



HAL
open science

Facettes et traitement documentaire de la photographie : la photothèque du Syndicat des Transports d'Ile-de-France

Marie Quadrini

► To cite this version:

Marie Quadrini. Facettes et traitement documentaire de la photographie : la photothèque du Syndicat des Transports d'Ile-de-France. domain_shs.info.docu. 2010. mem_00575066

HAL Id: mem_00575066

https://memic.ccsd.cnrs.fr/mem_00575066

Submitted on 9 Mar 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



CONSERVATOIRE NATIONAL DES ARTS ET METIERS
INSTITUT NATIONAL DES TECHNIQUES DE LA DOCUMENTATION

MEMOIRE pour obtenir le
Titre professionnel "Chef de projet en ingénierie documentaire" INTD
niveau I

Présenté et soutenu par

Marie Quadrini

le 15 novembre 2010

Facettes et traitement documentaire de la
photographie
La photothèque du Syndicat des Transports d'Ile-de-
France

Jury
Gabrielle DELAGRANGE
Cécile KATTNIG

Cycle supérieur Promotion XL

Remerciements

Je remercie Cécile KATTNIG pour ses précieux conseils.

Je remercie également toute l'équipe de la Direction de la communication du STIF, et en particulier ma responsable de stage Gabrielle DELAGRANGE, pour son accueil.

Merci enfin à ma famille ainsi qu'à Erwan et Marjorie pour leur soutien constant.

Notice

QUADRINI Marie. Facettes et traitement documentaire de la photographie : la photothèque du Syndicat des Transports d'Ile-de-France. 2010. 99 p. Mémoire, Sciences et techniques de l'information, INTD, 2010

Résumé

Ce mémoire évalue la pertinence d'un rapprochement entre les facettes et le traitement documentaire de la photographie. Dans une première partie, les particularités du document photographique et de son traitement sont examinées. Dans une deuxième partie, les différentes acceptions du terme « facette » sont étudiées, ainsi que la compatibilité théorique entre images et facettes. La troisième partie, destinée à évaluer concrètement cette compatibilité, est consacrée à un cas pratique : la construction d'outils linguistiques pour la photothèque du STIF (Syndicat des Transports d'Ile-de-France).

Mots clefs

Document image, image fixe, photographie, photothèque, analyse de l'image, facette, langage documentaire, classification à facettes, thésaurus, vocabulaire contrôlé, indexation, navigation, recherche, web 2.0, transport.

Table des matières

UII	Introduction	7
	Première partie La photographie et son traitement documentaire.....	10
1	La photographie et son traitement documentaire	11
1.1	La photo: un document particulier.....	11
1.1.1	Différences entre document textuel et image fixe	11
1.1.2	Spécificités de la photo.....	16
1.2	La photo et les nouvelles technologies	19
1.2.1	Particularités liées au numérique	19
1.2.2	Métadonnées: les standards actuels	24
1.3	Conclusion	28
	Deuxième partie Traitement documentaire et facettes.....	30
2	Traitement documentaire et facettes	31
2.1	Langages documentaires et théorie des facettes.....	31
2.1.1	Langages d'indexation.....	31
2.1.2	Systèmes classificatoires traditionnels et à facettes	33
2.1.3	La théorie des facettes après Ranganathan	36
2.2	Développement des navigations à facettes sur le web.....	45
2.2.1	Remarques préalables	45
2.2.2	Le web marchand	47
2.2.3	Applications documentaires.....	48
2.3	Conclusion	51
	Troisième partie La photothèque du STIF	53
3	La photothèque du STIF.....	54
3.1	Contexte	54
3.1.1	Le STIF / la direction de la communication.....	54
3.1.2	Le fonds et ses utilisateurs	55
3.1.3	La mission.....	57
3.2	Pertinence des facettes	58
3.2.1	Par rapport au fonds	58
3.2.2	Par rapport aux utilisateurs.....	59
3.2.3	Par rapport aux outils disponibles.....	60
3.3	Constitution des outils linguistiques	61
3.3.1	Méthode et déroulement	61

3.3.2	Construction du lexique	64
3.3.3	Mise en oeuvre	71
3.4	Conclusion	74
	Conclusion	76
	Bibliographie thématique	79
	Annexes	90
	Annexe 1 Thésaurus du STIF	91
	Annexe 2 Proposition de facettes de navigation	98

Liste des tableaux

Tab. 1 Facettes / Sections du lexique	69
Tab. 2 Thèmes / Sections du lexique	71
Tab. 3 Champs du logiciel D / Sections du lexique	72

Introduction

Dans le domaine de la documentation, le mot « facette » se rapporte depuis des décennies à un procédé de construction de classifications étendu par la suite aux thésaurus. Les langages ainsi construits sont applicables à tout type de document.

Or, aujourd'hui, de nombreux sites internet font usage d'une navigation dite « à facettes ». Celles-ci se rencontrent surtout dans le monde du commerce en ligne, mais également, à leur suite, dans les nouveaux OPAC des bibliothèques anglo-saxonnes. Toutefois, comme le laisse deviner le rôle précurseur des sites marchands, les facettes en question émanent rarement de classifications documentaires. Elles s'appliquent d'ailleurs à des objets variés.

Du point de vue des différentes étapes du traitement documentaire (catalogage, analyse et indexation puis mise à disposition), les facettes peuvent intervenir à deux niveaux :

- pour la constitution du langage (ou, éventuellement, à l'indexation si le choix des termes répond à cette logique mais non l'organisation du langage).
- pour la mise à disposition.

En parallèle, l'image occupe une place importante dans notre quotidien tant personnel que professionnel. Les nouvelles technologies en rendent d'ailleurs la manipulation – voire, pour la photographie, la production – très aisée.

Dans ce contexte, sachant que ni les langages à facettes ni la navigation du même nom ne sont réservés à un type de document spécifique, il est naturel de rencontrer des cas d'application des uns et des autres à l'image.

En tant que document, l'image, et notamment la photographie, présente des spécificités. Liées à sa nature, celles-ci déterminent son traitement.

Il est dès lors intéressant d'établir s'il existe des affinités particulières entre photographie et facettes. Si les interfaces de navigation à facettes continuent à se populariser, entre autres dans le domaine de l'image, une telle réflexion pourrait contribuer à les exploiter au mieux.

Afin de traiter cette question, il convient tout d'abord de recenser les particularités du document photographique. Les évolutions déterminantes apportées par le numérique et les outils associés méritent également d'être examinées.

Ce contexte une fois posé, il est possible d'aborder la notion de facettes, de la théorie documentaire originale aux navigations actuelles. Ce panorama vise à clarifier les champs d'application du terme. Il est nécessaire pour évaluer la pertinence théorique d'un rapprochement entre facettes et traitement de l'image.

En dernier lieu, un cas pratique permet d'évaluer concrètement cette pertinence, aux niveaux logique et pratique. Il s'agit de la mise en place d'une photothèque d'entreprise dans le domaine du transport.

Première partie
La photographie et son
traitement documentaire

1 La photographie et son traitement documentaire

1.1 La photo: un document particulier

1.1.1 Différences entre document textuel et image fixe

1.1.1.1 Les différents types d'images fixes

Dans sa définition première, l'image est une représentation d'une chose ou d'une personne. Cette représentation peut être réalisée par de multiples procédés : peinture, photographie, cinéma, etc. Selon le but recherché et le moyen utilisé, la représentation est plus ou moins fidèle au réel et revêt des formes variées.

Dans cette mesure, peut-on traiter les images comme un ensemble homogène du point de vue documentaire ? Il semble que non.

Tout d'abord, la distinction classique entre image fixe et image animée s'impose. Les méthodes d'analyse et les critères de description de l'image animée, dont ce travail ne traite pas, lui sont propres. Ils sont liés notamment à sa durée et à son accompagnement sonore.

La classe des images fixes comprend à son tour une grande diversité d'objets : photos, peintures mais aussi cartes, logos, affiches, illustrations, etc.

Cette variété fait dire à D. Bruckmann que « *l'image fixe n'existe pas* ». Dans cette mesure, il s'interroge sur la possibilité de « *traiter de la même façon des images dont la nature, le statut culturel, les amateurs sont aussi différents* », avant de conclure que les efforts de normalisation, dont il constate l'échec dans le domaine de l'indexation de l'image, « *doi[ven]t s'appuyer sur la pleine conscience qu'il n'y a pas une seule image fixe, mais des images fixes de nature radicalement différente* » ([4], BRUCKMAN, p. 7).

Il propose d'opérer une première distinction empirique entre :

- l'image fixe unique, qu'il est possible d'indexer finement. Les peintures ou dessins entrent souvent dans cette catégorie.
- l'image fixe multiple, c'est-à-dire formant des séries. Elle tend à proliférer. Une indexation plus sommaire peut lui suffire. Elle est d'ailleurs parfois la seule envisageable pour parvenir à traiter tout le fonds. Les photos sous forme de reportages et les affiches sous forme de séries appartiennent à cette catégorie.

1.1.1.2 Caractère analogique de l'image

Un trait commun à toutes les images fixes se dégage cependant : le caractère analogique de leur représentation du réel.

Cela signifie que, pour transmettre l'information, l'image ne se sert pas de code (contrairement au langage) mais copie son objet. Ses emplois les plus utilitaires illustrent ce point. Elle peut par exemple se substituer au texte pour véhiculer du sens plus largement et/ou plus rapidement :

- Quand une population est largement illettrée, l'image (enseignes, vitraux, sculptures, etc.) dispense de l'apprentissage d'un code commun, celui de la lecture, pour communiquer avec le plus grand nombre.
- Aujourd'hui, la production et la transmission d'images est largement simplifiée et démocratisée. Il n'est pas rare que des messages soient transmis directement sous cette forme plutôt que par du texte. Par exemple, une personne présente à une fête et munie d'un téléphone portable peut envoyer un message multimédia pour informer le destinataire sur l'ambiance, le lieu et les invités plutôt qu'un message écrit, dont il serait beaucoup plus long de penser et de mettre en œuvre la rédaction.

Cette différence fondamentale a une incidence sur le traitement documentaire de l'image et le rend distinct de celui des documents textuels. Ce dernier paraît comparativement simple et naturel : le documentaliste se réfère aux mots employés dans le document lui-même (le sommaire, l'introduction, la quatrième de couverture, etc.) pour repérer les mots clés ou rédiger un résumé. Documents primaire et secondaire partagent le même langage.

L'image au contraire, parce qu'elle est analogique, est en elle-même dépourvue de mots. Or c'est par le biais du langage que se fait typiquement l'indexation.

Quelques restrictions sont cependant à apporter à ces deux constats :

- Des textes accompagnent plus ou moins systématiquement les images. Ils en font parfois partie, pour les affiches ou les dessins humoristiques. Dans ce cas, le texte oriente l'interprétation et limite les questions.
- Il existe aujourd'hui des systèmes dits « basés sur le contenu » (content-based) qui traitent de façon automatique et non linguistique un certain nombre d'informations, tels que la couleur, la forme, la texture. Le principe est globalement le suivant : à partir d'une première recherche classique, par des mots, le système propose à l'utilisateur de trouver des images similaires, selon l'un ou plusieurs de ces critères formels. Cela dit, et bien que la recherche soit active dans ce domaine, le traitement de l'image par l'image ne répond pas à tous les besoins. Il ne permet pas encore de faire l'économie des mots.

C'est pourquoi la question demeure : comment analyser l'image, c'est-à-dire la traduire, l'interpréter, tout en lui restant le plus fidèle possible ?

1.1.1.3 Polysémie de l'image et grilles de lecture

Paradoxalement, ce n'est pas à un défaut mais plutôt à une surabondance de mots qu'est confronté le documentaliste qui doit analyser un document visuel. Elle-même muette, l'image vaut mille mots, parmi lesquels il faut sélectionner les plus appropriés.

Est-il pour autant juste de dire qu'elle est polysémique ? Pour Michel Mélot, « *il faut traiter les mots par les mots et laisser l'image à l'image* » ([9], MELOT, p. 362). Cela ne revient pas à abandonner l'indexation de l'image par le langage mais « *à indexer non pas l'image, qui glisse entre les mots comme l'eau entre [les] doigts, mais les questions qu'on pose à l'image* ». C'est dans la variété de ces questions, non dans l'image elle-même, que réside la véritable polysémie. C'est également ce que souligne P. Richard quand il dit que « *l'image apparaît comme un tout et [que] c'est l'opération d'analyse conduisant à l'indexation qui en fait émerger les éléments* » ([16], RICHARD, p. 14).

Bien qu'il n'existe pas de véritable normalisation de l'indexation de l'image, les pratiques ont fait émerger des cadres pour son analyse. Ils appartiennent à deux grandes familles.

La première est relative à celui qui regarde l'image. L'expérience menée par John Sunderland en 1982 illustre la polysémie produite par l'observation ([18], VEZINA) : il est proposé à un enfant de 12 ans, à un profane et à un historien de l'art de faire une indexation libre du tableau de John Everett Millais, « Le Christ dans la maison de ses parents » (1849).

Les résultats sont les suivants :

- l'enfant de 12 ans décrit les objets mis en scène et les caractéristiques physiques des personnages : un homme, une femme, un enfant blessé à la main, des copeaux, etc.
- le profane reconnaît la Sainte famille et l'atelier de Joseph. Il décrit également l'ambiance heureuse et aimante qui règne.
- l'historien de l'art reconnaît des éléments extérieurs (l'artiste, le titre de l'œuvre, sa date d'exécution, son style, etc.), ainsi que la référence biblique exacte et la symbolique du tableau.

Ces éléments démontrent l'importance de la prise en compte du profil des utilisateurs et des questions qu'ils sont susceptibles de poser. Pour reprendre les termes de Michel Mélot, il

n'est pas cohérent d'indexer « *des images muettes sans connaître l'usage qui en sera fait* » ([9], MELOT, p. 362).

La deuxième famille de cadres d'analyse se rapporte à l'image elle-même, dont la description s'articule autour de niveaux de lecture communément reconnus :

- l'aspect technique, qui correspond à la forme (ex : cadrage, temps de pose, lumière)
- l'aspect dénotation, qui correspond au contenu (ex : chien, homme)
- l'aspect connotation, qui correspond au signifié (ex : tristesse, joie)

L'aspect dénotation est lui-même cerné par les cinq questions suivantes : qui, quoi, où, quand et comment ([6], KATTNIG).

La connotation est, logiquement, l'aspect le plus subjectif. Pour la traiter, il est possible de se référer à des couples d'opposition utilisés dans le domaine de la communication et de la publicité, à savoir : abstrait/sensuel, actif/passif, ancien/moderne, apaisant/stimulant, artificiel/naturel, beau/laid, sérieux/frivole, chaud/frais, coloré/terne, gai/triste, comique/tragique, décontracté/angoissé, érotique/froid, ordonné/discordant ([6], KATTNIG, p. 76).

L'indexation de la connotation caractérise le traitement documentaire de l'image : comme le souligne Paula Berinstein ([2], BERINSTEIN, p. 85), il est beaucoup moins fréquent d'indexer les sous-entendus des documents textuels.

Enfin, des cadres de lecture complémentaires peuvent être utilisés. Celui que propose Isabelle Wertel-Fournier pour la photographie d'actualité, par exemple, est axé sur les aspects « je vois que », « je sais que » et « j'interprète sur le plan symbolique ». Chacun est subdivisé en différentes valeurs, associées à des lexiques dédiés. Par exemple, dans la section « je sais que », l'auteur prévoit un lexique pour la localisation géographique, un pour le titre de l'œuvre et les noms propres d'objets, un pour les personnalités morales, un pour les fonctions non visualisables et un pour les manifestations et événements ([6], KATTNIG, p. 77).

D'autres modèles à portée universelle ont été proposés. Certains sont basés, entre autres, sur l'idée que les objets et les scènes représentés sont simultanément génériques et spécifiques ([5], JORGENSEN). Sara Shatford Layne souligne d'ailleurs ce point. Elle donne l'exemple d'une photo du pont de Brooklyn : une telle photo peut être recherchée soit parce qu'elle représente ce pont précis, soit parce qu'elle représente simplement un pont ([17], SHATFORD LAYNE, p. 584).

Ces rationalisations de l'indexation ont pour but de limiter une recherche d'exhaustivité à la fois titanesque et peu utile.

En outre, l'image n'apparaît normalement pas seule mais accompagnée de données textuelles (pour la photographie, son auteur, sa date, sa localisation, sa légende). C'est là sans doute le premier appui du documentaliste pour traiter l'image.

Michel Mélot recommande d'ailleurs d' « *indexe[r l'image] sous les mots qui l'accompagnent, et qui viennent soit de son producteur, soit de ses utilisateurs ou de ses interprètes* » ([9], MELOT, p. 364). Une image totalement dépourvue de texte d'accompagnement est en effet difficilement exploitable.

Les manipulations ludiques auxquelles se livre Pierre Desproges dans son « Dictionnaire superflu à l'usage de l'élite et des bien nantis »¹ permettent de juger de l'importance des informations textuelles dans la compréhension de l'image. Dans ce livre, un même tableau figure devant chaque lettre du dictionnaire, assorti d'une légende différente qui en change complètement la lecture.

1.1.1.4 Rapidité de mémorisation de l'image

Enfin, toujours en raison de son caractère analogique, l'image se distingue fortement du document textuel par la rapidité avec laquelle l'observateur en prend connaissance.

Les répercussions sont importantes. L'utilisateur doit pouvoir aussi bien trouver une image précise qu'accéder à des groupes d'images homogènes et suffisamment étoffés pour

- comparer des images partageant des caractéristiques communes, ce qui est particulièrement utile dans des disciplines telles que l'histoire de l'art.
- pallier des difficultés de verbalisation claire de sa recherche.
- gagner en efficacité dans le cas d'une recherche pointue, plus facile à effectuer en consultant un lot d'images qu'en formulant une requête très précise ([17], SHATFORD LAYNE, p. 585).

Dans ces conditions, l'indexation des images gagne à ne pas être trop fine et le système à donner accès à des groupes d'images cohérents plutôt qu'à des images isolées. Pour les documents textuels, au contraire, le travail de sélection est rendu bien plus lent par la nécessité de déchiffrer au moins le titre et généralement d'autres informations comme le résumé. Il est donc d'autant plus efficace qu'il pèse moins sur l'utilisateur.

¹ DESPROGES, Pierre. Dictionnaire superflu à l'usage de l'élite et des bien nantis. Paris, Editions du Seuil, 1985. 145 p. ISBN 2-02-008658.1

1.1.2 Spécificités de la photo

1.1.2.1 La photographie et le réel

Il s'agit à présent de déterminer les spécificités de la photo et de son traitement documentaire.

Tout d'abord, la photographie semble être la plus analogique de toutes les images. Elle n'est pas, en effet, une recreation du réel mais sa reproduction. Comme le dit Roland Barthes :

« Dans la photographie, en effet - du moins au niveau du message littéral -, le rapport des signifiés et des signifiants n'est pas de « transformation » mais d'« enregistrement », et l'absence de code renforce évidemment le mythe du « naturel » photographique : la scène est là, captée mécaniquement, mais non humainement (le mécanique est ici gage d'objectivité) ; les interventions de l'homme sur la photographie (cadrage, distance, lumière, flou, filé, etc.) appartiennent toutes en effet au plan de connotation » ([1], BARTHES, p. 6).

La photo nous apparaît comme fiable et ayant presque valeur de preuve, ce qui la rend particulièrement capable de tromper.

Elle a notamment la capacité de « *masquer le sens construit sous l'apparence du sens donné* » ([1], BARTHES, p. 7). Sur une photographie publicitaire, par exemple, l'agencement des personnes et des objets peut sembler naturel alors qu'il résulte d'une mise en scène intentionnelle et soigneusement préparée.

Par ailleurs, comme pour les autres images fixes, son sens se construit en partie sur des éléments extérieurs : les textes d'accompagnement. Une photo mal légendée, intentionnellement ou non, peut totalement changer de sens. L'importance de l'exactitude du texte associé est proportionnelle à l'apparente fidélité de la photographie au réel : plus le cliché semble fiable, plus un éventuel contresens aura de portée.

