiques du patient concerné par l'étude et celles des patients déjà hospitalisés et enregistrés dans la base de données du service de réanimation. La distance que nous avons utilisée pour calculer les plus proches voisins est la distance Euclidienne. Le K définit le nombre de plus proches voisins à considérer pour l'élaboration de la décision du classifieur ; dans le but d'avoir le meilleur taux de discrimination, nous avons choisi trois valeurs de K qui sont (1, 3, 5), pour tester les classifieurs 1-PPV, 3-PPV et 5-PPV.

Pour la deuxième interface, il s'agit de calculer la probabilité d'avoir une IN en fonction des k plus proches voisins. La présentation de cette probabilité est faite en fonction de sa valeur. Si elle est supérieure à 50% c'est qu'il existe un grand risque d'avoir une infection nosocomiale ; pour cette raison la valeur de la probabilité se présente en rouge gras et avec une image d'alerte.

Ces interfaces Homme-Machine ont fait l'objet de plusieurs itérations, démonstrations et évaluations avec les utilisateurs.

Caractéristiques du cas cible

Nouveau patient

| Dossier médical | Durée de séjour | Age | Catégorie | Sexe | SABS II | Prise ATB | Durée VA | Durée KTC | InNos |
|-----------------|-----------------|-----|-----------|------|---------|-----------|----------|-----------|-----------|
| 518/06 | 6 | 63 | Médicale | B | 27 | Oui | 7 | 0 | Oui (60%) |

Caractéristiques des plus proches voisins

| | | | | | | | | | | |
|-------------------|--------|----|----|----------|---|----|-----|----|---|-----|
| Dossier médical 1 | 549/06 | 14 | 74 | Médicale | B | 43 | Oui | 13 | 0 | Oui |
| Dossier médical 2 | 439/06 | 15 | 63 | Médicale | B | | Oui | 15 | 0 | Oui |
| Dossier médical 3 | 397/06 | 17 | 73 | Médicale | B | 38 | Oui | 8 | 0 | Oui |
| Dossier médical 4 | 555/06 | 5 | 81 | Médicale | B | 47 | Oui | 4 | 0 | Non |
| Dossier médical 5 | 507/06 | 4 | 80 | Médicale | B | 50 | Oui | 4 | 0 | Non |

Identifiants des voisins

| | | |
|-------------------|----------|---|
| Dossier médical 1 | 00/45801 | A |
| Dossier médical 2 | 06/19038 | B |
| Dossier médical 3 | 06/19563 | C |
| Dossier médical 4 | 06/28485 | D |
| Dossier médical 5 | 06/25738 | E |

Fonction de combinaison

| | 1 | 2 | 3 | 4 | 5 |
|--------------------------|-------------|-------------|-------------|-------------|-------------|
| Nbre de voisins retenus | 1 | 2 | 3 | 4 | 5 |
| Identifiants des voisins | A | A B | A B C | A B C D | A B C D E |
| Réponses des voisins | 0 | 0 0 | 0 0 0 | 0 0 0 N | 0 0 0 N N |
| Décompte des réponses | Oui:1 Non:0 | Oui:2 Non:0 | Oui:3 Non:0 | Oui:3 Non:1 | Oui:3 Non:2 |
| Valeur retenu | Oui | Oui | Oui | Oui | Oui |
| Évaluation | 100% | 100% | 100% | 75% | 50% |

Combinaison [X] Quitter

Ce bouton permet de calculer la probabilité d'acquisition d'une IN en fonction des caractéristiques des K plus proches voisins (dans ce présent cas: k= 5 patients dont les 3 premiers ont acquis une IN. Donc le calcul des combinaisons donne une probabilité de 60%).

