



Date : 03/08/2006

Analyse des facettes des catégories utilisées dans les annuaires du Web : étude comparative

Kyung-Sun Kim

Sei-Ching Joanna Sin

Soo-Jin Park

Xiaohua Zhu

Jom Polparsi

School of Library and Information Studies

University of Wisconsin-Madison

Madison, USA

traduit de l'anglais par Françoise Bourdon, Bibliothèque nationale de France

Meeting:

136 Classification and Indexing

Simultaneous Interpretation:

No

WORLD LIBRARY AND INFORMATION CONGRESS: 72ND IFLA GENERAL CONFERENCE AND COUNCIL
20-24 August 2006, Seoul, Korea
<http://www.ifla.org/IV/ifla72/index.htm>

Résumé

La classification à facettes est considérée comme pertinente pour organiser les ressources numériques. Basé sur le modèle de classification à facettes suggéré pour les ressources du Web (Zins, 2002), cette étude analyse les annuaires du Web les plus répandus dans différents pays/régions d'Asie et examine les différences culturelles que reflètent leurs systèmes de classification. Trois annuaires du Web très utilisés dans quatre pays/régions (Chine, Hong Kong, Corée et Thaïlande) ont été sélectionnés et leurs systèmes de classification ont été analysés et comparés : un Yahoo local et deux annuaires élaborés sur place pour chaque région. Partant des résultats, l'étude propose un modèle plus adapté à la culture asiatique.

INTRODUCTION

L'information disponible sur le Web est de plus en plus diverse et son volume croît rapidement. En janvier 2005, on a estimé que le Web indexable atteignait au moins 11,5 billions de pages (Gulli & Signorini, 2005). Aujourd'hui, ces ressources sur le Web, volumineuses et sans cesse changeantes, sont le plus souvent organisées selon des systèmes de classification hiérarchiques. Cependant, de tels systèmes hiérarchiques peuvent ne pas être pertinents pour organiser des ressources interculturelles destinées à des groupes de culture et de langue différentes, puisqu'il peut y avoir des différences culturelles dans la catégorisation (Segall, 1990 ; Malt, 1995).

La présente étude examine et analyse les classes ou catégories utilisées dans douze annuaires du Web pour quatre pays asiatiques différents. S'appuyant sur les résultats, l'étude propose et commente un modèle de classification à facettes qui pourrait aider à organiser les ressources asiatiques du Web. L'objectif de l'étude est de prendre en compte aussi bien les caractéristiques uniques que communes trouvées dans les classifications du Web utilisées dans chacun des pays ou régions, et de proposer un modèle de classification à facettes qui pourrait faciliter l'accès aux ressources du Web à des utilisateurs asiatiques ayant des cultures et des langues différentes.

CLASSIFICATIONS A FACETTES vs CLASSIFICATIONS HIERARCHIQUES

Pour faciliter l'apprentissage et la recherche d'information, nous classons les choses en rassemblant celles qui sont similaires. La plupart des systèmes de classification que nous utilisons traditionnellement sont des systèmes hiérarchiques dans lesquels les sujets sont organisés selon une hiérarchie de classes, de divisions et de subdivisions. Dans de tels systèmes un sujet est classé à un seul endroit et ne peut être atteint que par un seul chemin. Sauf s'ils comprennent bien la structure de la hiérarchie, et savent exactement dans quelles classes principales (dans quelles divisions, subdivisions, etc.) se trouve un sujet particulier, cela exige des utilisateurs une série d'essais et d'erreurs pour localiser un sujet dans le système.

Les similitudes, ressemblances et différences invoquées par Aristote et que gère la plupart des systèmes de classification hiérarchique ne sont pas toujours des relations universellement acceptées (Olson, 2001). Elles sont influencées de façon significative par notre société et peuvent varier d'un groupe à l'autre en fonction de sa culture et de ses besoins. En outre, les gens peuvent ne pas être d'accord sur les ressemblances à considérer comme importantes et sur ce qui devrait être utilisé pour catégoriser les choses au niveau supérieur de la hiérarchie. Un problème majeur que posent les systèmes de classification hiérarchique est que les gens ne veulent pas toujours privilégier la même similitude : "nous ne voulons pas toujours ce que l'ordre d'énumération nous donne" (Olson, 2001). Il est nécessaire de tenir compte de la diversité des utilisateurs et d'appliquer une gamme de solutions optionnelles, partielles et locales.

A la différence des systèmes hiérarchiques qui attribuent aux sujets des séquences fixes, les systèmes de classification à facettes attribuent aux sujets les aspects, les propriétés ou les caractéristiques d'une classe ou d'un sujet spécifique, définis clairement, exclusifs les uns des autres et exhaustifs dans leur ensemble (Taylor, 2000). La classification à facettes offre plusieurs chemins pour localiser un sujet. En utilisant différents aspects/facettes d'un sujet, les utilisateurs peuvent restreindre la recherche et localiser le sujet plus facilement. Bien que chaque système hiérarchique et à facettes ait ses points forts et ses faiblesses (Kwasnik, 1999), la structure hiérarchique traditionnelle est de plus en plus considérée comme inefficace pour organiser les ressources du Web. Aujourd'hui l'approche par facettes est de plus en plus considérée comme étant la plus efficace et la plus appropriée pour organiser les ressources du Web (Ellis & Vasconcelos, 2000 ; Bates, 2002 ; Broughton, 2002).