Patrick Peccatte, qui s'est lancé dans un projet de redocumentarisation de photos prises en Normandie durant la Seconde Guerre mondiale intitulé PhotosNormandie, donne à ce sujet un exemple significatif. Sur l'un des clichés, la légende originale indiquait « *des personnes du village ainsi que des soldats américains se sont réunis* ». Or il ne s'agit pas de « personnes du village » mais d'infirmières allemandes faites prisonnières par les Américains. Cette erreur cause un glissement de sens non négligeable ([11], PECCATTE).

Le documentaliste joue donc à ce niveau un rôle déterminant. Il lui appartient de fournir à l'utilisateur les éléments textuels nécessaires à une consultation éclairée de la photographie. Contexte, légende, date et lieu de prise de vue sont autant d'informations essentielles qu'il lui faut collecter et vérifier.

Ce travail s'apparente parfois à une véritable enquête. Il est plus rare pour les documents textuels mais aussi, dans une moindre mesure, pour les autres images fixes. Certaines, moins foisonnantes que la photographie, sont plus systématiquement annotées. C'est le cas de la peinture, qui comporte généralement la signature du peintre, des cartes postales ou des illustrations, qui s'accompagnent naturellement de texte. D'autres, comme les affiches, comportent en elle-même des mots. Le risque de dissociation entre l'image et ses informations associées est alors moindre.

1.1.2.2 Les différents types de photos

Toutefois, les photos elles-mêmes ne forment pas un ensemble homogène.

Il faut tout d'abord distinguer la photographie documentaire de la photographie d'art. La première tend vers le réalisme : son but est de montrer, de donner à voir ce qui est. La seconde assume son point de vue par rapport au réel, voire avec un fort degré de subjectivité, comme n'importe quelle œuvre d'art.

La famille des photographies documentaires se compose des genres suivants, largement repris de Bayle² ([20], BAYLE, p. 17) :

- La photographie de mode
- La photographie de charme
- La photographie sportive
- La photographie d'architecture
- La photographie de portrait
- La photographie de reportage (ou photojournalisme)
- La photographie scientifique et technique

Ces catégories ne sont pas imperméables. Il est aisé d'imaginer, par exemple, qu'un portrait soit également une photo d'art. Elles permettent toutefois de se figurer les adaptations nécessaires au traitement de chaque fonds selon la nature.

Prenons l'exemple de la photographie scientifique et technique. Il est certain que la dénotation (relative au contenu) domine, tandis que la connotation (relative au signifié) est a priori absente. En outre, les légendes sont très pointues et doivent être rédigées par les experts du domaine ([8], KATTNIG, p. 27).

En parallèle, un traitement documentaire s'adapte non seulement au fonds mais à ses utilisateurs. A ce titre, il est intéressant de répertorier les entités qui gèrent des fonds photographiques :

- Les agences de presse et d'actualité
- Les agences d'illustration, dont certaines sont spécialisées dans un domaine précis
- Les agences de photographes et les photographes indépendants
- Les entreprises et établissements publics, notamment leurs services de communication
- Les journaux et groupes de presse
- Les musées et bibliothèques ([6], KATTNIG, p. 15-22)

Il est naturel qu'une agence d'actualités, par exemple, n'envisage pas le traitement de ses reportages de la même façon qu'un musée envisage celui de ses photographies d'art. Ce dernier réservera une place importante à la connotation, fera mention des courants artistiques, donnera des informations précises sur le support et les techniques, etc., non seulement parce que ce type de photographie le demande mais aussi parce que le public visé est susceptible de mener une recherche sur ces critères.

Dans sa dernière partie, ce travail se concentre sur le cas d'une photothèque d'entreprise lancée par un service de communication. Il est donc utile d'apporter quelques précisions sur cette famille de photos.

Il s'agit le plus souvent de reportages commandés par l'entreprise et, comme l'explique Cécile Kattnig, « *portant sur les divers aspects de [ses] activités et sur la vie de l'entreprise* » ([6], KATTNIG, p. 18). Ces photos illustrent des supports de communication et sont communiquées à la presse. La plupart se rapprochent de la photo journalistique. Un certain nombre relève du technique : la création d'une photothèque est souvent l'occasion de rassembler les fonds de plusieurs services.

1.1.2.3 Le reportage photographique

Enfin, la photographie documentaire est rarement un objet isolé. En règle générale, en effet, elle appartient à un reportage sur un sujet donné.

Il importe de conserver ce lien entre les photos d'une même série, pour deux grandes raisons :

- d'un point de vue strictement pratique, une partie du traitement de l'image peut se faire par lots.
- d'un point de vue théorique, l'ensemble du reportage constitue une unité de sens supplémentaire par rapport à la photo seule.

Prenons l'exemple d'un reportage effectué lors d'une inauguration. La photo d'une personnalité faisant un discours peut être recherchée soit pour elle-même, parce qu'elle

² Cette typologie a été réalisée dans le cadre d'un travail sur le site d'une galerie d'art.

représente précisément cette personne ou l'action de discourir, soit en tant qu'étape de l'événement, parce que l'utilisateur souhaite en connaître le déroulement.

D'ailleurs, quand l'unité d'une série a été brisée, sa reconstitution s'avère parfois longue et hasardeuse. Patrick Peccatte évoque ce travail dans le cadre du projet PhotosNormandie, mentionné plus haut (voir p. 16). Il donne l'exemple de trois photos issues d'un même reportage mais pour lesquelles

« aucune des légendes ne mentionne l'existence des deux autres photos en relation, si bien qu'un visiteur peut facilement ignorer que le cliché qu'il examine appartient à un reportage ou à tout le moins à une série » ([11], PECCATTE).

Dans un contexte historique tel que celui-là, la préservation du lien entre les photos d'une même série est particulièrement importante. Du point de vue du documentaliste, elle limite les erreurs d'interprétation. Du point de vue du visiteur, elle donne une idée globale de l'événement représenté et constitue une piste d'extension de la recherche.

Ces propriétés inhérentes à la photographie étant posées, il est nécessaire d'en examiner les évolutions actuelles.

1.2 La photo et les nouvelles technologies

1.2.1 Particularités liées au numérique

1.2.1.1 Le support numérique

La différence fondamentale entre photographies argentique et numérique réside dans leur mode de production. La première résulte d'un phénomène photochimique : la lumière s'« imprime » directement sur le support. La seconde naît d'un code qui traduit la luminosité captée par l'appareil ([3], BOISQUILLON, p. 9-10).

La photographie numérisée, quant à elle, occupe une place à part. C'est une copie d'une photo traditionnelle, créée à des fins de sauvegarde et/ou d'amélioration du confort de consultation. Dans ce cas, bien sûr, les changements induits par le numérique ne concernent que les pratiques associées au document, non la nature du document lui-même. Ce travail ne traite pas spécifiquement de la numérisation. Signalons simplement que c'est un procédé long et coûteux, qui demande une réflexion préalable importante, d'une part sur l'organisation du travail (choix des photos à numériser avec échelonnement par ordre de priorité, mise en place de la chaîne de numérisation, etc.), d'autre part sur les aspects techniques (nature des supports originaux, qualité de l'image numérique finale, etc.). ([6], KATTNIG, p. 56-66).

Il faut souligner que le numérique, apparu à la fin du siècle dernier, n'a pas modifié que les supports. Les pratiques associées ont également évolué. Ces changements peuvent être examinés au prisme des étapes de la chaîne documentaire.

1.2.1.2 Production

La première évolution est quantitative : avec le numérique, le nombre de photos s'est significativement accru.

Cela s'explique au minimum par les facteurs suivants :

- La démocratisation des outils :
 - appareils photos numériques (qui par ailleurs donnent aussi la possibilité de réaliser de petits films), dont le prix baisse et la qualité augmente.
 - appareils ayant une autre fonction originelle mais qui permettent de prendre et de diffuser des photos, comme les téléphones portables.
 - outils facilitant la lecture et l'échange des clichés, tels que des ordinateurs puissants, des logiciels adaptés et un accès à internet en haut débit.
- La simplicité. Les photos numériques sont facilement produites (la capacité d'une carte mémoire est plus importante que le nombre de poses d'une pellicule argentique), visualisées (l'étape du développement n'est plus nécessaire), triées, remaniées, supprimées.

En résumé, c'est la souplesse offerte par le numérique qui séduit et incite à une production plus importante ([15], RAULIN, CACALY, p. 59-60). D'ailleurs, les « *utilisateurs d'appareils photos numériques réalisent 5,5 fois plus de photos qu'avec un appareil argentique* » et « *les Français prennent en moyenne mille photos par an et par personne* » ([15], RAULIN, CACALY, p. 60).

Le phénomène de démultiplication des prises de vue se rencontre surtout chez les photographes amateurs mais n'épargne pas les photographes professionnels. C'est en tout cas le constat du responsable de la médiathèque de la SNCF, rencontré au cours de la mission de stage. Selon lui, les reportages ont tendance à compter beaucoup plus de photos quand ils sont réalisés en numérique. De ce fait, son conseil est de demander, en tant que commanditaire, un affinement du tri par le photographe avant livraison. Il est vrai qu'au STIF, par exemple, les reportages comptant plus de 200 photos ne sont pas rares.

Par ailleurs, en entreprise, certaines photos sont prises non par des photographes professionnels mais par des employés. Dans ce cas, la problématique se rapproche de celle de la photo amateur.

Si le tri n'est pas ou partiellement effectué par le photographe, la tâche échoit au documentaliste. Une définition préalable des critères de sélection devient dès lors nécessaire. Si possible, elle sera établie en concertation avec les utilisateurs. Ce travail de tri exige en outre du documentaliste un bon discernement et un regard avisé.

1.2.1.3 Gestion

Viennent ensuite les étapes du traitement documentaire à proprement parler.

1.2.1.3.1 Traitement par lot

A tous les niveaux de la chaîne documentaire, les logiciels permettent de traiter les images par lots, prédéfinis ou librement constitués. Il est ainsi aisé d'associer à toutes les photos d'un reportage certains éléments de description (date de prise de vue, localisation, contexte, etc.). L'utilisateur, pour sa part, peut télécharger en bloc une sélection personnalisée d'images.

1.2.1.3.2 Création, importation et extraction automatique de données

Au moment de l'intégration des photographies à la base, il est possible d'importer simultanément des informations annexes.

La partie suivante de ce travail (voir p. 24) est consacrée à ces informations, les métadonnées. Il suffit pour l'instant de dire que ce transfert automatique se fait grâce à l'encapsulation des métadonnées, c'est-à-dire leur intégration au fichier image lui-même. Certaines sont automatiquement fournies par l'appareil de prise de vue (date de prise de vue, taille de l'image, temps d'exposition, etc.), d'autres sont renseignées manuellement par l'auteur (titre, légende, etc.). Ce procédé pallie le problème, évoqué plus haut (voir p. 16-17), de la dissociation entre photographie et textes d'accompagnement.

Par ailleurs, il a été vu (voir p. 12) que des systèmes extraient automatiquement des données sur la forme, la couleur ou la texture de l'image. Ces fonctionnalités sont toutefois peu répandues à l'heure actuelle.

Enfin, signalons que des recherches sont actuellement menées dans le domaine de l'indexation automatique des documents visuels. Celle-ci est déjà expérimentée, entre autres, par Google. Elle s'appuie sur le contexte d'apparition des images ([14] RAULIN, p. 54). Malgré tout, l'indexation demeure aujourd'hui encore un travail largement humain.

1.2.1.3.3 Traitement matériel

Toutes les manipulations traditionnellement physiques du document sont désormais effectuées à partir du point unique qu'est l'ordinateur. En quelques clics, les images sont

triées, dédoublonnées ou retouchées. Selon les fonctionnalités disponibles, les retouches sont parfois réalisées dans le logiciel de gestion de photothèque lui-même.

1.2.1.3.4 Catalogage

A moins bien sûr qu'il y ait eu numérisation et conservation de l'original ou que l'original provienne d'un support externe qu'il y a lieu de garder, la notion de classement physique (cote et lieu de stockage) devient caduque.

En outre, en raison de l'évolution des supports, l'image numérique possède un vocabulaire propre : le nom des formats (tiff, gif, jpeg, png, etc.), l'expression de la taille (en pixels) ou de la qualité de la définition (en dpi, Dot per inch ou Point par pouce), etc.

1.2.1.3.5 Indexation

Les nouvelles technologies modifient l'indexation (c'est-à-dire la description intellectuelle des photographies) à la fois dans sa réalisation et, dans une moindre mesure, dans sa nature.

Les bases dites « full web », qui autorisent des contributions à distance, se développent. De ce fait, le travail d'indexation peut être en partie délocalisé et reporté sur le producteur de l'image. L'auteur intègre lui-même ses photos et fournit en même temps des informations qui servent de base de travail au documentaliste (ex : intitulé du reportage, lieu de prise de vue, etc.). C'est le fonctionnement qu'adopte par exemple le site Picture Tank³.

Cette participation accrue des auteurs (phénomène qui ne se limite d'ailleurs pas à l'image) est plus qu'une nouvelle organisation du travail : une indexation réalisée par des non spécialistes présente nécessairement, dans sa nature même, des particularités.

L'indexation collaborative, exclusivement réalisée par des non professionnels, illustre ce point. Dans le domaine de la photo, la plateforme Flickr⁴ est un bon exemple. Il s'agit d'un site web de partage de photos et de vidéos. Il y est fait usage de la folksonomie : les contributeurs « taggent » leurs images, c'est-à-dire qu'ils leur associent des mots clefs créés librement ou choisis parmi ceux des autres contributeurs.

Parmi les caractéristiques d'une telle indexation, la prolifération des synonymes et des formes fléchies est notable.

Par ailleurs, tous les mots clefs sont mis sur le même plan. Les différents niveaux de lecture de l'image, évoqués plus haut (voir p. 14), ne peuvent donc être rendus. Les termes relatifs

³ Picture Tank, <http://picturetank.com> [consulté le 24 septembre 2010]

⁴ Flickr, <http://www.flickr.com/> [consulté le 2 juillet 2010]

à la technique, à la dénotation ou à la connotation, en admettant que le contributeur parvienne à les dégager sans que ce cadre lui soit fourni, sont présentés indistinctement.

Dans un contexte professionnel, bien entendu, ces problèmes sont normalement limités par l'intervention des documentalistes. Ceux-ci réalisent l'indexation ou du moins la vérifient et la rectifient.

Enfin, dans les cas où l'indexation se traduit non par des mots clefs mais par une classification, les nouveaux outils autorisent le classement des documents sous plusieurs catégories sans création de doublon. L'efficacité d'une consultation par ce biais s'en trouve grandement renforcée.

En résumé, le traitement descriptif de l'image, catalogage et indexation, est globalement facilité par les nouvelles technologies. Celles-ci permettent une automatisation de la création, du transfert, voire de l'extraction de certaines données, et favorisent une participation plus active de l'auteur. Au-delà des conséquences pratiques, et dans les cas où la question se pose, ce gain de temps peut être une incitation à ne pas négliger le traitement de la photographie (et non uniquement de la série) comme unité documentaire.

1.2.1.4 Consultation et diffusion

Du côté des utilisateurs, les changements sont également nombreux.

Certaines avancées ne concernent pas que l'image :

- choix entre la simplicité de la recherche sur tous les champs (favorisée par des utilisateurs habitués à Google) et la finesse de la recherche avancée.
- facilité d'une consultation par rubrique, optimisée par la possibilité d'appartenance d'un même document à plusieurs rubriques.
- accès direct au document ou à sa visualisation.
- richesse de la navigation par rebond, qui permet de relancer automatiquement une recherche à partir d'un élément de description du document, et de l'autopostage, qui permet d'élargir une recherche sur un terme plus générique (ex : de Paris à Ile-de-France).

Du côté de la photographie, le confort de consultation a été grandement amélioré. Il est possible de procéder à des ajustements de la recherche en visualisant instantanément les résultats et de sélectionner des images pour constituer des lots personnalisés.

La diffusion des images est elle aussi simplifiée. Dans une base en ligne, tous les publics autorisés accèdent directement aux originaux, selon des modalités prédéfinies. Il n'est plus

nécessaire de reproduire la photographie et de la transmettre physiquement. Celle-ci se trouve en effet au même endroit que sa fiche descriptive et la multiplication des chemins d'accès à une photo ne contraint pas à la dupliquer elle-même.

Enfin, lors du transfert, les informations associées aux photographies sont automatiquement jointes. Ce système évite des déperditions d'information, ce qui est très sécurisant, notamment pour les questions de droits.

1.2.1.5 Conservation

Les modalités de conservation des documents numériques leur sont propres. L'enjeu est de garantir à la fois la conservation des données et la pérennité de la lecture de leur format. Notons que ces questions ne concernent pas uniquement les documents visuels.

Il est important de souligner que la dégradation des supports externes, tels que les CD ou les DVD, peut poser problème. Leur fragilité semble mal évaluée et les photos y sont couramment stockées.

1.2.2 Métadonnées: les standards actuels

Les métadonnées sont des informations fournies sur un document. Elles relèvent de sa description physique (titre, auteur, date, format...) ou intellectuelle (mots clefs, sujet...). Dans le domaine de l'image, il existe des métadonnées techniques et textuelles. Concrètement, elles se présentent sous deux formes : encapsulées, c'est-à-dire intégrées au fichier lui-même, ou externes, c'est-à-dire simplement associées au document, comme dans une base de données.

1.2.2.1 Métadonnées techniques

Il a été vu (voir p. 21) que les appareils de prise de vue fournissent automatiquement certaines métadonnées, qui sont alors toujours encapsulées.

Ces informations techniques portent le nom d'EXIF (EXchangeable Image File) et sont supportées par les formats jpeg et tiff. Elles comprennent, entre autres, la marque et le modèle de l'appareil, la date et l'heure de la prise de vue, les réglages, les dimensions et la résolution de l'image ([8], KATTNIG, p. 24). L'EXIF n'est pas à proprement parler un standard mais la plupart des appareils le proposent.

1.2.2.2 Métadonnées textuelles

1.2.2.2.1 L'IPTC-NAA-IIM

Des métadonnées renseignées manuellement peuvent également être encapsulées.

Elles répondent alors au standard IPTC-NAA-IIM (ou simplement IPTC-IIM). Le standard IIM (Information Interchange Model) a été créé en 1991 par l'IPTC (International Press and Telecommunications Council) et la NAA (Newspaper Association of America) ([8], KATTNIG, p. 24). Né dans le monde de la presse, ce standard est surtout adapté à la photographie d'actualité.

Il s'agit d'un jeu d'une centaine de rubriques qui reprend toutes les informations nécessaires à la description physique et intellectuelle des photographies (nom de l'auteur, date, localisation, mots clefs, etc.). A chaque rubrique est attribué un numéro, qui permet aux éditeurs IPTC (ex : Photoshop, PixVue) d'établir des correspondances avec leurs propres champs.

En règle générale, c'est la version allégée de trente-trois rubriques définie par Adobe que les logiciels de visualisation ou de traitement de l'image proposent⁵. Nous nous contenterons ici de signaler les éléments relatifs à la description intellectuelle de l'image :

- Une rubrique Subject reference, c'est-à-dire Catégorie, de 13 caractères minimum et 236 caractères maximum, répétable. Les catégories sont exprimées par des codes à 8 chiffres à choisir dans une liste de 17 catégories prédéfinies par l'IPTC-NAA. Cette rubrique vise à remplacer les champs Category et Supplemental category à partir de la version 4 de 1997, qui est la dernière.
- Une rubrique unique Keywords, c'est-à-dire Mots clefs, de 64 caractères maximum, répétable.
- Une rubrique Headline, c'est-à-dire Titre, de 256 caractères maximum, non répétable.
- Une rubrique Caption / Abstract, c'est-à-dire Description / Résumé, de 2000 caractères maximum, non répétable.
- La possibilité de créer jusqu'à vingt champs personnalisés (Custom fields).

L'IPTC-NAA-IIM autorise l'emploi d'accents et de signes diacritiques, mais ceux-ci ne sont pas toujours bien gérés par les logiciels ([12], PECCATTE).