Figure 8 : Extrait de l'interface Homme-Machine de l'étape 2 de fouille de données (calcul de la probabilité d'IN)

6. CONCLUSION

Cet article a proposé une approche centrée utilisateur de conception de système interactif d'aide à la décision basé sur un processus d'ECD. Depuis plusieurs dizaines d'années, une importante masse d'informations est stockée sous forme informatique dans les entreprises. Les systèmes d'information sont destinés à garder la trace d'événements de manière fiable et intègre. Ils automatisent de plus en plus les processus opérationnels, particulièrement, d'aide à la décision. Pour ce faire, l'extraction de connaissances à partir de données (ECD) est utilisée comme outil décisionnel permettant de découvrir, à partir de bases de données, des connaissances auparavant inconnues et potentiellement utiles pour la prise de décision. Un SIAD basé sur l'ECD est hautement interactif. En effet, comme la prise de décision implique l'acteur humain, qu'est le décideur, le processus d'ECD nécessite de sa part une interactivité Homme-Machine pour se réaliser. Une modélisation d'un tel système doit donc selon nous faire appel aux méthodes, modèles et critères de l'IHM. D'où la proposition d'une approche

de développement qui s'appuie sur le modèle en U. Une adaptation de celui-ci au processus d'ECD a fait l'objet de cet article.

L'approche proposée est actuellement mise en œuvre depuis plus de 2 ans. En effet, le SIAD en question est en cours d'évaluation et d'utilisation dans le service de réanimation du Centre Hospitalo-universitaire Habib Bourguiba à Sfax, Tunisie. Il a pour but d'aider les médecins du service, utilisateurs du système, à comprendre (en se référant à l'historique des patients enregistrés dans la base de données de l'hôpital), prédire (par l'application d'un algorithme de fouille de données) et par la suite (en fonction du résultat de prédiction) à prévenir les infections nosocomiales ; infections contractées par les patients pendant leur hospitalisation. Le SIAD développé doit faire l'objet dans les mois à venir d'un ensemble d'expérimentations.

Sous l'angle de l'IHM et en nous basant sur cette expérience dans le domaine médical, on se trouve confronté au besoin de développer une méthodologie spécifique pour l'évaluation de SIAD à base de l'ECD en s'inspirant des critères, méthodes et techniques d'évaluation issues conjointement du domaine de l'IHM et de celui de la fouille visuelle de données (visual data mining) ; c'est-à-dire avec intégration du traitement visuel d'information dans toutes les étapes du processus ECD. Cette fouille visuelle permet de présenter les données et les connaissances dans une forme visuelle permettant à l'utilisateur d'interpréter les données, tirer les conclusions ainsi qu'interagir directement avec ces données. On considère que les techniques de visualisation améliorent les techniques d'ECD courantes en augmentant l'implication de l'utilisateur et sa confiance à propos des constatations découvertes [Keim 2002]. Une telle méthodologie d'évaluation doit permettre d'étudier l'expérience cognitive et affective des utilisateurs du système d'aide à la décision pour la lutte contre les infections nosocomiales [Yoshikawa 2003].

REMERCIEMENTS

Les auteurs tiennent à remercier le Docteur H. Kallel pour l'intérêt qu'il a accordé au projet et le temps passé pour aider à concevoir, implémenter et évaluer le système d'aide à la décision. Ils remercient également les relecteurs anonymes de la revue JIPS pour leurs remarques constructives.

REFERENCES

- Abed, M. 1990. Contribution à la modélisation de la tâche par outils de spécification exploitant les mouvements oculaires: application à la conception et à l'évaluation des interfaces homme-machine. Thèse de doctorat, Université de Valenciennes et du Hainaut-Cambrésis, France.
- Abed, M., Bernard, J.M. et Angué, J.C. 1991. Task analysis and modelization by using SADT and Petri Networks. Proceedings Tenth European Annual Conference on Human Decision Making and Manual Control. Liege.
- Abed, M. 2001. Méthodes et Modèles formels et semi-formels de conception et évaluation des systèmes homme-machine. Mémoire d'HDR, Université de Valenciennes et du Hainaut-Cambrésis, France.
- André, J. 1993. Le cycle de vie en Y. Arche SQL.
- Barthet, M.F. 1988. Logiciels interactifs et ergonomie, Modèle et méthodes de conception, Dunod informatique, Paris.