Comparée à la classification hiérarchique, la classification à facettes est souple et ouverte à de nouvelles catégories. L'approche par facettes est aussi plus appropriée pour représenter des idées ou des objets complexes. Elle est plus particulièrement utile pour organiser des pièces ou des documents complexes dans un environnement multidisciplinaire (Broughton, 2002). Dans leur examen de l'utilisation de l'approche par facettes dans les bases de données en ligne, Inwersen et Wormell (1992) ont découvert que l'approche par facettes excelle dans la fourniture de points d'entrée structurés et multidimensionnels aux documents et aux objets. De tels points d'entrée multiples sont bons pour l'organisation des ressources du Web, car les utilisateurs du Web sont dans des contextes différents. Tout se passe comme si différents utilisateurs essayaient de localiser un sujet ou de naviguer sur un sujet à partir de différents points d'entrée. Les systèmes hiérarchiques traditionnels, tels que la CDD [Classification décimale Dewey] ou la LCC [Classification de la Library of Congress] ont été critiqués comme étant trop rigides et parfois ethnocentriques (Olson, 1998). En étudiant les termes parents dans différents schémas de classification, y compris dans la CDD et la LCC, Kwasnik et Rubin (Kwasnik & Rubin, 2003) ont démontré la difficulté à mettre en correspondance et à traduire les termes dans ces schémas de classification pour représenter de façon précise les concepts parents dans des cultures et des langues différentes. Une approche par

facettes, souple et ouverte à de multiples perspectives, devrait être plus appropriée pour effectuer le classement des ressources interculturelles et multilingues. (Kwasnik, 1999).

Les systèmes de classification à facettes ont aussi quelques faiblesses. Une notation un peu longue ou complexe avec laquelle les utilisateurs ne sont pas familiarisés en est une (Iyer, 1995). Pour les ressources du Web, cependant, cela n'est pas un problème puisque les documents ou les pièces sont accessibles par les liens hypertextes sans qu'il soit besoin d'utiliser la notation. Dernièrement plusieurs applications ont été développées, explorant comment implémenter efficacement l'indexation, la navigation et la recherche par facettes (Devadason, Intaraksa, Patamawongjariya, & Desai, 2001 ; Yee, Swearingen, Li, & Hearst, 2003 ; Ross & Janevski, 2005).

Cette étude vise à démontrer la capacité d'une approche par facettes à organiser les ressources du Web dans différentes parties de l'Asie, et à proposer un modèle de classification à facettes qui pourrait convenir à la culture asiatique. Nous décrirons d'abord la méthodologie utilisée dans cette étude. Puis, nous présenterons les résultats. Un modèle de classification à facettes, basé sur les résultats, sera présenté ensuite. Cette présentation se terminera par une discussion portant sur les résultats et leurs conséquences.

L'ETUDE

Le but de cette étude est d'élaborer un modèle préliminaire de classification à facettes pour organiser les ressources asiatiques du Web, à partir de l'analyse des catégories actuellement utilisées dans les annuaires les plus répandues dans quatre pays d'Asie. Les objectifs de l'étude comprennent : 1) tester et peaufiner le modèle à facettes de Zen pour organiser les ressources du Web ; et 2) étendre ce modèle en y ajoutant des sous-facettes qui pourraient s'avérer utiles.

L'échantillon

Un total de 12 annuaires du Web, largement utilisés dans quatre pays/régions d'Asie (y compris la Chine, la Région administrative de Hong-Kong, la Corée, et la Thaïlande) ont été sélectionnés et leurs schémas de classification ont été analysés dans cette étude. Les annuaires du Web sélectionnés pour cette étude sont : Sina Beijing et Sohu (Chine) ; Sina HK et Tim Way (Hong Kong) ; Empas et Naver (Corée) ; et Hunsa et Sanook (Thaïlande). Pour chaque région, le Yahoo local a été aussi sélectionné. De plus, les catégories de haut niveau utilisées dans Yahoo, l'Annuaire Google, et le Librarians' Internet Index des États-Unis ont été aussi analysées et comparées avec les autres pour identifier les similarités et les différences entre Asie et Amérique.

La collecte des données

Pour chaque annuaire du Web, toutes les pages affichant les classes ou les catégories de premier et de deuxième niveau ont été capturées le 10 juillet 2005. Les classes ou catégories utilisées dans les annuaires de chaque pays ont été traduites et analysées.

RESULTATS

Catégories de premier niveau dans les annuaires Yahoo locaux :

Quand on a comparé les Yahoo locaux, on a trouvé quelques différences mineures (voir Tableau 1). Le Yahoo de Thaïlande, par exemple, a une catégorie " Voyages" qui n'existe pas dans les autres. En Corée, la façon dont les sujets sont regroupés diffère légèrement. Les Lettres qui sont habituellement présentées avec les arts, sont, en Corée, regroupées avec les sciences sociales. En Chine et à Hong-Kong, la Politique a été ajoutée à la catégorie "Gouvernement". Mises à part ces quelques différences, les classes de premier niveau sont presque identiques. Une telle similarité entre les différentes versions locales de Yahoo n'est pas surprenante puisque le Yahoo des États-Unis peut exercer son contrôle sur les versions locales.

Tableau 1. Les catégories de premier niveau dans les Yahoo locaux.