Cet ensemble de métadonnées, largement utilisé, a été créé principalement pour les photographies d'actualité. Il n'est donc pas réellement adapté à toutes les photos. La rubrique Mots clefs, unique, impose les mêmes limites que dans Flickr : tous les termes sont mis sur le même plan, quel que soit le niveau d'analyse de l'image auquel ils se rapportent. Ce choix convient bien mieux à la photographie d'actualité, d'où la connotation est pratiquement absente, qu'aux photographies d'illustration ou d'art.

⁵ Pour une présentation complète de la version allégée, se reporter à l'article « Métadonnées : une initiation » ([24], PECCATTE)

Des différences existent néanmoins avec l'indexation collaborative. Dans celle-ci, il n'y a pas de vocabulaire contrôlé préexistant. Dans l'IPTC-NAA-IIM, les mots clefs peuvent être importés à partir d'un système qui en comporte un. C'est là que surgit une importante difficulté : toute la richesse éventuelle de ce vocabulaire (sa structuration, les relations entre les termes) est totalement perdue lors de l'intégration au champ Keywords.

Patrick Peccatte examine cette question dans l'un de ses articles ([13], PECCATTE). A partir d'exemples de légendes et de mots clefs employés pour des photos d'agence, d'actualité et d'illustration, il détecte des problèmes tels que :

- la surindexation, c'est-à-dire la présence de plusieurs mots clefs appartenant à la même chaîne hiérarchique (ex : Arbre et Olivier). En effet, le standard ne permet pas de « *hiérarchiser les mots clefs ou de les qualifier* ».
- la multiplication des formes fléchies et des quasi-synonymes, ainsi que l'indexation prolixe, c'est-à-dire l'utilisation redondante de termes équivalents (ex : Vélo et Bicyclette) insérés automatiquement par le système. L'explication est la suivante : les thésaurus permettent de distinguer un terme descripteur (ex : Confiserie), à utiliser effectivement, de ses équivalents non descripteurs (ex : Bonbon, Sucrierie), qui ne sont pas utilisés pour l'indexation et renvoient simplement vers le descripteur correspondant. Le standard, en revanche, ne fait pas de distinction entre descripteurs et non descripteurs.

Il y a donc une importante déperdition de sens lors du transfert des mots clefs d'une base de données aux champs IPTC-NAA-IIM.

Par ailleurs, dans un autre article, Patrick Peccatte signale qu'

« une enquête récente réalisée conjointement par le CEPIC [Center of the picture industry] et l'IPTC montre que dans le monde professionnel les descriptions textuelles (légendes, captions) sont plus utilisées que les mots-clés pour décrire les images ». ([10], PECCATTE)

Ce phénomène constitue un nouvel exemple de la spécificité d'un traitement documentaire réalisé par des non spécialistes. Cette donnée doit être prise en compte dans le cadre d'une recherche d'images.

1.2.2.2.2 L'XMP

En 2001, Adobe définit un nouveau format de métadonnées, le XMP (Extensible Metadata Platform) pour les images et les pdf.

Le XMP est créé à partir du langage de balisage libre XML (Extensible Markup Language). Il présente l'avantage d'améliorer le traitement des accents et de gérer le multilinguisme dans les légendes ([8], KATTNIG, p. 24). Il s'appuie sur le Dublin Core, un schéma de métadonnées composé de quinze éléments visant à décrire des ressources numériques. Le Dublin Core fait l'objet d'une norme internationale depuis 2003.

En 2005, l'IPTC Core for XMP est développé, dans le but de faire « *la transition entre l'IPTC IIM et le XMP tout en respectant les règles de description bibliographique de l'ISBD* » ([8], KATTNIG, p. 24). Notons que l'ISBD est un ensemble de normes de catalogage. L'une d'elles est consacrée aux « non-book material » (les « non-livres »), catégorie qui comprend l'image fixe.

L'IPTC Core règle une partie des problèmes liés à l'IPTC-NAA-IIM :

« pas de limitation de taille des champs, pas de problèmes d'accents (codage Unicode), possibilité de légendes multilingues, extensibilité et personnalisation des métadonnées » ([12], PECCATTE).

Néanmoins, en ce qui concerne la structuration des mots clefs, les progrès sont minces.

C'est ce que souligne Patrick Peccatte dans la conclusion de l'article « Quelques remarques sur l'indexation des photos » ([25], PECCATTE). Il cite les recommandations du Metadata working group (un consortium regroupant les entreprises leaders dans l'industrie numérique, Adobe, Canon, Microsoft, Nokia et Sony, et travaillant sur les questions d'interopérabilité) à ce sujet :

- la valeur sémantique des mots clefs devrait être conservée (l'exemple du mot « San Francisco », dont on devrait toujours savoir qu'il s'agit d'un lieu, est donné) et, à cette fin, le XMP autorise l'ajout de qualificatifs. Mais aucune explication complémentaire sur la façon de procéder n'est fournie.
- les relations hiérarchiques ne sont pas gérées mais certains outils (dont Adobe Lightroom ou Microsoft Expression media) offrent des possibilités dans ce domaine. Or Patrick Peccatte rappelle que, si ces solutions existent, leur mise en œuvre ne fait pour l'instant l'objet d'aucune standardisation. Leur utilisation risque dès lors de poser plus de problèmes qu'elle n'en résout.

Les métadonnées encapsulées présentent donc l'avantage de souder à l'image un grand nombre d'informations. Toutefois, les standards actuels n'assurent pas encore une transmission fidèle de la richesse du travail d'indexation et de la structuration du vocabulaire contrôlé initialement utilisé.

1.3 Conclusion

L'image se distingue du document textuel par sa représentation analogique du réel. En dépit d'avancées dans le traitement de l'image par l'image, l'utilisation du langage demeure globalement incontournable.

La première conséquence est que les questions posées à l'image, non prédéfinies par des mots, sont multiples. Elles relèvent à la fois de l'individualité de l'utilisateur et de la nature du document lui-même. Pour encadrer son travail face à cette richesse, le documentaliste adopte des grilles de lecture et prend en compte le public visé.

La seconde conséquence est que le document visuel est rapidement mémorisé. Il est de ce fait pertinent de faire reposer une partie du travail de sélection sur l'utilisateur, qui balaie facilement les résultats du regard.

Parmi les images fixes, la photographie se distingue par une plus forte adhésion au réel. Elle peut donner l'apparence du naturel à une mise en scène. Au-delà de ce trait général, il faut toutefois établir une distinction entre les différents types de photos et, au premier chef, entre photographies d'art et photographies documentaires. La prise en compte de ces différences, ainsi que du profil des utilisateurs, détermine un traitement documentaire adapté.

Une autre partie des caractéristiques de l'image découle du contexte actuel : l'apparition du numérique a amené de nombreuses évolutions.

La photographie numérique connaît un grand succès. Elle est financièrement abordable et le développement d'outils dédiés facilite sa gestion, de sa création à sa diffusion. Ses conditions de production peu contraignantes sont d'ailleurs à l'origine d'une forte augmentation du nombre de prises de vue. Ce phénomène implique un plus grand travail de sélection a posteriori, qui incombe au photographe ou au documentaliste.

Par ailleurs, l'association de ce nouveau support et de systèmes de gestion adaptés a significativement modifié le traitement documentaire. Les tâches se sont simplifiées et automatisées. Les pratiques actuelles ont parfois des conséquences sur la nature même du traitement intellectuel, comme l'illustre le cas particulier de l'indexation collaborative.

Enfin, les métadonnées de l'image sont de deux sortes : techniques et textuelles. Elles peuvent être encapsulées, ce qui limite les risques de perte d'informations lors de leur manipulation, ou externes. Si des normes pour le catalogage de l'image fixe ont bien été créées, ce n'est pas le cas pour son indexation. Il existe en revanche des standards pour la définition des métadonnées encapsulées qui prévoient des rubriques de description du contenu. Toutefois, malgré des améliorations, ces standards ne permettent pas de conserver

la finesse de l'indexation produite en externe. Celle-ci conserve de ce fait une forte valeur ajoutée.

Il convient à présent d'examiner ce que sont les facettes, afin d'évaluer quelle place elles peuvent prendre dans le traitement de la photographie.

Deuxième partie
Traitement documentaire et
facettes

2 Traitement documentaire et facettes

2.1 Langages documentaires et théorie des facettes

Pour comprendre dans quel contexte est née la théorie des facettes et quelles sont ses applications possibles, il faut revenir brièvement sur les principaux types de langages documentaires.

Bruno Menon ([29], MENON, diapositives 4-5) propose une première distinction, que nous adoptons ici, entre langages d'indexation (décrivant des thèmes, sujets ou concepts) et systèmes classificatoires (décrivant des catégories de thèmes, sujets ou objets). Il est important de préciser la différence entre concept et sujet, car il y sera fait référence dans la suite de ce travail. Comme l'explique Jacques Maniez, « *le concept fixe les contours du mot [tandis que] le sujet marque l'intérêt pour la chose dénotée par le mot. Assimiler un sujet matière à la somme des concepts qu'il utilise n'est donc pas foncièrement faux, mais réducteur* » ([26], MANIEZ, p. 130). En d'autres termes, un sujet peut porter sur un concept (ou plusieurs s'il s'agit d'un sujet composite, tel que « Le génome de l'homme de Néandertal »), qu'il aborde alors sous un certain angle et dans un contexte donné, tandis que le concept existe hors contexte, comme une définition du dictionnaire.

2.1.1 Langages d'indexation

2.1.1.1 Indexation analytique

L'indexation dite analytique (ou post-coordonnée) consiste à dégager les concepts contenus dans un document. Ce type d'indexation se fait soit en langage libre, c'est-à-dire à l'aide de termes qui ne font l'objet d'aucune normalisation préalable, soit à partir d'un langage contrôlé.

La folksonomie, évoquée plus haut dans le cas de Flickr (voir p. 22), est un exemple d'usage du langage libre : c'est l'utilisateur qui crée le terme qu'il désire employer.

Du côté du vocabulaire contrôlé, se trouvent les listes d'autorités. Elles prennent les formes suivantes :

- des listes de termes normalisés, noms propres compris, généralement présentées en ordre alphabétique.
- des thésaurus. Ce langage s'est développé en même temps que l'informatique, à la fin des années cinquante. Un thésaurus est un ensemble structuré de termes, c'est-à-dire que des relations sont établies entre les mots qui le composent. Il s'agit de relations

générique / spécifique (ex : Canidé / Chien), d'équivalence (ex : Vélo / Bicyclette) ou d'association (ex : Bois / Sylviculture). Quand des termes sont équivalents, l'un d'eux est choisi comme descripteur. Il est alors effectivement utilisé pour l'indexation. Les autres termes sont des non descripteurs. Ils sont répertoriés mais renvoient directement sur le descripteur correspondant, sans être eux-mêmes employés.

Quand un document est indexé par ce biais, les mots clefs n'apparaissent pas combinés les uns aux autres. Par exemple, dans la banque d'images du CNRS⁶, à côté d'une photo de plancton, ils sont simplement listés ainsi : « Plancton, Larve, Biodiversité ».

Ces langages sont souples et intègrent facilement de nouveaux concepts. Dans un contexte informatisé, ils présentent un fort potentiel : les termes sont librement associés à la recherche grâce aux opérateurs booléens ET, OU et SAUF. Pour cette raison, ils sont qualifiés de « combinatoires » ([28], MENON, p. 19).

2.1.1.2 Indexation synthétique

Au contraire, l'indexation est dite synthétique (ou pré-coordonnée) lorsque les termes sont proposés à l'utilisateur sous la forme d'une chaîne reconstruisant le sujet.

Les listes de vedettes matières, qui sont des listes de sujets, constituent une forme élaborée de listes d'autorités. Elles ont précédé les thésaurus. Les exemples mentionnés ci-dessous sont tirés de Rameau (Répertoire d'Autorité-Matière Encyclopédique et Alphabétique Unifié), illustration française de ce langage ([22], BNF).

Le fonctionnement en est le suivant :

- en amont, les termes sont plus ou moins pré-coordonnés (ex : Braille ou Vaisselle d'acier inoxydable).
- à l'indexation, ils sont agencés entre eux selon des règles strictes, de façon à reconstruire le plus exactement possible le sujet du document. L'indexeur produit ainsi une vedette matière finale telle que « Amazone, Vallée de l' -- Découverte et exploration françaises » (soit « tête de vedette -- subdivision + adjectif »).
- à la recherche, c'est la chaîne de termes, soit un syntagme et non un terme isolé, qui est utilisée ([26], MANIEZ, p. 303).
- ces listes sont fermées : les suggestions des utilisateurs sont prises en compte mais seuls les experts font effectivement évoluer le langage.

⁶ Banque d'images du CNRS, <http://phototheque.cnrs.fr> [consulté le 15 juin 2010]

2.1.2 Systèmes classificatoires traditionnels et à facettes

2.1.2.1 Classifications énumératives

Les classifications ne portent pas sur les concepts (comme les thésaurus) ou sur les sujets ou thèmes (comme les vedettes matières) mais sur des catégories de sujets ou de thèmes ([29], MENON, diapositive 6). Tout comme les vedettes matières, et contrairement aux thésaurus, plus universels, les classifications appartiennent souvent au monde des bibliothèques ([28], MENON, p. 19).

Les classifications les plus connues sont la Classification décimale de Dewey, la Classification de la bibliothèque du Congrès et la Classification décimale universelle (CDU). Il s'agit de classifications encyclopédiques, c'est-à-dire visant à couvrir tous les champs de la connaissance.

Dans ces systèmes, à partir de grandes classes de base, des divisions et subdivisions sont opérées, résultant d'affinements successifs. Cela donne, par exemple, dans la Classification Dewey :

600 - Techniques

640 - Vie domestique

641 - Alimentation

641.5 - Cuisine

641.57 - Cuisine pour les collectivités ([19])

Un document portant sur l'hygiène dans les cantines se verra donc attribuer l'indice 641.57. Notons que la notion d'hygiène n'est pas rendue.

Ces systèmes sont énumératifs : ils recensent les sujets en nombre fixe, prédéfini. Pour leurs détracteurs, ils s'avèrent donc « *incapables de tailler des sujets sur mesure adaptés à chaque document* » ([26], MANIEZ, p. 221).

2.1.2.2 Classification à facettes: la théorie de Ranganathan

2.1.2.2.1 Comparaison avec les classifications traditionnelles

C'est justement ce qui amène Shiyali Ramamrita Ranganathan, mathématicien et bibliothécaire indien, à proposer, dans les années vingt, un modèle de classification novateur destiné à pallier cette rigidité : la classification à facettes.

Ce système, dont il dit avoir eu l'idée en observant un jeu de Meccano, est conçu pour éviter qu'un sujet doive impérativement rentrer dans une case unique, préexistante et figée, au prix d'une éventuelle déperdition de sens.

Il maintient un premier niveau de division entre les grands domaines du savoir, comme dans les classifications traditionnelles, sous le nom de « facette principale ». Il leur adjoint cinq facettes fondamentales, qu'il estime être les catégories de base composant tout sujet, les facettes PMEST : Personality, Matter, Energy, Space, Time (Personnalité, Matière, Energie, Espace, Temps). Il faut souligner que Ranganathan déclare son modèle applicable à la fois aux « disciplines (éducation, sidérurgie) et [aux] sujets matières (« l'épidémie de sida en Afrique dans les années 1990 ») [car] la seule différence entre les deux modes de représentation est selon lui une différence de précision » ([26], MANIEZ, p. 220).

Pour illustrer l'innovation apportée par la classification de Ranganathan, Jacques Maniez cite l'exemple suivant ([26], MANIEZ, p. 221). Soit le sujet « La prévention des maladies virales du riz au Japon dans les années 70 ». Dans la classification Dewey, ce sujet aurait été classé sous l'indice 633.189.8, qui dérive de l'indice 633 – Céréales, par restrictions successives :

633 – Céréales

633.18 Riz

633.189 Maladies du riz

633.189.8 Maladies du riz d'origine virale

Deux problèmes apparaissent :

- La notion de prévention n'apparaît pas.
- Le système n'autorise pas de rapprochement entre les maladies virales du riz et celles d'autres céréales. La subdivision monohiérarchique comme seul mode de création de sous-classes ne permet pas de s'émanciper, aux niveaux inférieurs, de l'ensemble Riz, quels que soient les autres concepts contenus dans le sujet. Celui-ci ne peut être envisagé que sous un seul angle.

Dans le modèle de Ranganathan, au contraire, le sujet est tout d'abord découpé en unités, les facettes, selon le schéma PMEST :

- Facette principale : Agriculture
- Facette Personnalité : Riz
- Facette Matière : Maladie virale
- Facette Energie : Prévention
- Facette Espace : Japon
- Facette Temps : Années 70

Cette répartition dans les cinq facettes représente la phase analytique du travail.

Dans un deuxième temps, les indices attribués à chaque unité (EJ pour Agriculture, 381 pour Riz, 421 pour Maladie virale, 5 pour Prévention) sont assemblés dans un ordre prédéfini et selon une notation particulière qui signale pour chaque indice la facette à laquelle il appartient. La facette Energie, par exemple, est signalée par « : »⁷. De ce fait, l'indice final « EJ , 381 ; 421 : 5 : » de l'exemple, est construit un peu comme une phrase.

Cette recomposition du sujet représente la phase synthétique du travail.

En raison de la succession des phases analytique et synthétique dans le processus d'indexation, ce type de langage est qualifié d'analytico-synthétique.

Le principal avantage d'un tel système est sa flexibilité. Les indices, non prédéfinis, se créent au gré des sujets traités et de la combinaison de facettes qui les traduit. Cette souplesse permet ensuite à l'utilisateur de disposer non pas d'un seul mais de plusieurs accès à un sujet. Ainsi, le sujet donné en exemple pourrait être retrouvé dans le cadre de plusieurs recherches : sur les maladies virales, sur la notion de prévention ou sur le riz. Dans une classification classique, le seul accès serait le riz.

2.1.2.2 Limites de la théorie de Ranganathan

Les limites de ce modèle ont néanmoins été soulignées.

Tout d'abord, le terme « facette » est à l'origine de toutes sortes de dérives dans l'utilisation ultérieure du concept. Tiré du vocabulaire courant, ce mot fait l'objet d'interprétations éloignées du sens que lui donne Ranganathan. Certains peuvent considérer, par exemple, que tout aspect de la description d'un document, comme un champ de notice bibliographique, est une facette ([26], MANIEZ, p. 220).

Par ailleurs, il n'est pas dit si les termes composant la classification appartiennent à telle ou telle facette en fonction de leur valeur sémantique propre (expression d'une action, d'un état ou d'une entité) ou de leur fonction syntaxique dans un sujet donné (rôle d'action, d'objet ou d'instrument).

Par exemple, le terme « grille d'analyse » se réfère sémantiquement, en soi, à une entité. Il peut pourtant occuper une fonction syntaxique d'instrument dans un sujet précis (ex : « Indexation des documents à partir d'une grille d'analyse ») ([27], MANIEZ, p. 252). Il n'y a donc pas de coïncidence stricte entre ces deux aspects, sémantique et syntaxique, d'un même mot. Seule la fonction syntaxique d'action est toujours occupée par des termes de même valeur sémantique, qu'il s'agisse de verbes (ex : indexer) ou de substantifs dérivés de

⁷ « colon » en anglais, qui a donné son nom à cette classification, la Colon classification.

verbes (ex : indexation). L'inverse, en revanche, ne se vérifie pas : une valeur sémantique d'action n'empêche pas un mot d'endosser une autre fonction syntaxique. Ainsi « indexation » peut, selon les sujets, jouer un rôle syntaxique d'action, d'objet ou d'instrument.

Précisons que pour les facettes Espace et Temps, cette ambiguïté ne pose guère problème. L'expression de ces notions laisse en effet peu de place au doute.

Ainsi, « *Temps et Espace sont manifestement de nature syntaxique [mais] les trois autres facettes semblent suspendues entre catégories syntaxiques [...] et catégories sémantiques* » ([27], MANIEZ, p. 253).