- Bastien, C. et Scapin, D. 2001. Evaluation des systèmes d'information et Critères Ergonomiques. In Kolski C. (dir.), Environnements évolués et évaluation de l'IHM, Interaction homme-machine pour les SI 2, Hermès, Paris, 53-80.
- Ben Ayed, M., Feki, I. et Alimi, A.M. 2006. Optimisation de la technique de RBC pour la classification dans un processus de data mining. Atelier « Fouille de données complexes », Lille, France.
- Ben Ayed, M., Ltifi, H., Kolski, C. et Alimi, A.M. sous presse. A User-centered Approach for the Design and Implementation of KDD-based DSS: A case Study in the Healthcare Domain. *Decision Support Systems*, accepté et à paraître.
- Booch, G., Rumbaugh, J. et Jacobson, I. 2000. Le guide de l'utilisateur UML. Eds Eyrolles, Paris.
- Boehm, B. 1988. A Spiral Model of Software Development and Enhancement. *IEEE Computer*, 21 (5), 61-72.
- Breedvelt, I., Paternò, F., et Sereni, C. 1997. Reusable Structures in Task Models, *Proceedings Design, Specification, Verification of Interactive Systems*. Springer Verlag, 251-265.
- Brossette, S.E., Sprague, A.P., Hardin, J.M., Waites K.B., Jones W.T. et Moser S.A. 1998. Association Rules and Data Mining in Hospital Infection Control and Public Health Surveillance. *Journal of the American Medical Informatics Association*, 5, 4, 373-381.
- Brossette, S.E., Sprague, A.P., Jones, W.T., et Moser, S.A. 2000. A data mining system for infection control surveillance. *Methods Inf Med.*, 39, 4-5, 303-310.
- Cathelain, S. 2005. Contribution à la conception de systèmes coopératifs. Application au domaine du contrôle de trafic aérien. Thèse de doctorat, Université de Valenciennes et du Hainaut-Cambrésis, France.
- Chevrin V., Coiturier O., Mephu Nguifo E., et Rouillard J. 2007. Recherche anthropocentrée de règles d'association pour l'aide à la décision. *Revue d'Interaction Homme-Machine*, 8, 2, 117-146.
- Curtis, B., et Hefley, B. 1994. A WIMP No More: The Maturing of User Interface Engineering. *Interactions*, 23-34.
- Collins, D. 1995. *Designing Object-oriented user interfaces*. Readwoods City, CA: Benjamin/Cummings Publishing Inc.
- Ezzedine, H., et Kolski, C. 2004. Démarche d'évaluation d'IHM dans les systèmes complexes, application à un poste de supervision du trafic ferroviaire. *Revue d'Interaction Homme-Machine (RIHM)*, 5, 91-122.
- Evans, R.S. 2007. Case studies in clinical decision support: LDS hospital experience. *Clinical Decision Support*, 6, 143-167.
- Fayyad, U.M., Djorgovski, S.G. et Weir, N. 1996. Automating the Analysis and Cataloging of Sky Surveys. *Advances in Knowledge Discovery and Data Mining*. MIT Press, 471-494.
- Fayyad, U.M., Grinstein, G.G., et Wierse, A. 2002. Information visualization in data mining and knowledge discovery. Morgan Kaufmann.
- Frawley, W.J., Piatetsky-Shapiro, G., et Matheus C.J. 1992. Knowledge discovery in databases: an overview. *AI Magazine*, 57-70.
- Garlatti, S. 1996. Multi-media and interactive decision support systems in complex situation. In 43th meeting of the European working group "Multicriteria Aid for Decisions", Brest.
- Garner, J.S., Jarvis, W.R, Emori, T.G., Hogan, T.C. et Hugues J.M. 1988. CDC definitions for nosocomial infections. *Am. J. of Inf. Contr.*, 16, 3, 128-140.