Yahoo Etats-Unis	Yahoo Chine	Yahoo Hong Kong	Yahoo Corée	Yahoo Thaïlande
Arts & Lettres <i>(Arts & Humanities)</i>	oui	oui	Arts	oui
Affaires & Économie <i>(Business & Economy)</i>	oui	oui	oui	oui
Informatique & Internet <i>(Computers & Internet)</i>	oui	oui	oui	oui
Formation <i>(Education)</i>	oui	oui	Formation/Apprentissage <i>(Education / Learning)</i>	oui
Divertissements <i>(Entertainment)</i>	oui	oui	oui	oui
Gouvernement <i>(Government)</i>	Gouvernement & Politique <i>(Government & Politics)</i>	Gouvernement & Politique <i>(Government & Politics)</i>	oui	oui
Santé <i>(Health)</i>	Santé et Médecine <i>(Health & Medicine)</i>	Santé et Médecine <i>(Health & Medicine)</i>	Santé et Médecine <i>(Health & Medicine)</i>	oui
Informations & Médias <i>(News & Media)</i>	oui	oui	oui	oui
Loisirs & Sports <i>(Recreation & Sports)</i>	Loisirs et Vie <i>(Recreation & Life)</i>	Loisirs et Vie <i>(Recreation & Life)</i>	Vie de loisirs <i>(Leisure Life)</i>	Loisirs /Sports <i>(Recreation/ Sports)</i>
Références <i>(Reference)</i>	oui	oui	oui	oui
Régions <i>(Regional)</i>	oui	oui	oui	Provinces & Districts <i>(Provinces & Districts)</i>
Sciences <i>(Science)</i>	oui	oui	oui	oui
Sciences sociales <i>(Social Science)</i>	oui	oui	Lettres et Sciences sociales <i>(Humanities & Social Science)</i>	oui
Société & Culture <i>(Society & Culture)</i>	oui	oui	oui	oui
Nouvelles additions <i>(New additions)</i>	N/A	N/A	N/A	N/A
				Voyage <i>(Travel)</i>

Catégories de premier niveau dans les Annuaire Yahoo locaux basés sur le modèle de Zins

En analysant les principales catégories de premier niveau utilisées dans neuf portails Internet incluant des portails commerciaux (tels que AOL, Yahoo), Zins (Zins, 2002) a identifié huit facettes communément utilisées, reposant sur ce qui peut servir à organiser les ressources du Web : 1) sujets 2) objets 3) applications 4) utilisateurs 5) localisations 6) références 7) médias, et 8) langues

Ici la facette "sujets" caractérise un sujet ou une grande discipline (par exemple : les affaires, la formation, les sciences). La facette "objets" fait référence à tous les types d'objet ou d'entité du monde réel, y compris les organisations et les personnes. La facette "applications" couvre toutes les ressources associées à des applications spécifiques (par exemple, les clavardages [chats], les messageries électroniques). La facette "utilisateurs" indique quels sont les utilisateurs potentiels. La facette "localisations" correspond à toutes les ressources relatives à des localisations géographiques. La facette "références" renvoie à différentes ressources de référence comme des almanachs, des dictionnaires, des encyclopédies, etc. La facette "médias" indique tous les médias qui servent à exprimer le contenu d'une ressource (par exemple des graphiques, la radio). Enfin, la facette "langues" signale la langue utilisée dans les sites Web.

Lors de l'utilisation du modèle de Zins pour analyser les classes de premier niveau, on a constaté que la plupart des classes appartenait à la facette "sujets" (Voir Tableau 2). Les facettes "références", "médias" et "applications" étaient aussi représentées. Comme tous les annuaires du Web qui sont étudiés ici sont destinés à des pays donnés et à leurs propres citoyens, la facette "langues" n'existe pas. Les facettes "utilisateurs" et "objets" n'ont pas été trouvées non plus.

Tableau 2. Principales catégories dans les Yahoo locaux : analyse des facettes basée sur le modèle de Zins

facettes	sujets	localisations	références	médias	applications
Catégories	Arts & Lettres Affaires & Économie Informatique & Internet Formation Divertissements Gouvernement Santé Informations & Médias Loisirs & Sports Sciences Sciences sociales Société et Culture	Régions	Références	Informations & Médias	Voyage (Thaïlande)

Catégories de premier niveau dans les annuaires du Web élaborés localement

Ensuite, les classes de premier niveau ont été comparées dans tous les annuaires du Web élaborés dans quatre pays d'Asie. Les Yahoo locaux n'ont pas été intégrés dans cette analyse puisque leurs catégories ont déjà été étudiées et jugées similaires d'un pays à l'autre. Toutes les classes principales de deux annuaires locaux pour chaque pays ont été analysées et comparées : *Sina Beijing* et *Sohu* (Chine) ; *Sina HK* et *Tim Way* (Hong-Kong) ; *Empas* et *Naver* (Corée) ; et *Hunsa* et *Sanoook* (Thaïlande). En tant que références, les classes de premier niveau de Yahoo, de l'Annuaire Google et du Librarians'Internet Index des Etats-Unis ont été incluses dans cette étude. Au total les principales catégories de premier niveau de 11 annuaires du Web basés sur le modèle de Zins ont été analysées. Les résultats sont présentés dans le Tableau 3.