En raison de ces chevauchements inévitables entre syntaxe et sémantique, la répartition des termes de la Colon Classification dans les différentes facettes « *est assurée par le système, ce qui évite à l'indexeur des difficultés d'interprétation* », notamment entre les facettes Personnalité et Matière ([27], MANIEZ, p. 253). Les choix faits en amont n'en sont pas moins « *largement empiriques* » ([27], MANIEZ, p. 253). Par exemple, sous la facette principale Education, Ranganathan propose l'organisation suivante :

- Cours et programmes, sous la facette Personnalité.
- Propriétés institutionnelles (règlements, vacances, récompenses, etc.), sous la facette Matière.
- Moyens matériels d'enseignement (TV, documents, etc), sous la facette Energie ([27], MANIEZ, p. 253).

Bien qu'il ne fasse pas la distinction entre sémantique et syntaxe car il ne la juge pas pertinente, Ranganathan privilégie donc l'approche syntaxique : TV est placé sous Energie en tant que moyen de mise en œuvre de l'éducation et non sous Personnalité en tant qu'entité.

Enfin, les cinq facettes PMEST elles-mêmes, destinées à s'appliquer à tous les sujets possibles, rendent le système un peu abstrait. Son application à des domaines spécialisés paraît délicate.

2.1.3 La théorie des facettes après Ranganathan

2.1.3.1 L'école indienne

Ces manques ont suscité de nombreuses tentatives d'amélioration. En Inde, le Documentation Research and Training Center (DRTC) a été créé dans les années soixante. L'un de ses membres, Bhattacharyya, a développé le système POPSI (POstulate-based

Permuted Subject Index), dans lequel « *le système des facettes est conservé, mais considéré comme un outil de structuration sémantique et complété par une syntaxe* » ([27], MANIEZ, p. 254). Cela clarifie le critère de répartition des termes sous chaque facette, qui est cette fois celui de la valeur sémantique : « *les catégories fondamentales sont définies par rapport aux concepts, hors contexte* » ([27], MANIEZ, p. 254).

POPSI comporte les facettes suivantes :

- Facette Discipline, qui reprend la facette principale.
- Facette Entité, qui « *désigne tout objet identifiable* » ([27], MANIEZ, p. 254).
- Facette Action, qui reprend la facette Energie.
- Facette Propriété, qui reprend tous les « *attributs des entités ou des actions (intelligence, densité, efficacité)* » ([27], MANIEZ, p. 254).
- Facette Espace.
- Facette Temps.

La syntaxe n'intervient que dans la phase synthétique, quand il s'agit d'ordonner les concepts pour reconstituer les sujets. Elle est d'ailleurs basée sur un modèle plus souple que celui de la Colon classification : elle allie un ordre prédéfini d'enchaînement des facettes, comme le modèle PMEST, à la possibilité de traduire la subordination d'un concept à un autre par un lien dans la notation.

2.1.3.2 L'école britannique

Au Royaume-Uni, la pensée de Ranganathan a également beaucoup inspiré, en particulier à travers le Classification Research Group (CRG), créé en 1952.

L'un des ouvrages de référence du CRG est un livre de Brian Vickery publié en France au début des années soixante sous le titre « Les classifications à facettes : Guide pour la construction et l'utilisation de schémas spéciaux » ([30], VICKERY).

Il développe principalement l'idée qu'une classification à facettes à visée universelle, comme celle de Ranganathan, n'est pas le modèle le mieux adapté à des fonds spécialisés. Or, en dehors du contexte des bibliothèques, c'est un cas de figure extrêmement fréquent. Vickery ne propose donc pas des catégories fondamentales théoriquement applicables à tous les fonds mais simplement une méthode de travail. Celle-ci permet de dégager des facettes propres au domaine étudié, non prédéfinies, et d'organiser une classification complètement adaptée au secteur, selon le principe que « *l'analyse par facettes est essentiellement une analyse conceptuelle des matériaux propres au sujet* » ([30], VICKERY, p.16).

L'auteur préconise de répertorier les termes représentatifs du domaine à partir de la documentation existante puis d'établir pour chacun une définition. A partir de celle-ci une catégorie de rattachement, dans un premier temps librement formulée, est attribuée à chaque mot. L'exemple donné dans l'ouvrage concerne les sols, pour lesquels Vickery détermine que :

- Cohésion est une Propriété.
- Gravier est une Partie du sol.
- Minéralisation est un Processus, etc.

La catégorisation est volontairement arrêtée à un niveau intermédiaire, jugé adapté au secteur étudié, au lieu d'être poussée jusqu'aux facettes fondamentales.

Après avoir traité tous les termes recensés, les catégories sont normalisées. A ce stade, l'étude d'autres classifications dans des domaines parents peut mener à des réajustements. Les catégories ainsi formées sont considérées comme les facettes de la classification. Dans son exemple sur les sols, Vickery obtient dix-huit facettes, parmi lesquelles :

- Propriété du sol, applicable par exemple à Cohésion.
- Partie physique du sol, applicable par exemple à Gravier.
- Processus (à l'intérieur du sol), applicable par exemple à Minéralisation ([30], VICKERY, p.17).

Cette approche très pragmatique contraste avec celle de la Colon classification. Pour Vickery, il s'agit simplement d'identifier tous les aspects nécessaires à la reconstitution d'une classe de sujets dans un secteur donné. Les facettes en constituent les différentes clefs d'accès. Il précise cependant qu'une liste de catégories fondamentales peut servir de guide.

Il est à noter, à ce propos, que le CRG a également proposé une liste de facettes applicables à tous les domaines : Objet, Genre, Partie, Propriété, Matériau, Processus, Opération, Agent, Patient, Produit, Sous-produit (ou dérivé), Espace et Temps ([31], BROUGHTON, p. 53).

Comme le DRTC indien, le CRG fonde son approche sur une organisation des concepts par catégories sémantiques. Remarquons, en ce qui concerne les facettes Agent et Patient que leur intitulé évoque plus une fonction syntaxique qu'une catégorie conceptuelle. En réalité, la fonction d'agent ou de patient se situe bien au niveau sémantique : un outil, par exemple, est considéré comme un agent (ex : burin).

2.1.3.3 L'école nord-américaine

Le problème, mentionné plus haut (voir p. 35), que pose l'emploi du mot « facette » trouve ici son illustration. En Amérique du Nord, l'acception du terme est beaucoup plus large que ce qui a été décrit jusqu'ici. Selon la norme ANSI/NISO Z39.19-200X, il peut en effet s'appliquer « à divers aspects non sémantiques d'un document [notamment] : sujet, auteur, localisation, format, langue et lieu d'édition, [ce qui rend la notion de facette] plus ou moins équivalente à un champ de base de donnée » ([31], BROUGHTON, p. 61). Il est important de souligner ce point pour mieux comprendre l'emploi très variable du mot selon les systèmes, particulièrement sur le web, comme nous le verrons plus loin (voir p. 45).

2.1.3.4 En France

En France, la théorie des facettes ne semble pas avoir connu le même succès qu'en Inde ou que dans les pays anglo-saxons. Cela s'explique peut-être par une approche axée sur la version universaliste de ce modèle. Il est en tout cas significatif que l'AFNOR donne d'une facette la définition suivante « *catégorie de notions de même nature ou exprimées d'un même point de vue telle que phénomène, processus, propriété, outil, permettant un regroupement des notions indépendamment des disciplines traitées* » (AFNOR, « Vocabulaire de la documentation », cité par Jacques Maniez, [27], p. 261), alors que des ouvrages tels que celui de Vickery proposent justement des regroupements de notions propres aux disciplines traitées.

2.1.3.5 La théorie des facettes au-delà des classifications

Bien que leur domaine d'origine soit celui des classifications, les facettes se rencontrent dans d'autres langages documentaires, particulièrement les thésaurus. Dans ce cas, seule la partie analytique est mise à contribution : ce sont les concepts, non les sujets, qui sont classifiés.

La norme AFNOR Z-47 100 sur les thésaurus propose de regrouper les termes selon deux modalités : soit par thème (ex : agriculture, médecine, économie) soit par facette (ex : processus, matériaux, outils) ([27], MANIEZ, p. 258).

Il est toutefois également possible de combiner les deux modes de regroupement. Michèle Hudon suggère ainsi d'établir un « *regroupement préliminaire par thèmes et [une] subdivision en groupes de termes plus restreint sur la base de facettes* » ([24], HUDON, p. 84). Dans le champ sémantique de l'Agriculture, par exemple, le résultat serait :

- Facette Etre vivant : Agriculteur.
- Facette Processus : Plantation.
- Facette Outil : Faux, etc.

L'association du champ sémantique et de la facette présente l'avantage de conserver deux clefs d'accès pour un même terme : un accès vertical, par champ sémantique, et un accès horizontal, par facette.

Enfin, il faut souligner les liens forts entre classification et thésaurus pour comprendre comment les facettes ont pu être si facilement transposées de l'une à l'autre. Certes, leur objet n'est pas le même : la classification organise des catégories de sujets tandis que le thésaurus organise des concepts. Sur le fond, il n'y a pas d'équivalence stricte entre les deux notions (voir p. 31). Sur la forme, en revanche, étant donné que les uns et les autres sont traduits par des mots, les sujets apparaissent bien comme des concepts. Les théoriciens du CRG, d'ailleurs, répartissent les éléments de leurs classifications en fonction de la valeur sémantique des concepts, et ne rendent la syntaxe qu'ensuite, par la combinaison de ces éléments. Quant aux thésaurus-facettes de l'école britannique (dans lesquels la facette et non le thème représente le premier niveau de division), ils sont construits sur le principe qu'une « *classification quelle qu'elle soit est le meilleur point de départ pour la constitution d'un thésaurus* » ([31], BROUGHTON, p. 59). Ce point de vue a été développé par la théoricienne Jean Aitchison, spécialiste de la question.

2.1.3.6 Des définitions et des applications diverses

Les facettes recouvrent donc des réalités différentes et ont plusieurs domaines d'application, ce qui rend le concept délicat à cerner.

Jacques Maniez en donne toutefois une définition globale :

« composantes essentielles d'une structure complexe, dont la séquence ordonnée peut réaliser une recombinaison et un classement rationnel des entités soumises à cette structure ». ([26], MANIEZ, p. 228)

Il propose en outre le découpage suivant, qui n'exclut pas des « *passerelles* » :

- Facettes catégorielles. Elles traduisent des catégories sémantiques fondamentales « *telles que « entités », « actions », « états » applicables à toute classification de concepts indépendamment des domaines* » ([27], MANIEZ, p. 261).
- Facettes structurelles. Elles s'appliquent « *aux objets et aux sujets composites* » ([27], MANIEZ, p. 261). Appliquées aux documents en tant qu'objets, elles peuvent inclure les champs d'une notice bibliographique. Appliquées aux sujets, les facettes structurelles sont typiquement les facettes dégagées par Vickery pour ses classifications spécialisées (voir p. 37).

Ajoutons que le mot « *sous-facette* », parfois employé, se rapporte à un mode d'organisation des termes à l'intérieur d'une subdivision. Par exemple, dans un thésaurus,

sous le terme générique Jouet, les entrées Type de matériau (Jouet d'étoffe, Jouet de métal, etc.), Groupe d'âge (Jouet d'adulte, Jouet de bébé) et Mode d'activation (Jouet mécanique, Jouet électrique, etc.) représenteraient des sous-facettes ([27], MANIEZ, p. 259).

A première vue, il semble n'y avoir qu'une différence de degré entre facettes catégorielles et facettes structurelles, ces dernières étant plus spécifiques. Cela est globalement vrai. Pourtant, l'adaptation à un domaine précis est susceptible d'amener des distorsions. Jacques Maniez prend l'exemple d'une documentation musicale décrite à l'aide des facettes structurelles Compositeur, Exécutant, Forme de la composition, etc. ([27], MANIEZ, p. 256). Dans ce contexte, il est pertinent de recourir à deux facettes distinctes pour Compositeur et Exécutant. Du point de vue du concept, en revanche, une seule facette Artiste (issue de la facette catégorielle Entités, subdivisée en Vivants puis en Humains) serait plus juste, en prévision des cas où le compositeur serait également l'exécutant. La facette catégorielle est en effet basée sur la sémantique propre du concept (ici une personne et en l'occurrence un artiste), non sur son rôle dans une classe de sujets donnée.

Cette distinction rend plus compréhensibles les différences rencontrées dans les réflexions ou les systèmes se référant à la notion de facette.

2.1.3.7 Langages documentaires à facettes et image

Ces principes étant posés, il est possible d'examiner leur application éventuelle aux langages documentaires destinés à l'image.

2.1.3.7.1 Exemples de thésaurus iconographiques à facettes

Le thésaurus de l'INA ([44], INA), par exemple, est organisé selon un système mixte. Il est d'abord divisé en champs sémantiques, de manière traditionnelle. A l'intérieur de chacun, les descripteurs sont répartis sous des facettes : Action, Infrastructure, Matériel, Organisme, Personne, etc. Ainsi, dans le champ sémantique de l'Audiovisuel, la facette Infrastructure introduit les descripteurs Auditorium, Discothèque, Studio d'enregistrement, etc. Notons que ce thésaurus traite exclusivement le contenu des images.

Ce n'est pas le cas, par exemple, de l'Art and Architecture Thesaurus du Getty Institute ([43], GETTY INSTITUTE), aux Etats-Unis. Organisé à facettes dès le premier niveau, ce thésaurus regroupe sans les distinguer les descripteurs relatifs à l'objet et à son sujet :

- Concepts associés. Il s'agit de « *questions de théorie et de critique, idéologies, attitudes, mouvements culturels ou sociaux* » par exemple beauté, équilibre, métaphore, socialisme.

- Attributs physiques. Ce sont les caractéristiques physiques des œuvres, par exemple friabilité, rondeur, dureté.
- Styles et périodes.
- Agents. Ce sont les personnes, caractérisées par « *leur activité, leurs attributs physiques ou psychologiques, leur rôle ou condition sociale* » par exemple paysagiste, ordre religieux.
- Activités.
- Matériaux. Il s'agit des matériaux constituant l'œuvre, par exemple argile, ivoire artificiel.
- Objets, par exemple peinture, amphore, façade.

Dans ces deux exemples, les facettes choisies ne reflètent pas un schéma commun dédié à l'image plutôt qu'à un autre type de document.

2.1.3.7.2 *Rapprochement entre les aspects de la description de l'image et les facettes*

La question se pose donc : existe-t-il des affinités particulières entre la description intellectuelle des documents visuels et les facettes ?

Tout d'abord, il a été vu que l'image se caractérise par plusieurs niveaux de lecture. Si l'on exclut l'aspect technique, ces niveaux sont la dénotation (ou contenu), cerné par les questions qui ? quoi ? où ? quand ? comment ?, et la connotation (ou signifié) (voir p. 14).

Or il est possible d'opérer des rapprochements entre ces aspects et des facettes fondamentales. C'est le parti que prend Sara Shatford Layne en déclarant que les « *sujets d'une image peuvent être classés dans quatre facettes : temps, espace, activités et événements, et objets* » ([17], SHATFORD LAYNE, p. 584).

D'ailleurs, si les appellations de facettes fondamentales proposées par Jacques Maniez sont reprises, à savoir Entités (vivantes et inanimées), Etats, Processus, Espace et Temps ([27], MANIEZ, p. 255), les correspondances suivantes apparaissent :

- Dénotation
 - Qui : Entité vivante, humaine ou animale + Etat (ex : un chien endormi)
 - Quoi : Entité vivante végétale ou inanimée + Etat (ex : une pomme rouge)
 - Où : Espace (ex : Paris)
 - Quand : Temps (ex : en 2010)
 - Comment : Processus (ex : tissage)
- Connotation : Etat (ex : joie)

Les facettes peuvent bien sûr être affinées pour correspondre au contexte d'application.

En outre, organiser le lexique en fonction des aspects de la description de l'image est une option reconnue : en dehors de la question des facettes, Isabelle Wertel-Fournier propose ce type de découpage (voir p. 14).

2.1.3.7.3 Quelques exemples

Il convient à présent d'évaluer la pertinence de tels regroupements à l'aide d'exemples⁸.

2.1.3.7.3.1 Entités vivantes humaines

Pour les Entités vivantes humaines, imaginons une photo d'un petit groupe de professeurs prise à l'occasion d'une manifestation. S'il n'a pas accès à la légende, l'observateur ne sait pas nécessairement qu'il s'agit de professeurs et de manifestants. En revanche, il voit instantanément qu'il s'agit d'hommes ou de femmes, de personnes jeunes ou âgées, etc.

Un document textuel sur le même sujet, en revanche, aurait tout de suite placé le document dans son champ sémantique par le biais des mots « professeurs » ou « manifestation ». Les caractéristiques physiques des personnes, quant à elles, n'auraient sans doute pas été mentionnées. Il a été vu d'ailleurs avec l'exemple du tableau de Millais décrit par un enfant, un adulte profane et un historien de l'art, que ce sont sur ces caractéristiques premières que se fondent les descriptions fournies par l'observateur le moins averti (voir p. 13).

Dans un document visuel représentant des personnes, la facette (c'est-à-dire la catégorie fondamentale de l'élément observé) semble donc être identifiée avant le champ sémantique.

2.1.3.7.3.2 Entités inanimées

Pour les Entités inanimées, cette préséance est beaucoup moins nette. De nombreux objets sont spécialisés et propres à un champ sémantique. Ainsi, face à une photo de locomotive, il apparaît dans un même mouvement qu'il s'agit d'une machine et qu'elle appartient au champ sémantique du transport.

Toutefois, en voyant une photo non légendée d'un bâtiment sans enseigne visible, on décrira d'abord un bâtiment avant de savoir qu'il s'agit, par exemple, d'une bibliothèque et qu'elle appartient au champ sémantique Service d'information documentaire ou d'une piscine et qu'elle appartient au champ sémantique Equipement sportif.

⁸ Les champs sémantiques mentionnés dans cette partie sont tirés du thésaurus Motbis ([45], SCEREN-CNDP)

2.1.3.7.3.3 Processus

Pour les Processus, la situation est à peu près la même : ils sont souvent spécialisés et facilement rapportés à un champ sémantique.

Cependant, là encore, en l'absence de légende, la photo d'une personne attablée devant un livre et des feuilles évoque d'abord l'idée d'activité. Celle-ci peut être apparentée à la lecture mais non précisément identifiée. Le rattachement du processus à un champ sémantique (correction, indexation, traduction, etc.) ne se fait qu'avec un complément d'informations.

2.1.3.7.3.4 Etats

Pour les Etats, la situation est un peu plus complexe car ils se répartissent dans plusieurs niveaux de lecture. La facette fondamentale d'Etat recouvre en effet des termes concrets et abstraits. Les premiers peuvent être utilisés pour décrire des Entités (Qui et Quoi dans la dénotation) tandis que les Etats abstraits qualifient la connotation de la photographie, ce qu'elle évoque.

2.1.3.7.3.4.1 Etats abstraits

Les Etats abstraits (tristesse, gaieté, sensualité, etc.) sont normalement classés dans un champ sémantique dédié. Dans le thésaurus Motbis il s'agit d'Etude psychologique. Il est donc cohérent de les regrouper dans une même catégorie.

Ce regroupement répond d'ailleurs à un besoin de transversalité sensible dans la recherche d'illustrations : une recherche sur la gaieté pourra être satisfaite aussi bien par une photo d'enfants en train de jouer que par une photo de fête ou de discussion.

2.1.3.7.3.4.2 Etats concrets

Les Etats concrets (solidité, chaleur, sommeil, etc.), pour leur part, sont répartis dans différents champs sémantiques. Dans le thésaurus Motbis, par exemple, Chaleur est placé sous Grandeur physique et Sommeil sous Physiologie.

Est-il dès lors intéressant de regrouper tous les états concrets sous une même facette ? Une partie d'entre eux est en tout cas traitée en tant qu'ensemble par les systèmes d'extraction automatique d'informations relatives à la forme, la couleur et la texture des images. Leur utilisation comme critères de recherche est donc perçue comme ayant un sens.