- Guigue, L., et Donadey, T. 1999. Infections et systèmes d'aide à la décision. *Médecine et maladies infectieuses*, 29 (12), 758-767.
- Hartson, H.R., et Hix, D. 1989. Towards empirically derived methodologies and tools for Human-Computer development. *International Journal of Man-Machine Studies*, 31, 477-494.
- Hix, D., et Hartson, H.R. 1993. *Developing user interfaces: Ensuring usability through product & process*. Wiley professional computing, John Wiley & Sons, inc., New York, USA.
- Huart, J., Kolski, C., Bastien, C. 2008. L'évaluation de documents multimédias, état de l'art. In Merviel S. (Ed.), *Objectiver l'humain ? Volume 1, Qualification, quantification*, Hermes, Paris, pp. 211-250.
- Ivory, M. et Hearst, M. 2001. The State of the Art in Automated Usability Evaluation of User Interfaces. *ACM Computing Surveys*, 33, 4, 173-197.
- Jambu, M. 1999. *Introduction au Data Mining, analyse intelligente des données*. Eds Eyrolles, Paris.
- Jacobson, I., Booch, G., et Rumbaugh J. 1999. *Le processus Unifié de Développement logiciel*. Eds Eyrolles, Paris.
- Kallel, H., Bouaziz, M., Ksibi, H., Chelly, H., Hmida, C.B., Chaari, A., Rekik, N. et Bouaziz M. 2005. Prevalence of hospital-acquired infection in a Tunisian Hospital. *Journal of Hospital Infection*, 59, 343-347.
- Kaplan B. 2001. Evaluating informatics applications-clinical decision support systems literature review. *International Journal of Medical Informatics*, 64, 15-37.
- Kawamoto, K., Houlihan C.A., Balas E.A. et Lobach, D.F. 2005. Improving clinical practice using clinical decision support systems: A systematic review of trials to identify features critical to success. *British Medical Journal*, 330, 765-768.
- Keim, D. 2002. Information Visualization and Visual Data Mining. *IEEE Transactions on Visualization and Computer Graphics*, 8 (1), 1-8.
- Klein, M. et Methlie, L.B. 1990. *Expert Systems: A Decision Support Approach with applications in management and finance*. Eds Addison-Wesley Publishing Company.
- Kolodner, J.L., et Leake, D.B. 1996. *A Tutorial Introduction to Case-Based Reasoning: Experiences. Lessons & Future Directions*, AAAI press, MIT press, 31-65.
- Kolski, C. 1997. *Interfaces Homme-Machine, application aux systèmes industriels complexes (2ème édition revue et étendue)*. Editions Hermès, Paris.
- Kolski, C. 1998. A call for answers around the proposition of an HCI-enriched model. *ACM SIGSOFT Software Engineering Notes*, 23 (3), 93-96.
- Kolski, C., Ezzedine, H. et Abed, M. 2001. Cycle de développement du logiciel : des cycles classiques aux cycles enrichis sous l'angle des interactions homme-machine. In C. Kolski (Ed.), *Analyse et conception de l'I.H.M., Interaction Homme-Machine pour les S.I. 1*, Eds Hermes, Paris, 23-49.
- Kolski, C., Loslever P. 1998. An HCI-enriched model for supporting human-machine systems design and evaluation. *Proceedings 7th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design and Evaluation of Man-Machine Systems*, Kyoto, Japan, September.
- Lajnef, M., Ben Ayed, M. et Kolski, C. 2005. Convergence possible des processus du data mining et de conception-évaluation d'IHM : adaptation du modèle en U. *Proceedings of IHM'05, International Conference Proceedings Series*, ACM Press, Toulouse, 243-246.