Tableau 3. Principales catégories dans les annuaires du Web élaborés localement

	États-Unis	Chine	Hong-Kong	Corée	Thaïlande
	<i>Yahoo</i> [Y] <i>Google Directory</i> [G] <i>Librarians' Internet Index</i> [L]	<i>Sina Beijing</i> [N] <i>Sohu</i> [H]	<i>Sina HK</i> [N] <i>Tim Way</i> [T]	<i>Empas</i> [E] <i>Naver</i> [N]	<i>Hunsa</i> [H] <i>Sanook</i> [S]
sujets	Arts & Lettres (Y) (L) Arts (G)	Arts (H) (N) Littérature (<i>Literature</i>) (H) (N)	Arts (T) Le Monde des Lettres (<i>Arts World</i>) (N)	Culture, Arts & Religion (E) Culture & Arts (N)	Société, Arts & Culture (H)
sujets	Affaires & Économie (Y) Affaires (G) (L)	Affaires & Économie (H) (N)	Finance/Propriété (<i>Finance/Property</i>) (T) Affaires, achats & ventes (<i>Business, Buy & Sell</i>) (T) Économie (N)	Affaires & Centres commerciaux (<i>Business & Shopping Mall</i>) (E) Affaires et Shopping (<i>Business & Shopping</i>) (N) Économie & Investissements lucratifs (<i>Economy & Money-Making Know-How</i>) (E) (N)	Affaires (S) Affaires & Économie (H)
sujets	Informatique & Internet (Y) Informatique (G) (L)	Informatique & Internet (H) (N)	Informatique (T) Informatique & Internet (N)	Informatique & Internet (E) (N)	Informatique (S) Internet (S) Informatique & Internet (H)
sujets	Formation (Y) (L)	Formation & Emploi (<i>Education & Employment</i>) (N) Formation & Apprentissage (<i>Education & Training</i>) (H)	Formation (T) Formation & Emploi (<i>Education & Employment</i>) (N)	Formation & Établissements scolaires (<i>Education & Schools</i>) (E) Formation & Sciences (<i>Education & Sciences</i>) (N)	Formation (S) Sciences & Formation (<i>Science & Education</i>) (H)
sujets	Divertissements (Y)	Divertissements & Loisirs (<i>Entertainment & Recreation</i>) (H) (N)	Divertissements (T)	Divertissements (E) (N)	Divertissements & Loisirs (<i>Entertainment & Recreation</i>) (S) Divertissements (H)
sujets	Jeux (G)		Jeux & Passe-temps (<i>Games & Hobbies</i>) (N)	Jeux (E) (N)	
sujets	Gouvernement (Y) (L)	Politique/Droit/Forces armées (<i>Politics/Law/Military</i>) (H) (N)	Gouvernement & Politique (<i>Government & Politics</i>) (T) (N)	Agences gouvernementales & Société (<i>Government Agencies & Society</i>) (E) Société & Politique (<i>Society & Politics</i>) (N)	Gouvernement & Organisations (<i>Government & Organization</i>) (S) (H)
sujets	Santé (Y) (G) (L)	Médecine & Santé (<i>Medicine & Health</i>) (N) Santé (H)	Santé (T) Services & Traitements médicaux (<i>Health Care & Medical Treatment</i>) (N)	Santé & Médecine (<i>Health & Medicine</i>) (E)	Sciences médicales & Santé (<i>Medical Science & Health</i>) (S) Santé & Sciences médicales (<i>Health & Medical Science</i>) (H)

sujets	Maison (<i>Home</i>) (G) (L)				
sujets	Droit (<i>Law</i>) (L)	Politique/Droit/Forces armées (<i>Politics/Law/Military</i>) (H) (N)			
sujets	Informations & Médias (Y) Informations (<i>News</i>) (G)	Informations & Médias (H) (N)	Médias de masse (<i>Mass Media</i>) (T) Informations et Médias de masse (<i>News & Mass Media</i>) (N)	Informations & Médias (E) (N)	Informations & Médias (S) Médias et Informations (H)
sujets	Loisirs & Sports (Y) Loisirs (<i>Recreation</i>) (G) (L)		Loisirs/Détente (<i>Recreation/Leisure</i>) (T)	Détente & Sports (<i>Leisure & Sports</i>) (N)	Divertissements & Loisirs (<i>Entertainment & Recreation</i>) (S) Loisirs (<i>Recreation</i>) (H)
sujets	Sciences (Y) (G) (L)	Sciences & Techniques (<i>Science & Technology</i>) (H) (N)	Sciences (T) Sciences & Techniques (<i>Science & Technology</i>) (N)	Sciences (E) Formation & Sciences (<i>Education & Sciences</i>) (N)	Sciences et Formation (<i>Science & Education</i>) (H)
sujets	Sciences sociales (Y)	Sciences sociales (H) (N)			
sujets	Société & Culture (Y) Société (G)	Société & Culture (H) (N)	Société (T) Société & Culture (N)	Culture, Arts & Religion (<i>Culture, Arts & Religion</i>) (E) Société & Politique (<i>Society & Politics</i>) (N)	Les Gens, la Société & la Culture (<i>People, Society & Culture</i>) (S) Société, Arts & Culture (<i>Society, Arts & Culture</i>) (H)
sujets	Sports (G)	Sports & Entraînement physique (<i>Sports & Physical Training</i>) (H) (N)	Sports (T) (N)	Voyages & Sports (<i>Travel & Sports</i>) (E) Détente & Sports (<i>Leisure & Sports</i>) (N)	Sports (S) (H) Voyages & Nature (<i>Travel & Nature</i>) (H)
sujets		Demande d'emploi & Recrutement (<i>Job Application & Recruitment</i>) (N)			
sujets		Services sociaux (<i>Life Service</i>) (H) (N)	La Vie/Le Vivant (<i>Life/Living</i>) (T) Vie pratique (<i>Life Info</i>) (N)	Modes de vivre & Passe-temps (<i>Lives & Hobbies</i>) (E) Styles de vie (<i>Life Styles</i>) (N)	
sujets			Alimentation (<i>Eating</i>) (T)		
sujets					Activités & Événements (<i>Activities & Events</i>) (S)
sujets					Savoir/Connaissance & Informations (<i>Knowledge/Wisdom & Info</i>) (S)
sujets					Automobile (<i>Automobile</i>) (S) (H)
sujets					Immobilier (<i>Real Estate</i>) (S)
applications			Voyages (N)	Voyages & Sports (E) Le Monde & le Voyage (<i>World &</i>	Voyages (S) Voyage & Nature (<i>Travel &</i>