Quoi qu'il en soit, Etats concrets et Etats abstraits correspondent à des niveaux de lecture différents et représentent des critères de recherche distincts. Il semble de ce fait préférable de les maintenir séparés, même au sein d'une éventuelle facette Etat.

2.1.3.7.3.5 Espace et temps

Les termes relatifs à l'Espace et au Temps sont normalement toujours rassemblés dans des catégories propres.

2.1.3.7.3.6 Conclusion

Les regroupements de termes par facettes dans le domaine de l'image semblent assez naturels car ils correspondent globalement à ses niveaux de lecture.

Par ailleurs, un contenu peut être recherché pour son appartenance à un champ sémantique précis ou à sa facette. Quand l'identification des contenus n'est pas évidente, la facette peut parfois même primer.

De ce fait, une structuration à facettes du langage documentaire paraît intéressante. Elle donne un cadre de travail intégré à l'indexeur. Une fois toutes les facettes examinées, celui-ci est certain d'avoir traité tous les aspects : les facettes sont une matérialisation, à l'intérieur du lexique même, des réponses aux questions à se poser face à l'image. Cela n'exclut pas que les facettes fondamentales soient affinées pour s'adapter au domaine traité.

Il convient à présent de déterminer si une telle organisation en amont peut également être exploitée en aval, à la recherche.

2.2 Développement des navigations à facettes sur le web

2.2.1 Remarques préalables

Depuis quelques années, une augmentation significative du nombre de sites web proposant une navigation à facettes est observée. Ces interfaces reposent toutefois rarement sur des classifications à facettes documentaires. Elles s'appliquent d'ailleurs surtout à des objets non documentaires.

Elles utilisent une sélection de caractéristiques partagées par tous les objets ou sujets traités comme autant de filtres à croiser. Ce système permet d'exploiter pleinement toute structuration pluridimensionnelle de données. L'utilisateur choisit n'importe quelle dimension comme point d'accès et peut en combiner plusieurs pour générer automatiquement, de façon dynamique, des sélections correspondantes.

Concrètement, la navigation est généralement conçue de la façon suivante : la recherche commence par une recherche libre, à la Google. Les résultats sont affichés accompagnés d'une liste de caractéristiques communes, les facettes, à partir desquelles il est possible de d'affiner la sélection. Par exemple, pour une recherche sur le mot « pull », l'internaute se

voit proposer un affinage par taille, couleur, matière, prix, etc. A lui de combiner ces différents éléments comme il le souhaite.

Du côté des utilisateurs, les occasions de se familiariser avec des systèmes de ce genre et d'acquérir ainsi de nouvelles habitudes existent, même en dehors du web. La recherche par approche visuelle (view-based searching) de Windows, est connue de beaucoup. C'est l'outil qui permet d'affiner une recherche de fichiers à partir des différentes métadonnées des résultats (nom, date de modification, type, dossier, auteur, etc.). Sa manipulation courante peut contribuer à expliquer le succès d'interfaces similaires ([31], BROUGHTON, p. 62).

Du côté des webmasters, de nombreuses ressources sont disponibles pour ceux qui souhaitent mettre en place une navigation à facettes.

Certaines sont émises par des professionnels de l'information-documentation. C'est le cas par exemple des recommandations de William Denton. Celui-ci se réfère aux théories documentaires des facettes pour expliquer, dans un article en ligne, comment mettre en place une classification à facettes sur tout type de site. Il cite à ce sujet le travail de Louise Spiteri, qui après avoir étudié les théories de Ranganathan et de ses successeurs, en a tiré en 1998 un jeu de règles simplifié ([32], DENTON). Par ailleurs, en 2005, sept interventions sur soixante de la conférence annuelle du ASIST-IA⁹ Summit ont porté sur l'utilisation des facettes, ce qui confirme un regain d'intérêt pour cette conception.

Lorsqu'il s'agit d'applications documentaires, deux sortes de facettes sont à distinguer, bien qu'elles co-existent souvent :

- des facettes portant sur le document en tant que support. La problématique est à peu près la même que dans le cas d'objets non documentaires.
- des facettes portant sur le contenu intellectuel du document. Le langage documentaire utilisé présente alors nécessairement une composante d'organisation par facettes. Soit il a été originellement conçu sur ce principe, soit certaines de ses composantes peuvent être isolées comme facettes. Par exemple, dans une vedette-matière Rameau, la subdivision géographique peut représenter une facette Espace.

D'autres recommandations émanent de personnes a priori étrangères à l'information-documentation. C'est le cas sur le site Facetmap ([35], WILSON), où sont proposés aux webmasters des schémas de métadonnées ainsi qu'un logiciel pour générer une organisation à facettes.

⁹ American Society for Information Science and Technology - Intelligence Architecture

Enfin, notons que ce phénomène s'inscrit dans un mouvement global de développement, sur le web, de systèmes apparentés à des systèmes documentaires par des non documentalistes :

« Plusieurs phénomènes liés aux évolutions du WorldWideWeb et de ses avatars (intranets, extranets, places de marché électroniques) dans les années deux mille ont engendré des artefacts dont les structures et les fonctions s'apparentent à celles des langages documentaires, sans toutefois qu'ils soient issus des mêmes communautés de pratiques (bibliothèques ou centres de ressources documentaires) :

- le développement du Web marchand et celui des sites intranet d'entreprises s'accompagnent de la création de cadres de présentation et de découverte des ressources, les taxonomies ;

- la nouvelle génération du Web, avec le Web sémantique d'abord, puis avec le « Web 2.0 », est à l'origine de réponses originales, en apparence radicalement opposées, à la question de l'annotation / description des ressources numériques : les ontologies d'un côté, les folksonomies et l'indexation sociale de l'autre. » ([28], MENON, p. 22)

2.2.2 Le web marchand

Le premier domaine de pratique de navigation à facettes est le commerce électronique, qui l'intègre pour la consultation des catalogues.

D'après une étude menée en 2005, sur 75 sites marchands, 69% faisaient usage d'une forme de classification à facettes, selon des modalités variées :

- 67% ne proposaient différentes entrées que sur une seule page, généralement la page d'accueil.
- 28% les proposaient sur plusieurs pages, pour affiner une première recherche à l'aide de plusieurs critères.
- 4% offraient une vraie navigation à facettes, à partir d'options de filtrage multiples dès la page d'accueil ([31], BROUGHTON, p. 63-64¹⁰).

Ici, bien sûr, les facettes sont relatives à des classes d'objets, qui sont les produits à vendre.

Prenons l'exemple d'E-bay ([42], E-BAY). Les produits vendus étant multiples, il faut commencer par limiter la recherche à une catégorie d'objets homogènes, par exemple les vêtements de femme. Les critères suivants sont ensuite accessibles : taille, état (neuf ou occasion), prix, matière, longueur, couleur et entretien. Tous peuvent être combinés librement, par le biais de cases à cocher, pour ajuster la sélection.

¹⁰ Broughton rapporte les résultats d'une étude menée par Adkisson.

Pour reprendre la distinction établie par Jacques Maniez, il s'agit de facettes structurelles appliquées à une classe d'objets (voir p. 40). Aux yeux du CRG, ce ne sont d'ailleurs pas à proprement parler des facettes, celles-ci étant réservées à des concepts ou des sujets ([31], BROUGHTON, p. 16). Cette définition exclut de fait toutes les applications marchandes. Le recouplement de cette structuration des sites marchands avec la théorie documentaire des facettes est en tout cas largement fortuit.

2.2.3 Applications documentaires

2.2.3.1 Nouveaux OPAC

Les catalogues en ligne des bibliothèques portent le nom d'OPAC (Online Public Access Catalog). Les OPAC traditionnels proposent aux utilisateurs des modalités de recherche très proches de celles qu'utilisent les professionnels : opérateurs booléens, accès par sujet, etc. Outre des difficultés de compréhension, de tels systèmes risquent de générer du silence. ([34], MAISONNEUVE, TOUITOU, p. 14)

C'est pourquoi depuis quelques années de nouveaux OPAC se développent dans les pays anglo-saxons. Leur but est de s'adapter aux habitudes des usagers, désormais familiers des recherches à la Google, des nuages de mots et de la navigation à facettes. L'intérêt de cette dernière est qu'

« – elle réduit le « bruit » généré par les recherches trop larges. Les angles de vue des résultats proposés à l'utilisateur lui permettent de préciser sa demande en fonction des éléments contenus dans la base : il n'a pas besoin de connaître a priori l'architecture des données, qui lui sont présentées sous forme de facettes ;

– elle « donne à voir » les ressources gérées par la bibliothèque ;

– elle propose à l'utilisateur une stratégie d'affinement de la recherche s'appuyant sur les résultats de sa première requête ; elle permet une démarche itérative laissant l'utilisateur libre de consulter les résultats obtenus ou d'en restreindre la liste jusqu'à l'obtention d'un ensemble jugé suffisant ;

– elle suggère également des solutions auxquelles l'utilisateur n'avait pas initialement pensé. » ([34], MAISONNEUVE, TOUITOU, p. 15)

Il faut insister sur le fait que, paradoxalement, c'est bien le monde de l'information-documentation qui a subi l'influence du web et non l'inverse. Aux Etats-Unis, certaines bibliothèques utilisent même pour leur OPAC des logiciels initialement conçus pour des sites commerciaux ([33], LA BARRE, p. 85).

D'ailleurs, malgré un contexte professionnel spécialisé, certains de ces nouveaux OPAC ne proposent que des facettes relatives au document en tant qu'objet.

Parmi les six exemples d'OPAC de nouvelle génération que comparent Maisonneuve et Toutou (cinq américains et un danois), celui de la bibliothèque de Nashville, aux Etats-Unis, illustre bien ce point. Toutes les facettes y ont trait au support : sa disponibilité, son format, sa collection, sa localisation, sa langue et sa date de publication. Seule la facette Lieu est bien relative au sujet et présentée à part.

En revanche, la bibliothèque du Queens de New-York propose deux sortes de facettes :

- des facettes de notice bibliographique telles que format, auteur, année de publication, langue et collection.
- des facettes concernant le contenu des documents, telles que thème, communauté (c'est-à-dire les groupes humains, par exemple les Catholiques), lieu, personnes, sujet, cadre (ville, campagne...), période et personnage (leur nom ou trait caractéristique). Ce choix n'est pas sans évoquer une

« [...] décoordination des vedettes sujet : la restriction géographique, la restriction à une période historique et la restriction à un nom de personne trouvé dans une vedette matière. C'est la rencontre de la navigation à facettes et d'un répertoire d'autorités matières s'inspirant de la classification à facettes. [...] Répondant à un vrai besoin, cette possibilité de décoordonner des vedettes construites sera fort probablement largement exploitée dans les futurs catalogues. » ([34], MAISONNEUVE, TOUITOU, p. 15)

En France, le Visual Catalog de l'université d'Artois ([41], PAPY) fonctionne différemment. C'est également un OPAC d'un nouveau genre : l'utilisateur commence soit par explorer directement les indices de la classification Dewey, soit par faire une recherche en langage libre. Dans les deux cas, il accède à une page sur laquelle sont mis en regard les notices trouvées, les vedettes-matières Rameau et les indices Dewey correspondants. Dans ce cas, le mode de navigation est bien à facettes mais les facettes elles-mêmes sont particulières : deux facettes sont relatives au support du document (titre et auteur) et deux au langage documentaire (vedettes-matières Rameau et classification Dewey). En ce qui concerne les vedettes-matières Rameau, elles sont présentées sous une forme décoordonnée mais sans qualification de la valeur de chacune, c'est-à-dire sans facette. Ainsi, « Aristote (0384-0322 av. J.C) » et « -- Contribution à la sémiotique » apparaissent sur le même plan, sans que soit signalé qu'il s'agit d'une personne d'une part et d'un sujet d'autre part.

La navigation à facettes est donc un élément important des nouveaux OPAC, surtout dans le monde anglo-saxon. Les facettes y sont néanmoins de nature très différente : elles se rapportent soit au document en tant que support soit à son sujet, sans que cette distinction ressorte. Par ailleurs, il n'y a guère de classifications documentaires à facettes sous-jacentes. Celles-ci seraient pourtant encore plus efficacement exploitées par ces systèmes que les vedettes-matières.

2.2.3.2 Navigation à facettes et documents visuels

Les documents visuels n'échappent pas à ce mouvement d'expansion des interfaces de navigation à facettes. Comme pour tout document, leurs caractéristiques physiques (format, couleur, dimensions, définition...) et les aspects de leur contenu peuvent être traités comme autant de facettes.

Certaines banques d'images, comme Corbis ou Getty images, recourent à un tel système. Mais l'exemple le plus abouti est sans doute Flamenco, une interface développée à l'université de Berkeley ([38], HEARST).

Voici comment est organisée la page d'accueil de cet outil appliqué aux collections du musée des Beaux-arts de San Francisco : l'utilisateur commence par une recherche simple, en entrant un mot dans une case en haut à gauche ou, ce qui est plus rare sur ces interfaces, par une exploration des facettes. Les facettes sont :

- Support
- Localisation géographique
- Objets
- Bâtiments
- Animaux et plantes
- Terre et ciel
- Formes et couleurs
- Métiers
- Artistes

Elles se rapportent soit au contenu, soit au support de l'objet. Notons également qu'aucune facette ne se réfère à la connotation.

Par ailleurs, les facettes sont choisies en fonction du fonds traité. L'interface Flamenco pour les photos de Prix Nobel propose ainsi sexe, pays, affiliation, prix (discipline) et année¹¹.

Une fois les premiers résultats affichés, sous forme de miniatures, les facettes restent visibles pour affiner la recherche ou la relancer complètement sur un nouveau critère.

Ce type de navigation semble particulièrement bien adapté aux documents visuels. Il vise en effet, grâce à ses multiples possibilités de rebond, à éviter le silence. Il favorise en outre la consultation de groupes de résultats, faciles à réduire ou à élargir. Or il a été vu que la recherche d'images se singularise par la rapidité de mémorisation de ces documents (voir p. 15). Elle s'avère plus efficace quand l'utilisateur opère sa sélection finale à partir d'un groupe

¹¹ Flamenco - Nobel prize winners, <http://orange.sims.berkeley.edu/cgi-bin/flamenco.cgi/nobel/Flamenco> [consulté le 15 août 2010]

de résultats pertinent mais dont la présélection n'est pas trop fine. Une étude menée en 2003 dans le cadre du développement du projet Flamenco démontre d'ailleurs qu'à 91% les utilisateurs préfèrent ce type d'interface à une interface classique comme Google image (sauf lorsque la recherche porte sur un objet très précis, tel qu'une rose, auquel cas le taux baisse à 50%). L'une des personnes testées explique que « *la collection semble plus complète car [elle peut] dire combien de documents [sont] disponibles dans les différentes catégories à partir de la page d'accueil* » ([36], YEE, p. 407).

2.3 Conclusion

L'indexation est la description du contenu intellectuel des documents, qu'il s'agisse de leurs thèmes, sujets ou concepts.

L'indexation est dite analytique lorsque ce sont les concepts qui sont dégagés et qu'ils sont présentés sous forme de mots clefs non reliés les uns aux autres. Cette indexation se fait par le biais de langages post-coordonnés, c'est-à-dire prévus pour que les termes ne soient pas combinés en amont mais en phase de recherche. Elle peut être réalisée en langage libre, comme dans les folksonomies, ou à partir d'un vocabulaire contrôlé, comme des listes d'autorités ou des thésaurus.

L'indexation est dite synthétique quand, au contraire, les termes sont reliés entre eux par l'indexeur puis présentés comme un bloc à la recherche. Cette indexation se fait à l'aide de langages pré-coordonnés, tels que les listes de vedettes-matières, qui recomposent les sujets.

Les systèmes classificatoires, pour leur part, servent à décrire des catégories de thèmes, de sujets ou d'objets. A l'origine, les grandes classifications documentaires étaient uniquement énumératives, c'est-à-dire monohiérarchiques (ex : Classification Dewey, CDU). La Colon classification de Ranganathan repose sur un principe plus souple : les indices ne sont pas prédéfinis mais composés selon les sujets à partir de leurs unités minimales de sens, nommées facettes. Pour lui, ces unités minimales peuvent être regroupées en cinq catégories fondamentales : Personnalité, Matière, Energie, Espace et Temps.

Après Ranganathan, de nombreux théoriciens, en particulier indiens et britanniques, ont tenté d'apporter précisions et améliorations à sa théorie. Ils ont notamment tranché en faveur d'une nature sémantique et non syntaxique des facettes et ont précisé, ou étoffé, les catégories fondamentales. Au Royaume-Uni, l'idée de facettes non plus universelles mais adaptées au domaine traité a fait école. Aux Etats-Unis, les facettes ont une acception beaucoup plus large : non limitées au contenu, elles incluent les caractéristiques du document en tant que support.

Par ailleurs, en raison des passerelles naturelles existant entre classification et thésaurus, le principe des facettes s'est étendu à ce dernier langage.

Face à cette diversité de définitions et d'applications du terme, Jacques Maniez propose d'établir une distinction entre facettes catégorielles (fondamentales) et facettes structurelles (spécialisées).

Enfin, pour les images, il est intéressant de voir que les différents niveaux de leur description intellectuelle semblent bien s'accorder avec des facettes fondamentales.

Aujourd'hui, en outre, le terme de « facette » est largement répandu sur le web. Les navigations à facettes se rencontrent fréquemment sur des sites non documentaires. Les webmasters disposent de tous les outils et informations nécessaires pour les mettre en place et les utilisateurs sont disposés à s'en servir.

Dans ce domaine, les sites de commerce en ligne sont précurseurs. Ils proposent comme facettes les différentes caractéristiques de leurs produits. Ce type de facettes relève de l'acception la plus large du terme.

Le monde documentaire, à la suite des sites marchands, s'est lui aussi lancé dans le développement d'interfaces de recherche à facettes. La nouvelle génération d'OPAC anglo-saxons les intègre. Toutefois, les facettes choisies ne sont pas toujours de même nature : certaines se rapportent au document en tant que support et d'autres à son sujet. La proportion des unes et des autres varie selon les applications.

Dans le monde de l'image, l'interface Flamenco est un exemple très intéressant : les facettes sont choisies de façon pragmatique, en fonction de la nature du fonds traité.

Dans tous les cas, ces systèmes sont conçus pour éviter au maximum le silence et constituer des groupes de références. Ils sont de ce fait particulièrement bien adaptés aux documents visuels : vite balayés par l'œil, ceux-ci gagnent à être sélectionnés par l'utilisateur parmi un plus grand nombre de résultats que les documents textuels.

Il convient à présent d'examiner un cas concret de traitement de l'image pour voir s'il est théoriquement pertinent et techniquement réalisable d'organiser le langage documentaire à facettes et de l'exploiter par une navigation du même nom.

Troisième partie

La photothèque du STIF

3 La photothèque du STIF

3.1 Contexte

3.1.1 Le STIF / la direction de la communication

Le Syndicat des Transports d'Ile-de-France, établissement public administratif, est chargé d'organiser, coordonner et financer le transport public de voyageurs de la région. A ce titre, ses premiers partenaires sont les transporteurs : la SNCF, la RATP et les exploitants de lignes de bus privées regroupés dans le réseau OPTILE. Le STIF est régionalisé depuis 2005. Le Président du conseil régional, actuellement M. Jean-Paul Huchon, est à sa tête. L'organisme compte aujourd'hui un peu plus de 200 agents. Il n'y a pas de documentaliste ou de personne chargée de la gestion de l'information.

Le STIF met en œuvre les décisions du Conseil d'administration, où siègent principalement des représentants des collectivités territoriales. Ses activités revêtent de ce fait un aspect politique assez marqué.