- Lefébure, R. et Ventari, G. 2001. Data Mining : Gestion de la relation client, Personnalisation des sites Web. Eds Eyrolles.
- Leplat, J. 1985. Erreur humaine, fiabilité humaine dans le travail. Paris, A. Colin.
- Lepreux, S., Abed, M. et Kolski, C. 2003. A human-centred methodology applied to decision support system design and evaluation in a railway network context, *Cognition Technology and Work*, 5, 248-271.
- Lepreux, S. 2005. Approche de Développement centré décideur et à l'aide de patrons de Systèmes Interactifs d'Aide à la Décision. Thèse de doctorat, Université de Valenciennes et du Hainaut-Cambrésis.
- Lepreux, S. 2006. Supports pour la prise en compte des experts et utilisateurs dans le développement de Systèmes Interactifs d'Aide à la Décision. *Ergo-IA'06*, 139-146.
- Lepreux, S. 2007. Réflexions en vue d'une meilleure prise en compte des experts et utilisateurs dans le développement de SIAD : apports des patrons et de la composition d'IHM. *Revue d'Interaction Homme-Machine (RIHM)*, 8, 1, 61-90.
- Ltifi, H., Ben Ayed, M., Kolski, C., et Alimi, A.M. 2008. Prise en compte de l'utilisateur pour la conception d'un SIAD basé sur un processus d'ECD", D. Galaretta, P. Girard, J.C. Tucoulou, M. Wolff (Ed.), Actes de la Conférence ERGO'IA 2008, L'humain au coeur des systèmes et de leur développement (Bidart/Biarritz, 15-17 octobre), ESTIA, 85-92.
- Ltifi, H., Ben Ayed, M., Kolski, C., Alimi, A. 2009. HCI-enriched approach for DSS development: the UP/U approach. *IEEE Symposium on Computers and Communications (ISCC'09, July 5-8)*, IEEE Press, Sousse, Tunisia, 895-900.
- Lévine, P., et Pomerol, J. 1989. Systèmes interactifs d'aide à la décision et systèmes experts. Eds Hermès, Paris.
- Long, J.B., et Denley, I. 1990. Evaluation for practice. Tutorial, Ergonomics society annual conference.
- Matheny, M.E., et Ohno-Machado, L. 2007. Generation of knowledge for clinical decision support: Statistical and machine learning techniques. *Clinical Decision Support*, 10, 227-248.
- McDermid, J., et Ripkin, K. 1984. Life cycle support in the ADA, environment. Cambridge University Press, Cambridge, UK.
- Millot, P. 1990. Coopération homme-machine : exemple de la téléopération. Actes des Journées du GR Automatique. Strasbourg.
- Nielsen, J. 1993. Usability engineering. Academic Press, Boston.
- Pacaux-lemoine M.P. 1998. Coopération Hommes-Machines dans les procédés complexes : modèles techniques et cognitifs pour le contrôle de trafic aérien. Mémoire de Doctorat, Université de Valenciennes et du Hainaut-Cambrésis, Valenciennes.
- Paternò, F. 1999. Model Based Design and Evaluation of Interactive Applications. Springer Verlag, Berlin.
- Pelayo, S., Bernonville, S., Kolski, C., et Beuscart-Zépher, M. 2009. Applying a human factors engineering approach to healthcare IT applications: example of a medication CPOE project. J.G. McDaniel (Ed.), *Advances in Information Technology and Communication in Health*, IOS Press, 334-339.
- Riesbeck, C. et Shank, R. 1989. Inside Case-Based Reasoning. Lawrence Erlbaum.
- Robert, J.M. 2003. Que faut-il savoir sur l'utilisateur pour réaliser des interfaces de qualité?, In G. Boy (Ed.), *Ingénierie cognitive, IHM et cognition*, Edition Hermes-Lavoisier. Paris, 249-283.

- Royce, W. 1970. Managing the development of large software systems: Concepts and techniques. WESCON technical papers.
- Sandoval, V. 1997. L'informatique décisionnelle. Eds Hermes, Paris.
- Seffah, A., Kolski, C., Idoughi D. 2009. Persona comme outil de design de services interactifs : principes et exemple en e-maintenance. Proceedings of IHM 2009, 21ème Conférence de l'Association Francophone sur l'Interaction Homme-Machine (Grenoble, France, 13-16 octobre 2009), International Conference Proceedings Series, ACM Press, Grenoble, 333-336.
- Senach, B. 1990. Evaluation ergonomique des interfaces homme-machine. Une revue de la littérature. Rapport n°1180, INRIA.
- Simon, H.A. 1977. The new science of management decision. Prentice Hall, New Jersey.
- Tricot, N. 2005. Conception et évaluation de systèmes coopératifs avancés d'aide à la conduite automobile. Mémoire de Doctorat, Université de Valenciennes et du Hainaut-Cambrésis, Valenciennes, France.
- Turban, E. 1993. Decision Support and Expert Systems. Eds New York, Macmillan.
- Wilson, J.R., et Corlett E.N. 1996. Evaluation of human works: a practical ergonomics methodology. Eds Taylor & Francis, London.
- Yoshikawa, H. 2003. Modeling Humans in Human-Computer Interaction. In Jacko, J.A., Sears, A. (eds.), The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications, Mahwah, New Jersey, Lawrence Erlbaum Associates, 118-149.