				<i>Travel</i> (N)	<i>Nature</i> (H)
applications	Shopping (G)		Affaires, achats & ventes (<i>Business, Buy & Sell</i>) (T)	Affaires et Shopping (<i>Business & Shopping</i>) (N)	Shopping (S) Shopping & Services (<i>Shopping & Service</i>) (H)
langues	Régions (Y) (G) (L)	Comtés & Régions (<i>Counties & Regions</i>) (H)			
langues	Le Monde (<i>World</i>) (G)			La Corée & le Monde (<i>Korea & World</i>) (E) Le Monde & le Voyage (<i>World & Travel</i>) (N)	
médias		Pages d'accueil personnelles (<i>Homepage Personal</i>) (N)	Pages d'accueil personnelles (<i>Homepage Personal</i>) (T)		
médias	Informations & Médias (Y) Médias (L)	Informations & Médias (H) (N)	Médias de masse (<i>Mass Media</i>) (T) Informations et Médias de masse (<i>News & Mass Media</i>) (N) Films, TV & Musique (<i>Film, TV, Music</i>) (N)	Informations & Médias (E) (N)	Informations & Médias (S) Médias & Informations (H)
objets		Entreprises (<i>Companies</i>) (H)	Sociétés de services (<i>Service organizations</i>) (T)		Banques & Institutions financières (<i>Banks & Financial Inst.</i>) (S) Gouvernements & Organisations (<i>Government & Organization</i>) (S) (H)
objets	Les Gens (<i>People</i>) (L)			Les Gens (<i>People</i>) (N)	
objets			Vedettes (<i>Stars</i>) (T)		
références	Références (Y) (G) Références & Données immédiates (<i>Ready Ref. & Quick Facts</i>) (L)	Références (N)		Dictionnaires & Références (<i>Dictionaries & Reference</i>) (E)	
utilisateurs			Adultes/Sexe (<i>Adult/Sex</i>) (N)		
utilisateurs	Enfants & Adolescents (<i>Kids & Teens</i>) (G)	Enfants (<i>Kids</i>) (N)		Naver pour les Jeunes (<i>Junior Naver</i>) (N)	
	Nouvelles Additions (Y)				Autres (<i>Others</i>) (H)

Etats-Unis : *Yahoo* [Y], *Google Directory* [G] et *Librarians' Index* [L] ; Chine : *Sina Beijing* [N] et *Sohu* [H] ; Hong-Kong : *Sina HK* [N] et *Tim Way* [T] ; Corée : *Empas* [E] et *Naver* [N] ; Thaïlande : *Hunsa* [H] et *Sanook* [S]

Les résultats montrent que la plupart des classes de premier niveau (69%) possède une facette "sujet". Toutes les autres facettes, mise à part la facette "langue" sont présentes. Il est intéressant de constater que certaines différences existent entre les cultures américaines et asiatiques dans la façon de classer par catégories. Les "Lettres" par exemple sont souvent regroupées avec les "Arts" dans les annuaires des États-Unis, alors que les annuaires asiatiques ont tendance à séparer les "Lettres" des "Arts", ou bien à classer les "Lettres" sous "Culture" ou sous "Sciences". A la différence des annuaires américains qui ont une catégorie "Gouvernement", la plupart des annuaires asiatiques ont tendance à avoir une catégorie combinée "Gouvernement et Politique". Les catégories "Voyages" et "La Vie/Le Vivant" sont très répandues dans les annuaires asiatiques autres que les annuaires chinois.

Certains pays asiatiques semblent partager une catégorisation similaire à celle des Américains. Dans les annuaires américains et chinois, les catégories "Sciences sociales" et "Droit" sont toutes les deux présentes. La catégorie "Jeux" est importante dans les annuaires des États-Unis, de Hong-Kong et tout particulièrement dans ceux de Corée. Les annuaires américains et coréens ont aussi une catégorie "Le Monde".

Certaines catégories semblent n'être répandues que dans certains pays. "Automobile" par exemple, est une des catégories les plus répandues en Thaïlande, alors qu'elle n'est pas courante dans d'autres pays. La catégorie "Maison" ne se trouve que dans les annuaires américains. Il semble que dans les annuaires asiatiques les ressources correspondant à "Maison" sont classées sous "Vie" qui est une catégorie importante dans ces annuaires.

Les catégories de deuxième niveau dans les annuaires du Web asiatiques

Dans le but d'identifier les sous-facettes les plus courantes qui pourraient être utilisées pour organiser les ressources du Web asiatiques, les catégories de deuxième niveau utilisées dans tous les annuaires des quatre pays asiatiques ont été analysées. Pour chaque pays, trois annuaires du Web ont été étudiés, incluant un Yahoo local et deux annuaires élaborés localement. Un total de douze annuaires a été pris en compte dans l'étude.