Son champ d'action est large : non seulement tous les modes de transport, à l'exception de l'aérien, sont concernés (ferré, routier, fluvial), mais les questions d'aménagement et d'urbanisme sont également largement abordées. En effet, l'élaboration et l'évaluation du PDU (Plan de Déplacements Urbains) d'Ile-de-France lui reviennent. Ce plan, qui détermine l'organisation globale des déplacements dans la région, inclut notamment les modes de transport individuels, motorisés ou non, et le transport de marchandises. Il n'est pas inutile, pour bien comprendre l'étendue de la question, de présenter les neuf objectifs du dernier PDU d'Ile-de-France :

- « Agir sur les formes urbaines, l'aménagement et l'espace public
- Rendre les transports collectifs plus attractifs
- Redonner à la marche de l'importance dans les modes de déplacements
- Donner un nouveau souffle à la pratique du vélo
- Agir sur les conditions d'usage des modes individuels motorisés
- Rendre accessible l'ensemble de la chaîne de déplacement
- Rationaliser l'organisation des flux de marchandises, favoriser le transfert modal
- Construire un système de gouvernance responsabilisant les acteurs pour la mise en oeuvre du PDUIF

- Faire des Franciliens des acteurs responsables de leurs déplacements »¹²

L'équipe de la Direction de la communication, où s'est déroulée la mission, est composée de dix personnes : une secrétaire, une assistante de communication, sept chargés de projet et un directeur. Les activités du service sont classiques : publication de supports de communication externe et interne, coordination de la gestion des différents sites internet et de l'intranet, organisation d'événements et de campagnes de communication, etc. Un grand nombre est mené en partenariat avec des prestataires extérieurs, en particulier des agences de communication.

3.1.2 Le fonds et ses utilisateurs

3.1.2.1 Le fonds

La photographie joue un grand rôle dans un service de ce type car les besoins en illustrations y sont naturellement importants. Ils varient selon le public visé : photos d'agents ou d'événements internes pour les supports de communication interne, photos de véhicules, de points d'arrêt, de la ville, de voyageurs et d'événements externes pour les supports de communication externe ou pour la presse.

Pour y répondre, la Direction de la communication dispose d'un fonds d'environ 10 000 photos, toutes nativement numériques, au format jpeg ou tiff, et stockées jusqu'ici sur des CD-rom ou sur un serveur partagé interne. Certaines comportent des métadonnées EXIF et parfois IPTC. Les plus anciennes datent de 2003. Le nombre de reportages annuel a significativement augmenté depuis 2008. Quand les photos disponibles en interne ne répondent pas aux besoins, il arrive que la RATP ou la SNCF soient sollicitées, souvent par le biais des agences partenaires.

Les auteurs des photos sont des agents du STIF, qui font ou non partie de la Direction de la communication, ou des photographes professionnels. Ces derniers sont mandatés par la Direction de la communication, plus rarement par d'autres services, ou par des agences partenaires.

Les photographes professionnels réalisent des reportages sur des sujets tels que :

- Des événements internes : vœux de bonne année, réunions du conseil d'administration avec portraits de ses membres, visites techniques.

¹² STIF. Rapport d'activités 2009. In [Site internet du STIF] [en ligne], Paris, 2010 [consulté le 15 juillet 2010]. Accès libre, http://www.stif.info/IMG/pdf/RA2009_INTERNET_Complet.pdf. p. 25

- Des événements externes : inaugurations, conférences de presse, salons, réunions avec des organismes partenaires, signatures de contrats, remises de prix.
- Des reportages thématiques : sur l'accessibilité, l'intermodalité, les espaces urbains avant le lancement d'un projet, etc. Beaucoup portent sur un type de véhicule (visuels intérieurs et extérieurs, véhicules à l'état de prototypes, lors de leur lancement ou en service) ou sur un mode (bus, métro, tramway, etc.).

Certains événements représentent une étape d'un projet, qui peut lui-même faire partie d'un programme plus global. Les reportages sur le Francilien (nouvelle rame automotrice du réseau Transilien), par exemple, ont été réalisés à l'occasion des essais, d'une conférence de presse ou de l'inauguration : ces événements marquent les différentes étapes du projet Francilien, qui s'inscrit à son tour dans une démarche générale de modernisation et d'innovation.

Les agents du STIF, quant à eux, prennent également des photos sur ces sujets. Toutefois, l'essentiel de leur production porte sur d'autres événements internes, tels que séminaires, voyages du comité d'entreprise ou travaux d'aménagement des locaux. Quand leur qualité le permet, ces photos sont utilisées comme celles des professionnels.

Mentionnons également la présence, dans un autre service, d'un fonds de photographies plus techniques, également prises par des agents du STIF. Pour l'instant, ces clichés ne sont pas adjoints au fonds de la Direction de la communication. Ils ne sont donc pas traités ici.

3.1.2.2 Les utilisateurs

Les utilisateurs sont essentiellement :

- Les agents de la Direction de la communication et parfois d'autres services du STIF.
- Les agences de communication partenaires.
- Les collectivités territoriales et les transporteurs (pour leur propre communication).
- Les journalistes.

Actuellement, pour faire leurs demandes de photos, les partenaires extérieurs et les journalistes contactent le chargé de projet avec lequel ils sont en relation. Celui-ci grave les clichés demandés sur un CD et le leur fait parvenir (sauf quand la demande porte sur un petit nombre de photos, auquel cas la messagerie électronique ou une plateforme de transfert de fichiers peut convenir). Précisons que les journalistes travaillent pour la presse généraliste ou spécialisée, pour des titres tels que « Ville, rail et transports » ou « Bus et car magazine ».

Globalement, les recherches se font à partir des reportages récents, en particulier pour les photos de personnes. Il y a néanmoins des besoins ponctuels de photos plus anciennes, pour illustrer le changement apporté à un véhicule par une rénovation par exemple.

Il apparaît donc que, selon leur usage, ces photos peuvent être considérées comme des photographies d'illustration ou d'actualité (quand le but est l'illustration, il ne s'agit jamais de notions purement abstraites). Enfin, quelques-unes se rapprochent du technique.

3.1.3 La mission

Des problèmes majeurs de gestion des photographies se posaient :

- Dispersion des photos. Certaines se trouvaient sur des CD-rom, d'autres sur le serveur partagé.
- Pauvreté des métadonnées. Bien souvent seul l'intitulé du reportage et sa date étaient mentionnés sur le CD lui-même ou dans le nom du fichier. Le nom du photographe n'apparaissait pas toujours et parfois le titre du reportage ou la date manquait également. Dans la majorité des cas, ces informations ont pu être retrouvées sur les factures, du moins pour les photographes professionnels, ou collectées auprès des agents du service.
- Incertitude sur les droits associés. La situation allait de l'absence de mention à une mention détaillée mais ne répondant pas aux conditions désormais exigées par le STIF.
- Dégradation des supports. Certains fichiers stockés sur CD-Rom n'étaient plus lisibles.
- Lourdeur de la transmission des photos aux partenaires extérieurs. Il était généralement nécessaire de graver un CD, ce qui posait un double problème : le procédé est fastidieux et les métadonnées non encapsulées risquent de ne pas être jointes.

Pour ces raisons multiples, la Direction de la communication a souhaité mettre en place un logiciel de gestion de photothèque (projet piloté par l'une des chargées de projet). Celui-ci serait accessible aux personnes autorisées via un navigateur web. Sa mise en place, son hébergement et sa maintenance technique seraient assurés par un prestataire, ainsi que la formation des gestionnaires et utilisateurs internes. En fonction de ses droits, l'utilisateur pourrait visualiser ou non les différentes séries de photos, les commander puis les télécharger en ligne. Il serait demandé aux photographes (ou aux agences servant d'intermédiaire le cas échéant) d'alimenter la base directement, à distance, et de renseigner en même temps un certain nombre de métadonnées. Deux chargés de projets valideraient les informations fournies et complèteraient l'indexation. Le prestataire devrait en outre assurer l'indexation du fonds existant, partager son expertise et être disponible pour l'indexation des photos à venir au cas où les deux chargés de projet se trouveraient débordés.

Dans ce cadre, la mission comportait trois volets :

- Rédaction d'une version préparatoire d'un cahier des charges.
- Regroupement des photos et de toutes les informations utiles les concernant, notamment sur les droits.
- Constitution des outils linguistiques pour l'indexation.

C'est ce dernier aspect qui est traité ici.

3.2 Pertinence des facettes

3.2.1 Par rapport au fonds

3.2.1.1 Pour la constitution des outils linguistiques

En ce qui concerne la nature du fonds du STIF, deux points peuvent être soulignés :

- il s'agit d'un fonds iconographique. Il a été dit que les différents niveaux de lecture de l'image, notamment les cinq questions à se poser pour décrire son contenu, peuvent être rapprochés des facettes fondamentales :
 - Entités (entités animées, dont les humains, ou inanimées) – Qui et Quoi
 - Etat – Attributs des Entités ou du contenu de la photo
 - Espace - Où
 - Temps - Quand
 - Processus – Comment (voir p. 42)
- il s'agit d'un fonds spécialisé. Brian Vickery propose de constituer des classifications spéciales pour ce type de fonds : des catégories de termes sont créées à partir du vocabulaire en usage dans le secteur. Dans le cas présent, il semble qu'une classification à proprement parler ne soit pas vraiment appropriée, en particulier au vu des outils en usage (voir p. 62). Cependant, certains théoriciens considèrent ces classifications comme des points de départ adéquats pour la constitution de thésaurus à facettes spécialisés (voir p. 40). Dans cette optique, la méthode de Vickery est exploitable pour la construction d'un thésaurus.

Par ailleurs, l'exemple de l'INA montre qu'un thésaurus de l'image peut très bien comporter, au même titre que les autres, des regroupements de termes par facettes (voir p. 41).

Il existe donc des affinités entre, d'une part, la nature iconographique du fonds et une construction du thésaurus qui tient compte des facettes fondamentales et, d'autre part, sa nature spécialisée et l'utilisation de facettes adaptées.

3.2.1.2 Pour la consultation de la base

En parallèle, la question se pose de la pertinence d'une navigation à facettes pour ce fonds.

Les systèmes de ce genre, notamment Flamenco, semblent bien adaptés aux documents visuels (voir p. 15 et 50). Les images gagnent en effet à être présentées en groupes. Il est en outre pratique d'élargir ou de restreindre une recherche à partir d'un lot de résultats toujours visible. Il semble d'ailleurs que les utilisateurs préfèrent une navigation à facettes à une interface classique de recherche d'images (voir p. 51).

Il faut néanmoins souligner que ces interfaces se caractérisent par la possibilité de croiser plusieurs critères prédéfinis comme autant de filtres, tout en visualisant directement les résultats. Pour que ces filtres présentent un intérêt, ils doivent s'appliquer à un nombre initialement assez important d'éléments, faute de quoi ils sont certainement superflus.

Or le fonds du STIF est pour l'instant de taille modeste. Une telle sophistication paraît donc un peu surdimensionnée voire contre-productive : la multiplication des points d'accès, dont certains donneraient peu de résultats, risquerait de désorienter l'utilisateur. Pour l'instant, de grands thèmes peu subdivisés, offrant chacun un seul point d'entrée, suffisent à la consultation guidée de la photothèque. Cela est d'autant moins contraignant que les logiciels permettent d'affecter plusieurs thèmes à une même photo.

Toutefois, qualitativement, une navigation à facettes répondrait certainement aux besoins. Prenons l'exemple d'une demande formulée par l'une des chargées de projet : pour illustrer un support de communication, il lui fallait une photo de tramway, de préférence arrêté à une station, et sur laquelle figureraient des voyageurs. Quelques photos répondant strictement à ces critères lui ont été proposées, ainsi que quelques autres représentant l'intérieur d'un tramway avec des passagers à bord mais en train de rouler. Celles-ci l'ont également intéressée. Une navigation à facettes permettrait de mener facilement à bien une telle recherche. Les critères essentiels (ici le tramway et les voyageurs) seraient croisés puis les critères secondaires (ici l'arrêt à une station) simplement ajoutés ou supprimés en fonction des résultats obtenus.

3.2.2 Par rapport aux utilisateurs

Tous les utilisateurs de la base, du moins dans la phase de gestion courante, seront des non documentalistes :

- les photographes, qui déposeront leurs clichés et fourniront une première série d'informations (en langage libre uniquement).
- les deux chargés de projet, qui valideront les données fournies par les photographes, indexeront les photos à l'aide de mots clefs et pourront en modifier la liste.

- les utilisateurs finaux, qui consulteront la base.

En amont, pour les indexeurs, il est donc important que l'organisation du lexique soit la plus lisible possible. Aussi pratiques qu'elles puissent être, il est dès lors exclu de faire apparaître les facettes fondamentales sous leur appellation standard, soit par exemple Entité, Processus, Espace et Temps. Les intitulés d'éventuelles facettes propres au domaine traité semblent plus naturels et plus adaptés. Elles peuvent d'ailleurs être compatibles avec des facettes fondamentales. A cette condition, une telle organisation du lexique présente un double avantage : faire apparaître dans sa structure même les différents aspects à aborder pour décrire à la fois tout type d'image et toute photo de ce fonds précis. Cela permet de guider le travail de l'indexeur, ce qui est sans doute d'un intérêt particulier quand il s'agit pour lui d'une tâche annexe et non de son métier.

En aval, en l'absence de navigation à facettes, les utilisateurs finaux ne seraient confrontés à cette organisation du lexique qu'en recherche avancée (à condition que le système permette de faire ressortir chaque facette en lui attribuant un champ spécifique). Toutefois, peu de non documentalistes se servent de la recherche avancée. Ils recourent en priorité à la recherche simple à la Google, c'est-à-dire en langage libre dans un champ unique, ou à la consultation guidée par un accès thématique. Dans un cas comme dans l'autre, les facettes ne constituent pas ici un outil de recherche intéressant.

3.2.3 Par rapport aux outils disponibles

Les plupart des outils de gestion de photothèque actuellement disponibles ne prévoient pas d'organisation à facettes, que ce soit dans la structuration du lexique ou dans la navigation.

Ce sont essentiellement les quatre logiciels sélectionnés par le STIF avant le choix final qui ont été examinés en détail : A, B, C et D, qui a finalement été retenu, dans une version partiellement pré-paramétrée. Sur la question qui nous intéresse, ces logiciels semblent plutôt représentatifs. Les remarques qui suivent sont basées sur la lecture de l'offre qu'ils ont faite au STIF, la consultation de leur site web et des versions de démonstration en ligne et/ou des photothèques de leurs clients quand celles-ci sont en libre accès.

3.2.3.1 Organisation concrète du lexique et facettes

Il est globalement possible d'organiser le lexique à sa convenance. Tout d'abord parce que les logiciels proposent d'établir entre les termes toutes les relations existant dans un thésaurus (hiérarchie, équivalence, association). Ensuite parce que le lexique peut être réparti dans plusieurs champs, que ceux-ci soient prédéfinis (le logiciel D, par exemple, propose dans sa version pré-paramétrée six champs multi-valués pour les mots clefs) ou

entièrement paramétrables. Si le lexique comporte des facettes au premier niveau, il est dès lors facile de les présenter distinctement les unes des autres, dans la limite du nombre de champs disponibles, à l'indexation et éventuellement à la recherche.

En revanche, quand le système ne gère pas les relais virtuels¹³, ce qui est le cas du logiciel D, il ne peut y avoir de facettes à l'intérieur des grandes sections du lexique. L'emploi de sous-facettes ou de conjonctions entre thèmes et facettes est, de ce fait, exclu (voir p. 39 et 40). Les sous-facettes génèrent en outre de la polyhiérarchie, qui n'est pas gérée non plus.

Toutefois, ces facettes sont surtout utilisées quand il existe un grand nombre de termes spécifiques. Dans le cas présent, leur absence n'est pas problématique.

3.2.3.2 Consultation de la base et facettes

Les interfaces de consultation, pour leur part, ne prévoient pas de navigation à facettes.

Elles proposent généralement sur la première page une recherche simple, un accès par thèmes, un lien vers la recherche avancée et un autre vers le thésaurus.

Signalons le fonctionnement particulier du logiciel B : en plus de la recherche simple, les différents critères sont présentés dans une colonne à part, affichée sur toutes les pages. Sur le même principe, le logiciel A offre un accès aux thèmes à toutes les étapes, ce qui permet de relancer la recherche sur ceux-ci à n'importe quel moment de la consultation. Il ne s'agit cependant pas de navigations à facettes car les données ne peuvent être croisées : cliquer sur l'une d'elles relance complètement la recherche.

Pour l'instant, la mise en œuvre d'une navigation à facettes dans le monde de l'image avec un outil d'usage courant semble donc difficile. Cela n'est pas surprenant si l'on considère que Flamenco (voir p. 50), interface novatrice, ne date que de quelques années.

Il convient à présent d'examiner les outils linguistiques pour déterminer la place effective des facettes dans leur construction et leur utilisation, à l'indexation et à la recherche.

3.3 Constitution des outils linguistiques

3.3.1 Méthode et déroulement

La description qui suit ne reprend pas dans le détail toutes les étapes du travail effectué mais est simplement destinée à cerner ses grandes lignes et à situer la réflexion menée sur les facettes.

Le choix du type de langage a dépendu du fonctionnement des logiciels : un lexique d'une part et une liste de thèmes pour une consultation guidée d'autre part. Par ailleurs, les outils sélectionnés gèrent les relations hiérarchiques, d'équivalence et d'association, et permettent l'autopostage à l'indexation et à la recherche. De ce fait, le lexique a été conçu comme un thésaurus.

Le premier objectif était le lexique, à partir duquel les thèmes seraient constitués. Les sujets des reportages ont d'abord été listés à partir des informations textuelles fournies et d'une visualisation rapide des photos. Une fois tous les reportages examinés, des familles de sujets assez larges ont été formées : événements, internes ou externes, et reportages thématiques sur un type de véhicule, un mode de transport ou un axe de réflexion. Dans le même temps, des informations sur l'utilisation des photos et les utilisateurs eux-mêmes ont été collectées auprès des chargés de projet du service. Ont ainsi été posés a priori l'inutilité d'un rendu de la connotation et des besoins réduits en matière de description des caractéristiques formelles des clichés.

Ensuite, une centaine de photos de chaque famille a été examinée dans le détail. Tous les termes utiles à leur description ont été notés, en tenant compte des cinq questions qui, quoi, où, quand et comment, ainsi que des aspects techniques. En ce qui concerne ces derniers, seuls les plus saillants ont été relevés (portrait, prise de vue nocturne, vue intérieure ou extérieure, notamment pour les véhicules, vue semi-aérienne, etc.), toutes familles d'aspects confondues (cadrage, angle de prise de vue, période, etc.). Pour la dénotation, à partir de la première série de mots, des regroupements sémantiques et des rapprochements entre les synonymes ont été opérés.

Une première répartition des termes relatifs au contenu en modes, véhicules, infrastructures et équipements, personnes, projets, événements et phases d'événements s'est rapidement dessinée. Il est apparu qu'un même sujet pouvait être abordé selon un ou plusieurs de ces angles de vue et surtout que leur assemblage (assorti de la localisation géographique et de la date précise de la prise de vue) suffisait à décrire toute photo du fonds, même si tous les aspects n'étaient pas effectivement utilisés pour chacune.

La théorie des facettes a servi de guide de travail : la méthode de Brian Vickery (voir p. 37) a permis de dégager des facettes adaptées au fonds et les facettes fondamentales ont encadré une organisation du lexique apparemment pertinente pour l'image.

A partir de là, la consultation de l'ensemble des photos ainsi que la lecture des publications du STIF (sites internet, intranet, plaquettes spécialisées, rapport d'activités...), des

¹³ Dans les thésaurus, les relais virtuels sont des termes servant à structurer les descripteurs

recherches sur internet, et notamment sur Wikipédia pour la typologie des véhicules, ainsi que des questions posées à l'équipe ont permis de compléter ces catégories et d'éclaircir les liens entre les concepts.

L'ensemble des images a alors été survolé pour vérifier que la liste de termes n'omettait aucun aspect important de leur contenu. A ce stade, la principale difficulté portait sur le niveau de détail à adopter : le fonds était réduit mais certains utilisateurs seraient pointus dans le domaine (journalistes spécialisés, agents du STIF, transporteurs). Un aspect de la dénotation s'est toutefois trouvé restreint d'emblée, celui du vocabulaire technique relatif aux éléments des véhicules, très peu représentés.