Hela Ltifi est actuellement doctorante en Informatique. Elle est membre de deux laboratoires : REGIM (REsearch Group on Intelligent Machines) de l'École Nationale d'Ingénieurs de Sfax Tunisie et LAMIH (Laboratoire d'Automatique, de Mécanique et d'Informatique industrielles et Humaines) de l'Université de Valenciennes et du Hainaut-Cambrésis ; et ceci dans le cadre d'une thèse en cotutelle. Ses activités de recherche s'articulent autour des systèmes interactifs d'aide à la décision basés sur l'extraction de connaissances à partir des données devant être développés selon des démarches centrées utilisateur ; les résultats fournis par ces systèmes devraient permettre d'aider les médecins dans la lutte contre les infections nosocomiales. Elle est co-auteur de plusieurs articles dans des conférences internationales.



Mounir Ben Ayed est docteur en Science. Il est membre de l'équipe REGIM à l'École Nationale d'Ingénieurs de Sfax ENIS, Université de Sfax, Tunisie. Il est actuellement un enseignant des systèmes de gestion des bases de données, des datawarehouses et du data mining. Ses activités de recherche s'articulent autour des systèmes interactifs d'aide à la décision basés sur l'extraction de connaissances à partir des données. Il a dirigé des travaux sur les entrepôts de données, les OLAP et le data mining comme outils pour l'aide à la décision. La plupart de ses travaux sont appliqués dans le domaine médical. Il est co-auteur de plusieurs articles dans des conférences nationales et internationales. Mounir Ben Ayed a été membre du comité d'organisation du premier congrès international de Signaux Circuits et Système (SCS 2004) ; ainsi que coprésident du comité d'organisation de ACIDCA-ICMI en 2005. Il a organisé en plus des séminaires, écoles et workshops.



Christophe Kolski a obtenu le doctorat en 1989 et l'Habilitation à Diriger des Recherches (HDR) en 1995. Il est spécialisé en Interaction Homme-Machine, plus particulièrement en méthodes et modèles pour les systèmes interactifs et en interaction homme-machine intelligente. Il est professeur en informatique à l'Université de Valenciennes et du Hainaut-Cambrésis (UVHC) où il enseigne le génie logiciel et l'interaction homme-machine. Il dirige le groupe de recherche "Raisonnement Automatique et Interaction Homme-Machine" au LAMIH. Christophe Kolski est auteur, co-auteur, éditeur ou co-éditeur de nombreux livres, chapitres de livres, numéros spéciaux de revues, articles dans des revues et conférences nationales et internationales. Il a été co-éditeur de la revue d'Interaction Homme-Machine. Il a été président du comité d'organisation de la conférence CADUI'2002 et co-président des comités scientifiques des conférences IHM'2003 et ERGO-IA'2006.



Adel M. Alimi est professeur en Informatique industrielle titulaire en Génie Electrique (Informatique industrielle) depuis Décembre 2006. Il est fondateur et premier responsable du Groupe de Recherche sur les Machines Intelligentes (REGIM). Adel M. Alimi est le directeur de l'Ecole Nationale d'Ingénieurs de Sfax depuis Juin 2005, ENIS, Université de Sfax, Tunisie. Il a été le président de la conférence internationale ACIDCA-ICMI'2005 International Conference on Machine Intelligence en 2005, co-président de la conférence internationale SCS'2004 International Conference on Signals, Circuits & Systems, et président de la conférence internationale ACIDCA'2000, International Conference on Artificial & Computational Intelligence for Decision, Control and Automation. Il a organisé plusieurs sessions spéciales et écoles. Il a été aussi membre de comités de programme de plusieurs conférences internationales. Adel M. Alimi est auteur, co-auteur, éditeur ou co-éditeur de chapitres de livres, numéros spéciaux de revues, articles dans des revues et conférences internationales.