Après l'étude préliminaire basée sur le modèle de Zins, nous avons découvert quelques points qui devaient être résolus avant de poursuivre les analyses. L'un d'entre eux était relatif à la facette "utilisateurs". Nous avons constaté qu'il était souvent difficile de distinguer entre les personnes considérées comme "sujets" et les personnes considérées comme "utilisateurs". Par exemple la catégorie "Enfants & Adolescents" contient non seulement les sites Web qui sont utilisés par les enfants, mais aussi ceux qui traitent des enfants, qui sont créés par des enfants, et qui sont utilisés par des enseignants ou des parents s'adressant aux enfants. Il est évident que la facette "personnes" ["people"] serait d'une utilisation plus facile. Nous avons décidé de supprimer la facette "utilisateurs", et de créer à la place une facette "personnes" pour couvrir les ressources relatives aux personnes ayant une fonction d'utilisateur, de créateur, de sujet, etc. Parce que la facette "objets" est supposée couvrir les organisations et les personnes (qui sont déjà couvertes à présent par la nouvelle facette "personnes"), la facette "objets" a été éliminée. Pour couvrir les organisations qui se trouvaient à l'origine sous la facette "objets", une facette "organisations" a été créée.

Nous avons trouvé que la facette "références" n'était pas claire tout particulièrement dans un environnement Web. Par exemple, les sites Web fournissant des données factuelles sur les pays – comme le font les répertoires géographiques ou les encyclopédies – sont souvent classés sous des vedettes spécifiques telles que "profil du pays" ou "information par pays", sans précision du type de référence (par exemple, répertoire géographique, encyclopédie). Ce type de vedettes est tout à fait courant et il nous a semblé nécessaire de traiter cette question. La facette "information générale" a été créée pour couvrir de telles ressources. La facette "médias" est aussi problématique. Zins (2002) indique que les graphiques, les images et la radio sont des exemples de ce qui est couvert par la facette "médias". Il nous semble que cette facette "médias" couvre le type d'information aussi bien que le type de média. Comme une grande variété de types, de supports, de médias, et d'outils concerne la fourniture de l'information, et que c'est un défi pour les utilisateurs que de faire la distinction entre eux, nous avons décidé de créer une facette assez générique pour couvrir le tout. La facette "médias/types/supports/outils" remplace la facette "médias" pour couvrir les différents médias de diffusion (par exemple : radio, TV), les supports (par exemple : son, vidéo), les types (par exemple : citations, encyclopédies) et les outils (par exemple : calculateur, moteur de recherche).

Le modèle Zins a été modifié pour prendre en compte les huit facettes suivantes : 1) sujets, 2) applications, 3) localisations 4) médias/types/supports/outils 5) organisations 6) personnes 7) information générale, et 8) langues. Bien que les annuaires du Web que nous avons étudiés n'aient pas de facette "langues", nous avons considéré qu'il était important de conserver une facette "langues" principalement pour les annuaires s'adressant à un groupe

hétérogène d'utilisateurs. Si les annuaires du Web étaient élaborés pour un pays constitué de beaucoup d'immigrants (comme les États-Unis) ou pour des utilisateurs internationaux, la facette "langues" serait indubitablement l'une des plus importantes.

Tableau 4. Classification à facettes pour les ressources du Web asiatiques : modèle de Zins et modèle modifié

Modèle de Zins	Modèle modifié
sujets	sujets
applications	applications
localisations	localisations
médias	médias/types/supports/outils
objets	organisations
	personnes
utilisateurs	
références	information générale
langues	langues

En utilisant l'ensemble des facettes ainsi modifié, les catégories de deuxième niveau utilisées dans douze annuaires du Web ont été étudiées. A partir de l'analyse des catégories utilisées, nous avons essayé d'identifier les sous-facettes qui seraient utiles pour organiser les ressources sous chacune des facettes principales (voir Tableau 5).

Tableau 5. Sous-facettes identifiées à partir des catégories de deuxième niveau

Facettes	Exemple de catégories sous les facettes	Groupes de sous-facettes
sujets	Automobile Basket Caractères coréens et ordinateurs Collections [<i>Collectibles</i>] Conditionnement [<i>Packaging</i>] Droit des affaires Droits de l'homme Droits des consommateurs Émotion, mariage et amour Enseignement & apprentissage Environnement & nature Études culturelles Examen d'entrée dans l'enseignement supérieur Histoire de la littérature Impression sur tissu et peinture Informatique Investissements lucratifs Jouets Linguistique et langages humains Littérature contemporaine Musique populaire Musique traditionnelle thaï Mythes et fables Nanotechnologie Pétrole / Gaz naturel Régimes alimentaires Relations internationales Santé des animaux domestiques Santé et médecine Soins capillaires Théorie critique	<ul style="list-style-type: none"> • Chose concrète, Entité (ex. : automobile ; jouets) • Action & Procédé (ex. : Impression sur tissu et peinture ; conditionnement) • Type d'agent (ex. : Environnement & Nature ; Droits de l'homme ; Santé des animaux domestiques ; Pétrole / Gaz naturel) • Temps [Récence] [<i>Recency</i>] (ex. : Musique traditionnelle thaï ; Nanotechnologie) • Finalité [Fonction] – pour les affaires, les informations quotidiennes ; la formation ; les divertissements (ex. : Droit des affaires ; Soins capillaires ; Examen d'entrée dans l'enseignement supérieur ; Musique populaire) • Universitaire / Populaire
applications	Approvisionnement professionnels et services	<ul style="list-style-type: none"> • Fonction – pour affaires ; vie quotidienne