Une fois la liste jugée complète, un certain nombre de synonymes a été ajouté et la normalisation des termes a commencé. A partir des sources déjà évoquées, du thésaurus de l'UIC¹⁴ et du thésaurus Urbamet¹⁵, spécialisé dans l'urbanisme, certains termes techniques ont été revus.

Les relations hiérarchiques ont ensuite été créées. Les théories sur l'utilisation des facettes dans les thésaurus ont servi d'appui : facettes de premier niveau, croisement de thèmes et de facettes, sous-facettes. Cependant, la question de leur mise en œuvre concrète restait en suspens car il n'était pas encore certain que le logiciel les accepte.

Les relations d'association ont alors été complétées, plusieurs ayant déjà été établies au fur et à mesure.

Sur cette même période, les médiathèques de la RATP et de la SNCF ont été contactées pour comparer leurs outils linguistiques avec le thésaurus produit. Le thésaurus de la SNCF, un outil dédié développé en interne, a été consulté.

Les grands thèmes ont ensuite été définis à partir du lexique.

Enfin, une rencontre d'une journée avec une documentaliste consultante du prestataire a servi à définir le paramétrage de la base. Elle a également été l'occasion de prendre l'avis d'une professionnelle expérimentée sur le thésaurus et d'y apporter quelques rectifications.

mais qui ne sont pas eux-mêmes répertoriés.

¹⁴ UIC (Union internationale des chemins de fer) - CEMT Conférence européenne des ministres des transports. Thésaurus transport [en ligne]. Paris, 1995, mis à jour le 25 mai 2010 [consulté le 10 juillet 2010]. <Accès libre, <http://www.uic.org/spip.php?article1702>>

¹⁵ Ministère de l'écologie, de l'énergie, du développement durable et de l'aménagement du territoire – Commission thésaurus de l'association Urbamet. Thésaurus Urbamet [en ligne]. Paris, Association Urbamet, 1986, mis à jour le 7 mai 2010 [consulté le 10 juillet 2010]. <Accès libre, <http://www.urbamet.com/thesaurus/thesaurusurbamet.htm>>

La mission ayant pris fin avant la mise en place définitive de la photothèque et l'indexation de l'existant, il manque une phase d'évaluation pratique de la pertinence des langages produits. Signalons à ce sujet que le logiciel autorise l'emploi de candidats descripteurs et donne des informations sur les recherches n'ayant pas abouti.

3.3.2 Construction du lexique

Il convient à présent d'examiner le rôle concret des facettes, dans leurs différentes acceptions, dans la construction du thésaurus. Celui-ci peut être consulté dans son intégralité en annexe (voir annexe 1, p. 91).

Les facettes du domaine ont été dégagées en allant du particulier au général, selon la méthode de Brian Vickery ([30], VICKERY). Mais dans le cas étudié il n'y a pas de contradiction, seulement une différence de degré, entre ces facettes et les facettes fondamentales. La présentation qui suit est donc basée sur une première organisation par niveaux de lecture de l'image / facettes fondamentales, cette correspondance rendant le modèle généralisable. Sous chaque entrée, les facettes spécialisées sont détaillées, ainsi que les concepts qu'elles recouvrent.

3.3.2.1 Dénotation

3.3.2.1.1 Qui ? (facette Entités > Entités animées > Humains)

- Personnes physiques : elles sont ici nombreuses. Certaines sont nommées, d'autres non. Elles se caractérisent principalement par :
 - leur fonction, qui inclut ou non le rattachement à un organisme (ex : agent du STIF, agent des transporteurs, membre du conseil d'administration, élu, personnel d'accueil).
 - leur activité sur la photo (ex : voyageur).
 - leur aspect physique (ex : usager en fauteuil roulant, femme enceinte).

Non complètement exclusifs les uns des autres, ces regroupements représentent des facettes de point de vue.

- Personnes morales : elles sont répertoriées pour les cas où elles suffisent à qualifier les personnes physiques qui les représentent et pour ceux où elles revêtent une importance contextuelle (ex : signature de contrat, visite de locaux de partenaires).

3.3.2.1.2 Quoi ? (facette Entités > Entités inanimées)

3.3.2.1.2.1 Les objets observés

- Des véhicules : automobiles, deux-roues, train, RER, métro, tramway, bus, catamaran, etc. Dans de rares cas, des éléments des véhicules sont représentés (ex : bogie,

soufflet). Notons que la distinction entre prise de vue intérieure et extérieure doit être marquée car certaines recherches l'incluent. Elle a été rattachée aux aspects techniques de l'image.

- Des infrastructures et des équipements : aménagements urbains, bâtiments, chantiers, différents types d'équipements (de circulation, de service, de sécurité ou de confort), ouvrages d'art (tunnels, ponts, passages souterrains, etc.), points d'arrêt (gares, stations de bus, de métro ou de tramway) et voirie. Ces infrastructures et équipements appartiennent généralement au transport mais pas systématiquement, notamment quand le reportage porte sur la ville en général.
- Des objets courants sans rapport avec le monde du transport ou de l'urbanisme, tels que des aliments ou des parapheurs, y compris en gros plan.

3.3.2.1.2.2 Les modes de transport

A un niveau moins directement matériel, les modes de transport sont un aspect essentiel du secteur. Quel rapport un mode de transport entretient-il avec les objets listés au point précédent ?

A priori, un mode est déterminé par les véhicules et les infrastructures qu'il utilise. Il faut toutefois nuancer ce constat :

- Les véhicules. En général, les différents types de véhicules s'identifient au premier regard et sont désignés par le même terme que le mode de transport auquel ils se rapportent (bus, métro, automobile...). Ce n'est pourtant pas toujours le cas. Le tram-train, par exemple, est un véhicule d'aspect indéterminé, capable de circuler à la fois sur les voies de tramway et de train, mais il se rattache uniquement au mode tramway et en l'occurrence à la ligne 4 (T4). Par ailleurs, certains modèles de trains se rapportent à la fois au mode train et au mode RER, comme le modèle Z 20500, une rame automotrice qui circule à la fois sur les lignes H, J, K, L, P et R du réseau Transilien et sur les lignes C et D du RER. Il n'est donc pas possible de lier systématiquement l'aspect d'un véhicule ou un modèle donné à un mode précis. La concordance peut néanmoins s'établir au niveau de la famille de modes : par exemple, au niveau Train/RER pour les automotrices concernées par cette double affectation ou au niveau Transport ferroviaire pour le tram-train.
- Les infrastructures de transport. Certaines sont spécialisées (ex : une voie de tramway) mais d'autres concernent plusieurs modes, y compris certaines qui semblent dédiées à un seul, comme les voies ferrées. Une gare de banlieue, par exemple, peut être un point d'arrêt à la fois pour les trains, les RER et les tram-trains. De la même façon, une route peut servir aux automobiles, aux motos, aux bus, etc. Il est donc également difficile

d'établir un rapport strict entre des infrastructures de transport et un mode précis, bien que là encore le lien puisse se faire avec une famille de modes.

Précisons pour finir que les lignes, quant à elles, correspondent toujours à un mode.

En résumé, l'association d'un mode à des véhicules est globalement possible tandis que celle d'un mode à des infrastructures de transport est plus lâche. D'ailleurs, tous les équipements et infrastructures représentés ici ne se rapportent pas au transport.

3.3.2.1.2.3 Transport de personnes et transport de marchandises

Enfin, en dehors de la question du mode, le transport se définit par ce qui est transporté : des personnes ou des marchandises. Le transport de marchandises figure de façon annexe dans le fonds mais il faut lui réserver une place.

L'appellation Offre de transport représente une solution. Elle permet en effet d'englober les deux points de vue : transports par mode et transports par entités transportées. Le premier point de vue domine car la notion de transport de personnes est presque toujours sous-entendue.

3.3.2.1.2.4 Organisation des termes à l'intérieur de la facette Entités inanimées

A présent que les liens entre les différents concepts sont posés, l'organisation des termes à l'intérieur de la facette peut être examinée.

Dans le thésaurus Motbis, les modes de transports appartiennent au champ sémantique Transport tandis qu'il existe un champ sémantique dédié, Véhicule, pour le matériel éponyme.

Toutefois, bien que leur association demande de la prudence, modes et véhicules entretiennent des liens forts. Dans le thésaurus de l'INA, des facettes transversales à plusieurs champs sont mêlées à des facettes propres à un champ sémantique. Par exemple, Infrastructure et Matériel apparaissent à la fois dans Sport et dans Audiovisuel, tandis qu'Épreuve et Sport dérivé n'appartiennent qu'au champ sémantique du Sport. Sur ce modèle, Véhicules et Lignes permettraient ici de répartir sous chaque mode ou famille de modes les descripteurs des lignes d'une part et ceux des véhicules d'autre part.

Les infrastructures, quant à elles, sont réparties dans le thésaurus Motbis sous Infrastructure de transport, qui constitue un champ sémantique propre, et Infrastructure urbaine, qui appartient au champ sémantique Urbanisme. Les autres termes utiles à la description des infrastructures et équipements du fonds du STIF sont disséminés dans leurs champs sémantiques respectifs : Ascenseur, par exemple, appartient à Matériel et machine, Escalier

(à la place duquel il faut employer Élément du bâtiment) à Technique professionnelle, via le terme générique BTP, et Pictogramme à Sémiologie.

Pour une liste de termes réduite, il est donc plus simple d'aborder ces objets sous l'angle de leur catégorie fondamentale propre, celle d'Entités inanimées et, plus finement, d'Infrastructures et équipements. Dans le thésaurus de l'INA, d'ailleurs, les entrées Matériel et Infrastructure jouent bien le rôle de facettes. La jonction entre la facette Infrastructure et le thème Audiovisuel, par exemple, introduit les termes Auditorium et Discothèque.

3.3.2.1.2.5 En résumé

Sous la facette fondamentale Entités inanimées, qui répond à la question Quoi ?, se trouvent deux grandes catégories de concepts :

- L'Offre de transport, qui englobe :
 - Les Modes de transport, sous lesquels Lignes et Véhicules peuvent servir à distribuer les termes spécifiques.
 - La distinction entre Transport de marchandises et Transport de personnes.
- Les Infrastructures et équipements. Il est plus facile de les placer à part en raison de leurs liens trop lâches avec les différents Modes et même avec le transport.

3.3.2.1.3 Où ? (facette Espace)

Comme les modes, les lieux sont déduits de l'observation des objets ou éventuellement des personnes représentées.

A un premier niveau, les lieux ne sont pas des entités nommées mais des gares, des salles de conférence, des bureaux, des rues, etc. Ce sont en fait des infrastructures. A ce titre, ils sont joints à cette catégorie.

A un second niveau, les lieux sont nommés individuellement et font l'objet d'une localisation géographique plus ou moins précise, par exemple Place de la Bastille, Noisy-le-Sec ou Boulevard des Maréchaux. Il est important de faire entrer ces noms propres dans une hiérarchie, par exemple Ile-de-France > Paris > Place de la Bastille, afin d'autoriser des regroupements. Ici, une règle d'écriture a simplement été définie pour créer les lieux au fur et à mesure. Ceux-ci ne font donc pas réellement partie du lexique pour l'instant.

3.3.2.1.4 Quand ? (facette Temps)

A la question Quand ? répondent la date de la prise de vue et, le cas échéant, de l'événement (qui lui peut s'étaler sur plusieurs jours). Ces données, non prédéfinies, ne font pas partie du lexique : elles sont renseignées au fur et à mesure dans les champs appropriés.

Précisons que les mentions de périodes, comme les saisons ou les moments de la journée, sont peu utiles ici. Seule la nuit peut faire l'objet de recherches spécifiques. De ce fait, elle a été intégrée à la catégorie des aspects techniques, sous l'appellation Vue de nuit.

3.3.2.1.5 *Comment ? (facette Processus)*

Cette question porte sur les activités observées et l'action dans laquelle elles s'inscrivent ([6], KATTNIG, p. 76).

Dans le cas étudié, les activités sont surtout intéressantes quand l'action correspondante est un événement. Elles en constituent alors des phases : discours, interview ou cocktail sont autant de moments d'une inauguration, d'une conférence de presse ou d'un séminaire d'entreprise. Ces événements, à leur tour, peuvent représenter les étapes d'un projet, tel que le prolongement ou la création d'une ligne, la modernisation du matériel roulant, etc. Trois niveaux de réponse pertinents à la question Comment ? existent pour le fonds du STIF :

- Phase d'événement
- Événement
- Projet

Précisons qu'au projet il faut adjoindre la notion d'axe de réflexion, plus abstraite. Les termes qui la traduisent sont moins nombreux (ex : accessibilité, qualité de service). Ils n'expriment pas en eux-mêmes des Processus mais des Etats. Les inclure dans les Processus constitue une entorse au principe de la facette fondamentale, selon lequel les concepts appartiennent à une facette en fonction de leur sémantique propre, non de leur rôle dans un domaine. Néanmoins, dans le cas présent, les notions d'amélioration ou de maintenance sont toujours sous-entendues (ex : amélioration de l'accessibilité, amélioration de la qualité de service) car en elles réside la mission du STIF.

La qualification d'autres activités présente peu d'intérêt :

- Les activités métiers. Actuellement, le fonds n'a pas pour objectif d'illustrer les métiers du transport. De ce fait, la description des éléments ou personnes associés suffit. Par exemple, il est plus pertinent d'indexer une photo d'un agent en train de conduire un métro avec les mots Métro et Conducteur qu'avec Conduite.
- Les activités qui ne se rattachent ni à un métier ni à un événement. Ce ne sont pas des critères de recherche pertinents pour des photos d'entreprise (ex : rire, lire, parler, manger, etc.).

3.3.2.2 Tableau récapitulatif

QUESTION	FACETTE	SECTION DU LEXIQUE
Qui ?	Entités > Entités animées > Humains	Personnes
Quoi ?	Entités > Entités inanimées	Offre de transport Infrastructure - équipement
Où ?	Espace	Lieu de prise de vue
Quand ?	Temps	Date de prise de vue
Comment ?	Processus	Événement - projet

Tab. 1 Facettes / Sections du lexique

3.3.2.2.1 La facette Etat

La facette Etat ne trouve pas de place ici. Il a été vu qu'elle ne correspond pas à part entière à l'une des cinq questions : selon que les états sont abstraits ou concrets, ils peuvent être liés à la connotation pour les premiers ou aux Entités, en tant qu'attributs, pour les seconds (voir p. 44).

Dans le cas présent, il n'y a pas de connotation et les propriétés désignées par les états concrets (ex : rugosité, chaleur, etc.) restent secondaires. Les objets représentés sont suffisamment définis par leur identité (modèle de véhicule, route, ascenseur, etc.), sans qu'il soit besoin de décrire leurs propriétés. Il ne s'agit pas, en effet, d'une photothèque technique ou artistique. Elle n'est pas non plus d'un volume si important qu'il faille créer, par exemple, des regroupements par couleur pour les véhicules.

3.3.2.3 Aspects techniques

Enfin, les aspects techniques doivent être pris en compte. Beaucoup sont fournis automatiquement par l'appareil, sous forme de données EXIF (voir p. 24). Il est intéressant d'indiquer d'autres caractéristiques (ex : Portrait, Vue aérienne, Vue de nuit), sans trop entrer dans le détail. Le choix entre Vue intérieure et Vue extérieure doit également être ajouté car il intervient dans les recherches de photos de véhicules.

Ce regroupement empirique de termes n'est pas associé à une seule facette. Il contient par exemple des Etats (ex : Vue aérienne) ou des termes relatifs au Temps (ex : Vue de nuit). Il serait bien sûr possible de créer des subdivisions mais le nombre réduit de termes, une dizaine, rend cette démarche superflue.

3.3.2.4 Difficultés rencontrées

Au-delà de la question des facettes, les difficultés rencontrées dans la construction du thésaurus sont les suivantes :

- Concilier justesse des termes et vocabulaire courant et trouver le bon niveau de détail. Certains utilisateurs sont plus spécialisés que d'autres et l'indexation ne doit pas prendre trop de temps aux personnes à qui elle échoit. Les décisions ont été prises au cas par cas, sans valoriser une approche par rapport à l'autre.
- Faire coïncider normalisation du thésaurus et simplicité d'utilisation. Les relations hiérarchiques ont ainsi parfois posé problème. Les lignes A, B, C, D et E, par exemple, ont été placées sous le mode RER. Or il n'est théoriquement pas possible d'établir une relation hiérarchique entre ces termes. En effet, les lignes ne sont pas une sorte de RER, ce qui exclut une relation générique. Une relation partitive n'est pas autorisée non plus car les normes la réservent aux cas suivants : systèmes et organes du corps, localisation géographique, domaines de la connaissance et structures sociales hiérarchiques ([25], ISO, p. 6-7). Toutefois, dans la mesure où le lexique du STIF n'est utilisé qu'en interne, sa normalisation, qui garantirait son interopérabilité, importe sans doute moins que sa praticité immédiate.

3.3.2.5 Thèmes

Pour se conformer au fonctionnement des interfaces, il était nécessaire d'offrir comme alternative à la recherche par mot clef une consultation guidée par thèmes. En raison de la taille modeste du fonds, ceux-ci gagnaient à être larges et peu subdivisés. Rappelons qu'en l'absence de navigation à facettes, les différents thèmes ne peuvent être croisés mais que le système permet d'attribuer plusieurs thèmes à une même photo. Par ailleurs, il a été établi avec le prestataire que ces grands thèmes seraient des cadres souples, susceptibles d'évoluer avec les nouveaux reportages.

Au vu des images et des demandes, les catégories suivantes ont été constituées :

THEMES	Commentaires
<ul style="list-style-type: none">• Événement<ul style="list-style-type: none">• Événement interne• Événement externe	Cette section reprend uniquement les événements de la partie Événement-Projet, qui répond à la question Comment ? (facette Processus)
<ul style="list-style-type: none">• Personnes<ul style="list-style-type: none">• Direction STIF• Agents STIF	Cette section reprend uniquement les personnes physiques de la question Qui ? (facette Entités > Entités animées >

<ul style="list-style-type: none"> • Elus • Partenaires • Voyageurs 	Humains)
<ul style="list-style-type: none"> • Transport <ul style="list-style-type: none"> • Billettique • Bus • Métro • Modernisation / innovation • Mode actif • RER • Services <ul style="list-style-type: none"> • Accessibilité • Information voyageurs • Sécurité • Train • Transport de marchandises • Transport individuel motorisé • Voguéo • TZEN • Ville 	<p>Ces thèmes reprennent à la fois :</p> <ul style="list-style-type: none"> • des Projets, qui répondent à la question Comment ? (facette Processus) • des Modes et des Infrastructures – équipements, qui répondent à la question Quoi ? (facette Entités > Entités inanimées). <p>Remarque : le transport de marchandises n'est pas un Mode de transport. Pour simplifier et parce qu'il occupe une place annexe dans le fonds, il a été mis sur le même plan que les Modes.</p>

Tab. 2 Thèmes / Sections du lexique

Notons que dans la catégorie Événement, le dernier niveau est celui du reportage. La série est importante pour la photographie journalistique (voir p. 18). Dans l'éventualité d'une future interface de consultation à facettes, notons qu'il s'agirait d'un point d'entrée indépendant et non du résultat d'un recoupement de critères car c'est une unité documentaire à part entière.

Il est à présent possible de voir quelle place les facettes occupent effectivement dans la base.