	<p>Blogues Cartes virtuelles [E-cards] Clavardages [Chats] Forums/Babillage électronique [BSS] Jeux en ligne Logos pour téléphone mobile et téléchargement de sonneries Partage de fichiers Services au consommateur @ Services financiers Shopping Téléchargements [Downloads]</p>	<p>(ex. : Services financiers ; Cartes virtuelles ; Clavardages)</p> <ul style="list-style-type: none"> • En ligne / En différé • Payant / Gratuit • Transmission à sens unique / multiple (ex. : Téléchargements ; Clavardages ; Blogues) • Produits / Services
localisations	<p>Asie Bangkok Comté/Ville/Division en Chine Corée Océans Par pays Par région Provinces et districts Régions et pays Vie rurale Vie urbaine Zones industrielles et économiques</p>	<ul style="list-style-type: none"> • Découpage physique ou géographique (ex. : Océans ; Corée) • Découpage politique (ex. : Par pays ; Provinces et districts) • Découpage économique (ex. : Zones industrielles et économiques) • Degré de ruralité (ex. : Vie rurale ; Vie urbaine) • Echelle (ex. : Pays ; Région)
médias/types/supports/outils	<p>Annuaire Cartes Concours, enquêtes et sondages Evaluation et classement Galeries de photos Horaires des transports Journaux Magazines MP3 @ Petites annonces Prévisions météorologiques Revue Sites Web personnels</p>	<ul style="list-style-type: none"> • Forme de publication (ex. : livres, revues, cartes) • Type de document de référence (ex. : répertoires, journaux) • Forme de l'information (ex. : texte ; image ; son) • Officiel/Non officiel [Formel/Informel] • Fonctions (ex. : Calculateur ; moteurs de recherche ; sondages) • Médias (ex. : radio, TV, Web)
organisations	<p>Agences gouvernementales Ambassades et consulats Banques de commerce Cercles d'admirateurs [Fan clubs] Écoles de commerce Entreprises étrangères Instituts de recherche Main d'œuvre/Organisation du travail Organisations étudiantes Organisations internationales</p>	<ul style="list-style-type: none"> • Officiel (ex. : agences gouvernementales) / Non officiel (ex. : Cercles d'admirateurs) • Fonction • Avec/Sans espace physique (ex. : Écoles de commerce ; Cercles d'admirateurs)
personnes	<p>Artistes Avocats Consommateurs Cultures et groupes Educateurs Enfants Familles Femmes Femmes, femmes au foyer Gays et lesbiennes Joueurs d'échecs asiatiques</p>	<ul style="list-style-type: none"> • Age (ex.: Enfants) • Culture/Ethnie (ex.: Musiciens coréens) • Niveau de formation (ex. : formation supérieure) • Niveau d'expérience (ex. : experts) • Tendance sexuelle (ex. : hommes ; gays et lesbiennes)

	Leaders nationaux et de parti Musiciens coréens People Sa Majesté le Roi et la Famille royale Savants	<ul style="list-style-type: none"> • Profession (ex. : artistes ; avocats) • Statut (ex. : famille royale ; chef de direction) • Groupes (ex. : clubs) • Rôle (ex. : parents ; femmes au foyer)
information générale	Information Information pour les petites et moyennes entreprises Information sur le mariage Ressources	<ul style="list-style-type: none"> • Champ d'application (ex. : général vs spécifique)
langues		<ul style="list-style-type: none"> • Découpage géographique

La liste des sous-facettes a été développée en partant des quatre catégories de deuxième niveau les plus répandues dans les systèmes de classification utilisés dans les annuaires du Web. Elle n'est en rien exhaustive, mais elle est pertinente pour les catégories utilisées dans les annuaires sélectionnés.

DISCUSSION & FUTURS TRAVAUX

Les résultats de l'analyse des catégories de premier niveau laissent entendre que les annuaires du Web élaborés dans les pays asiatiques présentent des similitudes entre eux, et aussi avec ceux élaborés aux États-Unis. La plupart des catégories de premier niveau utilisées dans les annuaires américains sont présentes dans les annuaires du Web asiatiques. En même temps, les annuaires locaux incorporent quelques catégories uniques qui peuvent s'avérer utiles pour comprendre la société, la culture et les valeurs réelles. Cependant, il faut noter que la plupart des annuaires du Web qui ont été étudiés sont des annuaires commerciaux, et que certaines catégories peuvent avoir été ajoutées intentionnellement pour des finalités commerciales. C'est pourquoi il n'y aura pas de remarques relatives à la culture de chaque pays, bien qu'un modèle unique ressorte des données collectées. La discussion portera à la place sur un modèle à facettes qui pourrait être utilisé pour organiser les ressources du Web en Asie particulièrement.

Le modèle de Zins utilisé pour l'étude préliminaire a été utile parce qu'il a fourni un cadre relativement simple pour commencer. Cependant il a dû être modifié pour prendre en compte des caractéristiques uniques de ressources disponibles sur le Web et pour rendre le modèle plus convivial. Bien qu'une classification à facettes de pure forme nécessite des facettes exhaustives et s'excluant les unes des autres, il a été décidé de ne pas viser un système à facettes aussi pur car il risquait d'être irréaliste. Ainsi, un système à facettes pur pourrait contenir des facettes que les utilisateurs ne pourraient pas vraiment comprendre et utiliser. Comme notre objectif était d'élaborer un modèle qui facilite l'organisation et la recherche des ressources du Web, nous avons essayé d'élaborer un modèle clair et pratique à partir d'une approche par facettes.

Notre modèle de classification à facettes comprend huit facettes et plus de trente sous-facettes qui sont en cours de développement :

1. sujets (chose concrète, entité ; action et procédé ; type d'agent ; récence [recency] ; finalité et fonction ; universitaire / populaire)
2. applications (fonction ; en ligne/en différé ; payant/gratuit ; transmission à sens unique/multiple ; produits/services)
3. localisations (découpage physique ou géographique ; découpage politique ; découpage économique ; degré de ruralité ; échelle)
4. médias/types/supports/outils (forme de publication, type de document de référence, forme de l'information, [formel/informel]; fonctions ; médias)
5. organisations (officiel/non officiel ; fonction ; avec ou sans espace physique)
6. personnes (âge, culture/ethnie ; niveau de formation ; niveau d'expérience ; tendance sexuelle ; profession ; statut ; groupes ; rôle)
7. information générale (champ d'application)
8. langues (découpage géographique)

Pour certaines facettes il est relativement facile d'identifier des sous-facettes communes. Mais pour d'autres facettes, comme la facette "sujets" c'est plus difficile. La facette "sujets" couvre une grande variété de thèmes

différents dans des contextes et des disciplines variés. Bien que certaines facettes que nous pensions pouvoir être utiles ont été identifiées et incluses dans le modèle, il est évident qu'une recherche plus approfondie est nécessaire pour élaborer un ensemble plus complet de sous-facettes bien définies et utiles. On pourrait s'inspirer de la classification codée de l'information [Information Coding Classification] de Dalberg (Dalberg, 1994), système d'ordonnement universel par disciplines et domaines, pour identifier des sous-facettes pour regrouper différents domaines. Il convient de noter que la sous-facette "fonction" apparaît sous au moins quatre facettes différentes. Bien que les types de fonctions varient sous chaque facette, les différences sont minimales. Peut-être serait-il légitime de faire de la facette Fonction l'une des principales facettes. Il faudra faire des vérifications et des tests pour déterminer si elle constitue l'une des principales facettes utiles, et pour rendre cette facette plus claire et pratique.

Partant de l'analyse de douze annuaires du Web asiatiques et de trois américains, la présente étude propose une approche par facettes pour classifier les ressources du Web. En tout huit grandes facettes ont été développées sur la base du modèle de Zins. Cette présentation a approfondi l'analyse en identifiant des sous-facettes pour chaque grande facette. C'est l'ébauche d'un système de classification à facettes susceptible d'être utilisé dans différents pays asiatiques. Des travaux supplémentaires sont nécessaires pour rendre le système plus complet et plus utile. Les futures études pourraient se faire dans trois directions. Premièrement des études empiriques pourraient tester le modèle pour vérifier qu'il est applicable et efficace pour classifier les ressources du Web d'autres pays d'Asie. La souplesse de l'approche par facettes pourrait prendre en compte un système de classification plus accueillant culturellement, et on pourrait tester ce modèle pour vérifier s'il est applicable à des pays non asiatiques. Deuxièmement, des études sur les catégories de bas niveau seront nécessaires pour identifier d'autres facettes. Troisièmement, un prototype de l'interface de navigation entre les facettes, basée sur le modèle proposé, pourrait être construit et testé. Les libellés des facettes ont besoin d'être testés et clarifiés pour les utilisateurs. Un modèle à facettes peut aider à organiser les ressources du Web dans la mesure où il facilite l'accès par des utilisateurs différents par la culture et la langue. Pour atteindre un tel objectif l'implication des utilisateurs serait utile.

BIBLIOGRAPHIE

Bates, M. J. (2002). After the dot-bomb: Getting web information retrieval right this time *First Monday*, 7(7) retrieved on November 8, 2005, from: http://firstmonday.org/issues/issue7_7/bates/index.html

Broughton, V. (2002). *Facet analytical theory as a basis for knowledge organization tools in a subject portal*. Proceeding of the Seventh International ISKO Conference, Granada, Spain, 135-141.

Dalberg, (1994). Domain interaction: theory and practice. In: Albrechtsen, H. & Oernagar, S. (eds.). Knowledge organization and quality management: Proceeding of the third ISKO conference, Copenhagen, Denmark, p. 60-71.

Devadason, F., Intaraksa, N., Patamawongjariya, P., & Desai, K. (2001). Search interface design using faceted indexing for web resources. In *Asist 2001: Proceedings of the 64th Asist annual meeting, vol 38*, 224-238.

Ellis, D., & Vasconcelos, A. (2000). The relevance of facet analysis for World Wide Web subject organization and searching. *Journal of Internet Cataloging*, 2(3/4), 97-114.

Gulli, A., & Signorini, A. (2005). *The indexable Web is more than 11.5 billion pages*. Proceedings of the 14th International Conference on World Wide Web, Chiba, Japan, 902-903.

Iyer, H. (1995). *Classificatory structures : Concepts, relations and representation*. Frankfurt/Main: Indeks Verlag.

Kwasnik, B. H. (1999). The role of classification in knowledge representation and discovery. *Library Trends*, 48(1), 22-47.

Kwasnik, B. H., & Rubin, V. L. (2003). Stretching conceptual structures in classifications across languages and cultures. *Cataloging & Classification Quarterly*, 37(1/2), 33-47.

Malt, B. C. (1995). Category coherence in cross-cultural-perspective. *Cognitive Psychology*, 29(2), 85-148.

Olson, H. A. (2001). Sameness and difference: A cultural foundation of classification. *Library Resources &*

Technical Services, 45(3), 115-122.

Ross, K. A., & Janevski, A. (2005). Querying faceted databases. In *Semantic web and databases*, Vol. 3372, 199-218.

Segall, M. H. (1990). *Human behavior in global perspective: an introduction to cross-cultural psychology*. New York: Pergamon Press.

Yee, K.-P., Swearingen, K., Li, K., & Hearst, M. (2003). *Faceted metadata for image search and browsing*, Ft. Lauderdale, FL, United States.

Zins, C. (2002). Models for classifying internet resources. *Knowledge Organization*, 29(1), 20-28.