3.3.3 Mise en oeuvre

3.3.3.1 Intégration du thésaurus au logiciel

La présence de six champs dédiés aux mots clefs dans la version pré-paramétrée du logiciel D permet de présenter le lexique en autant de parties distinctes. Les facettes principales du thésaurus telles qu'elles ont été envisagées sont Offre de transport, Infrastructure - équipement, Événement - projet, Personne et Lieu de prise de vue. Il faut ajouter une section pour les aspects techniques. L'organisation finale a été la suivante :

CHAMP DU LOGICIEL D	SECTION DU LEXIQUE ENVISAGEE
Personnes	La section Personne du lexique est respectée.
Evénement - projet	La section Evénement – projet du lexique est respectée.
Propriétés [de l'image]	La section correspondante du lexique, celles des aspects techniques, est respectée.
Mots clefs	<p>Ce champ reprend les sections :</p> <ul style="list-style-type: none"> • Offre de transport • Infrastructure – équipement <p>En outre, pour améliorer la lisibilité, certains termes ont été mis directement sous Mots clefs : Bâtiment et Point d'arrêt.</p> <p>Pour la même raison, Infrastructure et Equipement ont été séparés.</p> <p>Le terme transversal Chantier, qui peut s'appliquer à différentes infrastructures, est également placé directement sous Mots clefs.</p> <p>Pour des raisons techniques (voir ci-dessous), Véhicule a également été placé directement sous Mots clefs.</p>
Lieu de prise de vue	La section Lieu de prise de vue du lexique est respectée.

Tab. 3 Champs du logiciel D / Sections du lexique

Il n'est pas possible, comme cela avait été prévu, d'utiliser Véhicule sous chaque mode ou famille de modes pour introduire les descripteurs correspondants (voir p. 67). Tout bien considéré, extraire Véhicule d'Offre de transport présente l'avantage de le mettre en valeur. Ainsi, face à une photo de véhicule, l'indexeur perçoit mieux qu'il doit renseigner non seulement le mode mais aussi le modèle du véhicule. Pour compenser cette séparation, des relations d'association ont été établies entre les descripteurs de véhicules et les modes. En revanche les lignes, qui entretiennent un rapport d'exclusivité avec le Mode auquel elles se rapportent, ont été placées sous celui-ci.

Par ailleurs, toutes les Entités inanimées ont finalement été regroupées dans le champ Mots clefs, ce qui, au prix d'une déperdition de sens, permet de conserver un intitulé court. En raison de ce regroupement, le lexique perd un peu de sa capacité à guider l'indexeur.

3.3.3.2 Pérennité d'une structuration à facettes en dehors de la base

3.3.3.2.1 Du point de vue du lexique

La sécurité offerte par l'encapsulation des données IPTC a été évoquée (voir p. 24) : lorsque les photos sont transmises à un utilisateur, ou en cas de migration d'une base à une autre, les informations ne risquent pas d'être perdues. Toutefois, ce standard présente pour l'instant tous les mots clefs au même niveau, sans traduire leur structuration initiale (voir p. 26). Une organisation à facettes du langage documentaire, comme toute autre, se perd donc quand le fichier est transmis seul.

Dans le cas d'une migration d'une base à une autre, en revanche, le paramétrage des tables suffirait à reproduire le schéma.

3.3.3.2.2 Du point de vue d'une éventuelle navigation à facettes

Dans le futur, il se peut que les interfaces de navigation à facettes continuent à se populariser. Une adaptation de la base à ce modèle pourrait alors être jugée pertinente, à plus forte raison si le volume du fonds augmente. Cela paraît d'ailleurs probable dans la mesure où, comme cela a été dit, le développement du numérique et des outils associés favorise une augmentation du nombre de prises de vue (voir p. 20).

Dans cette éventualité, il serait envisageable d'utiliser a minima les facettes de navigation suivantes :

- une adaptation des facettes de premier niveau du lexique (Personne, Offre de transport et Véhicule, Infrastructure - équipement, Événement - projet et Lieu de prise de vue, plus la Date de prise de vue).
- une sélection d'aspects techniques concernant à la fois l'image et son support.
- le nom du photographe et des informations sur les droits associés.

Notons que la définition nord-américaine des facettes, selon laquelle celles-ci s'appliquent également aux objets, serait ici sollicitée pour qualifier les catégories d'informations relatives au support.

Pour les facettes portant sur le sujet, une réutilisation systématique des termes du lexique n'est pas souhaitable, car ceux-ci sont trop précis. Il conviendrait plutôt de créer des regroupements adaptés, basés sur ces termes, de façon à ce que les résultats soient

suffisamment étoffés. Un exemple de facettes de navigation portant sur le contenu est proposé en annexe (voir annexe 2, p. 98). Il repose sur l'idée que les facettes sont des caractéristiques communes à tous les éléments traités. En effet, toutes les photos ont :

- un lieu
- une date de prise de vue
- un sujet matérialisé par des objets ou des personnes
- une occasion

Deux adaptations sont toutefois utiles :

- une même photo pouvant représenter à la fois des personnes et des objets, les deux catégories sont proposées au même niveau pour permettre leur addition. Il est en effet important de pouvoir croiser, par exemple, Voyageur et Bus.
- les modes sont placés sous une facette dédiée (en dépit du fait que toutes les photos ne sont pas associées à un mode) car ce concept est central.

3.4 Conclusion

La Direction de la communication du STIF, autorité organisatrice des transports d'Ile-de-France, voyait son fonds photographique s'accroître, ce qui en alourdissait d'autant la gestion. Ce fonds entièrement et nativement numérique, stocké principalement sur CD-rom, a pour auteurs des photographes professionnels et des agents du STIF. Les thématiques abordées sont bien sûr le transport mais également des domaines annexes tels que l'urbanisme et l'événementiel. Ces clichés sont utilisés comme photos d'actualité ou d'illustration. En dehors du service de la communication lui-même, les utilisateurs sont des agences partenaires, des collectivités ou des transporteurs, et des journalistes.

Dans ce contexte, la mission de stage était centrée sur la mise en place d'un logiciel de photothèque accessible à distance. Il s'est agi de préparer un cahier des charges, de rassembler physiquement les photos et les informations les concernant, de travailler sur les questions de droits et de construire les outils linguistiques.

En théorie, pour le fonds considéré et son utilisation, l'utilisation de facettes paraît plutôt pertinente pour le lexique : les facettes fondamentales semblent convenir à un fonds iconographique car elles recourent les différents aspects du contenu de l'image et les facettes structurelles conviennent à un fonds spécialisé. Elle l'est moins pour la consultation, en raison du nombre relativement réduit de photos.

En pratique, les outils couramment en usage ne sont pas spécifiquement adaptés à une structuration du lexique par facettes. Il n'est toutefois pas impossible de la traduire, du

moins pour des facettes de premier niveau. En revanche, à la consultation, il est difficile d'exploiter aussi bien les éventuelles facettes du lexique que des facettes relatives au document en tant que support car les interfaces ne s'y prêtent pas.

Pour construire les outils linguistiques, une démarche allant du particulier au général a été adoptée, proche de celle préconisée par Vickery pour les classifications spéciales.

Elle a permis de dégager des facettes spécialisées. Celles-ci se sont avérées compatibles avec les différents aspects de la dénotation de l'image, qui correspondent à des facettes fondamentales. Organiser le lexique à l'aide de facettes est donc possible sur le principe et, si l'on se base sur le cas étudié, assez naturel.

Du point de vue de la mise en œuvre concrète, des adaptations à l'outil utilisé ont été nécessaires. Dans l'avenir, une navigation à facettes pourrait être développée, notamment à partir d'une adaptation des facettes du lexique.

Conclusion

Contrairement au texte, l'image ne fait pas usage d'un code pour transmettre l'information. Son fonctionnement est qualifié d'analogique. Cela est particulièrement sensible dans la photographie documentaire, qui semble simplement enregistrer le réel.

Pour placer le traitement documentaire de l'image dans son contexte, il faut par ailleurs insister sur le développement du numérique. Les nouvelles technologies ont fait évoluer le volume, parfois le contenu, les métadonnées et la gestion des photographies.

En termes d'indexation, l'image se caractérise par la multiplicité des questions qu'elle suscite. Le documentaliste s'appuie sur des cadres de lecture pour la décrire : caractéristiques techniques, contenu (dénotation) et signifié (connotation). L'analyse de la dénotation repose elle-même sur cinq questions fondamentales : qui, quoi, où, quand et comment.

Or il semble exister des affinités entre ces axes de description et la théorie des facettes. Initialement développée pour les classifications, dans le monde des bibliothèques, cette théorie a été largement retravaillée et étendue ensuite. Elle est basée sur l'idée que les sujets des documents se composent d'aspects fondamentaux, tels qu'Entité, Action, Etat, Espace et Temps. Les facettes peuvent également être des catégories plus spécialisées, propres à un domaine particulier, sans qu'il y ait contradiction avec les facettes fondamentales. Sur ce double modèle, les facettes suivantes ont été dégagées au STIF :

- Entités : Offre de transport, Véhicule, Infrastructure - équipement et Personnes.
- Action : Événement – projet.
- Espace : Lieu de prise de vue.
- Temps : Date de prise de vue (ou de l'événement).

La facette Etat n'a pas trouvé d'utilité.

La recherche et la consultation des images fixes sont elles aussi caractéristiques. Il est en effet possible de les identifier, du moins sommairement, en les balayant du regard. De ce fait, les résultats peuvent être plus nombreux que pour des documents textuels. Laisser la sélection finale à l'utilisateur à partir d'un groupe d'images pertinent s'avère même particulièrement efficace.

Or des navigations dites à facettes se développent actuellement sur le web. Initialement créées pour les sites marchands, elles se rencontrent désormais également dans le monde documentaire, notamment aux Etats-Unis. Notons à ce propos qu'en Amérique du Nord, une facette peut désigner une caractéristique du document en tant que support, c'est-à-dire en tant qu'objet (par exemple son format, son prix, son poids). Ces interfaces permettent de croiser différentes informations sur les objets traités, à la manière de filtres, tout en gardant

toujours les résultats sous les yeux. L'un de leurs buts est en effet d'éviter le silence. Ce mode de navigation convient bien aux documents visuels, qui gagnent à être recherchés à partir de résultats étoffés. Le récent outil Flamenco est un bon exemple d'application de ce système à l'image.

Il semble donc que l'image présente bien des affinités théoriques avec les facettes, du point de vue de sa description et de sa consultation. Un cas pratique permet d'évaluer concrètement la pertinence de ce rapprochement. La mise en place d'une photothèque d'entreprise spécialisée dans le transport, celle du STIF, a ainsi mené aux constats suivants :

- Pour la constitution du lexique, les aspects de la dénotation et les facettes fondamentales sont facilement mises en rapport (la connotation, non pertinente ici, a été exclue). Cette organisation constitue un guide intégré à l'indexation, ce qui est certainement un avantage pour des utilisateurs non documentalistes. Par ailleurs, il n'y a pas de hiatus entre ces facettes et les facettes spécialisées du secteur étudié qui ont été utilisées. Les outils ne sont toutefois pas particulièrement adaptés à une telle structuration du lexique.
- Pour la consultation, une navigation à facettes ne semble pas très pertinente. Elle alourdirait inutilement l'interface de consultation de ce fonds de volume modeste. Par ailleurs, le fonctionnement des outils en usage ne s'y prête pas.

Au regard des évolutions actuelles, il paraît possible que bon nombre de photothèques adoptent ce modèle de navigation dans le futur. Dans ce cas, il serait intéressant d'exploiter comme facettes non seulement les données relatives à l'image en tant que support mais aussi les différentes unités de sens qui composent sa description. Distinguer ces unités dans le langage documentaire même permet de guider l'indexation (une fois toutes les unités examinées, comme autant de facettes d'un prisme, l'indexation est complète) et fournit une base de travail pour la constitution de facettes de navigation.

Bibliographie thématique

Traitement de l'image

[1] BARTHES Roland. Rhétorique de l'image. Communication, n° 4, 1964, p. 40-51. ISSN 1189-3788

Cet article s'intéresse au sens de l'image et, en l'occurrence, à celui d'une photographie publicitaire. L'auteur met en exergue les caractéristiques de la photographie par rapport aux documents textuels, notamment son lien particulier avec le réel.

[2] BERINSTEIN Paula. Do you see what I see ? Image indexing for the rest of us. Online [en ligne]. [consulté le 20 juin 2010] , mars-avril 1999, p. 85-88. Accès payant,

<https://cassiope.cnam.fr/http/web.ebscohost.com/ehost/pdfviewer/pdfviewer?vid=4&hid=109&sid=cc0bedfd-163b-41f1-a841-7f0c7a50fea9%40sessionmgr113>. ISSN 0146-5422

Cet article américain présente, de façon très pragmatique, les particularités de l'indexation de l'image. L'auteur reprend et commente notamment certaines propositions de Sara Shatford Layne relatives aux facettes.

[3] BOISQUILLON Annie. La Production de la photographie : aspects techniques. In MARTINEZ, Ruth. Des clichés et des clics. Paris , ADBS Editions, 2007. p. 9-22, ISBN 978-2-84365-094-9

Comme l'indique son titre, cette contribution présente un récapitulatif des aspects les plus techniques de la photographie numérique (les formats et leurs caractéristiques, la résolution, la couleur, etc.). Cela permet de mieux comprendre des termes souvent rencontrés mais dont le sens exact peut échapper.

[4] BRUCKMANN D. Le catalogage de l'image fixe. In Le Traitement documentaire de l'image fixe. Paris, Bibliothèque publique d'information, Dossier technique n° 3, [1987]. p. 4-9

L'auteur recense ici les différents types d'images fixes et leurs utilisateurs. Face à leur variété, il souligne la difficulté d'appliquer à toutes un même traitement documentaire.

[5] JORGENSEN Corinne, JAIMES Alejandro, BENITEZ Ana B., CHANG Shih-Fu. A conceptual framework and empirical research for classifying visual descriptors. Journal of the American society for information science and technology [en ligne]. [consulté le 20 juin 2010], 2001, vol. 52, n° 11, p. 938-947. Accès payant,

<https://cassiope.cnam.fr/http/www3.interscience.wiley.com/cgi-bin/fulltext/85008946/PDFSTART>. ISSN 1532-2882

Cet article rend compte d'un test destiné à évaluer la pertinence d'un schéma universel d'indexation de l'image, sous forme de pyramide. Son existence montre que des recherches sont menées dans le sens d'une normalisation des pratiques. En outre, ce modèle systématise une distinction intéressante entre valeur spécifique et générique des objets représentés sur les images.

[6] KATTNIG Cécile. Gestion et diffusion d'un fonds d'image. Paris, ABDS, Nathan, 2002. 127 p. 128, Information documentation . ISBN 2-09-191150-X

Cet ouvrage pratique permet d'identifier clairement les étapes de la gestion documentaire de l'image. Il comporte notamment un chapitre sur son traitement intellectuel.

[7] KATTNIG Cécile. Indexation et systèmes complexes d'indexation. In Les Imagies 2007 : "De l'image au document" Traitements documentaires de l'image et enseignement [vidéo], Montpellier, CRDP Académie de Montpellier, 25 octobre 2007, [consulté le 15 juillet 2010]. Accès libre, <http://www.canalc2.tv/video.asp?idVideo=6908&voir=oui>

Cette intervention de Cécile Kattnig donne une vision complète de l'indexation de l'image aujourd'hui.

[8] KATTNIG Cécile. L'Identification des photographies . In MARTINEZ, Ruth. Des clichés et des clics. Paris , ADBS Editions, 2007. p. 23-32, ISBN 978-2-84365-094-9

Face à la place croissante de l'image dans les pratiques professionnelles, cet article souligne l'importance de la fiabilité de son identification et de sa description. L'auteur fait un point sur les standards actuels de métadonnées, revient sur les modalités d'indexation et finit par quelques conseils.

[9] MELOT Michel. L'image n'est plus ce qu'elle était. Documentaliste-Sciences de l'information [en ligne]. [consulté le 25 juin 10] , 2005, vol 42, n° 6, p. 361-365. Accès payant,

http://www.cairn.info/article.php?ID_REVUE=DOCSI&ID_NUMPUBLIE=DOCSI_426&ID_ARTICLE=DOCSI_426_0361. ISSN 0012-4508

Cet article analyse les rapports entre texte et image. L'auteur plaide pour une indexation basée uniquement sur les textes d'accompagnement car les mots sont impuissants à traduire l'image. Certains aspects de l'évolution de l'image sont abordés.

[10] PECCATTE Patrick. Description des images et rhétorique documentaire . In Gunthert, André. Culture visuelle [en ligne], Paris, Laboratoire d'histoire visuelle contemporaine, mis en ligne le 30 juin 2010, [consulté le 10 juillet 2010]. Accès libre, <http://culturevisuelle.org/dejavu/255>

Cet article s'intéresse aux légendes des photographies et à leur importance, par rapport aux mots clefs, dans le monde de la presse.

[11] PECCATTE Patrick. Heuristique et sérendipité : un exemple en images. In Du bruit au signal (et inversement) [en ligne], Villemoisson-sur-Orge, mis en ligne le 18 janvier 2010, [consulté le 20 juin 2010]. Accès libre, <http://blog.tuquoque.com/post/2010/01/18/Heuristique-serendipite>

Ce billet propose une méthodologie de recherche d'images. Il prend appui sur les travaux menés par l'auteur dans le cadre d'un projet de redocumentarisation de photos historiques. A cette occasion, des remarques sont faites sur l'indexation existante des photos étudiées. Elles mettent en lumière l'importance des informations textuelles liées à l'image.

[12] PECCATTE Patrick. Métadonnées: une initiation - Dublin Core, IPTC, Exif, RDF, XMP, etc.. In Philosophie, informatique, mathématiques [en ligne], Villemoisson-sur-Orge, publié en août 2002, mise à jour le 13 décembre 2007, [consulté le 25 juin 2010]. Accès libre, <http://peccatte.karefil.com/Software/Metadata.htm>

Cet article fait le point sur les standards de métadonnées appliqués, entre autres, à l'image.

[13] PECCATTE Patrick. Quelques remarques sur l'indexation des photos. In Du bruit au signal (et inversement) [en ligne], Villemoisson-sur-Orge, mis en ligne le 28 avril 2009, [consulté le 20 juin 2010]. Accès libre, <http://blog.tuquoque.com/post/2009/04/28/Quelques-remarques-sur-l-indexation-des-photos>

Patrick Peccatte étudie ici les mots clefs d'un échantillon de photos d'agences tels qu'ils sont rendus par le standard IPTC / IIM : la structuration du lexique initialement utilisé y est perdue. Le résultat est donc souvent peu satisfaisant. L'auteur conclut par un point sur les évolutions apportées par les nouveaux standards et en souligne les limites.

[14] RAULIN Antoine. Image numérique et nouvelles technologies. In MARTINEZ, Ruth. Des clichés et des clics. Paris , ADBS Editions, 2007. p. 43-56, ISBN 978-2-84365-094-9

Cette contribution examine les changements apportés à la chaîne documentaire de l'image par l'arrivée du numérique et des technologies associées.

[15] RAULIN Antoine, CACALY Serge. La diffusion des photographies. In MARTINEZ, Ruth. Des clichés et des clics. Paris , ADBS Editions, 2007. p. 57-76, ISBN 978-2-84365-094-9

Les nouveaux usages liés à l'arrivée de la photo numérique (notamment la multiplication des prises de vue) sont exposés ici. L'auteur présente un récapitulatif des différents types d'images numériques en entreprise.

[16] RICHARD P. L'indexation de l'image fixe. In Le Traitement documentaire de l'image fixe. Paris, Bibliothèque publique d'information, Dossier technique n° 3, [1987]. p. 10-20

Cet article pose, entre autres, la question de l'unité documentaire dans le monde de l'image.

[17] SHATFORD LAYNE Sara. Some issues in the indexing of images. Journal of the American society for information science [en ligne]. [consulté le 15 juin 2010], Septembre 1994, Vol. 45, n° 8, p. 583-588. Accès payant, <https://cassiope.cnam.fr/http/www3.interscience.wiley.com/cgi-bin/fulltext/10050078/PDFSTART>. ISSN 0002-8231

Cette auteure américaine se penche sur les particularités de l'indexation de l'image. Elle souligne notamment sa mémorisation rapide, qui rend pertinente la consultation de groupes d'images. Les critères de constitution de ces groupes font l'objet d'une réflexion.

[18] VEZINA Kumiko. Survol du monde de l'indexation des images. Coursus [en ligne]. [consulté le 20 juin 2010], automne 1998, vol. 4, n° 1. Accès libre, <http://www.ebsi.umontreal.ca/cursus/vol4no1/vezina.htm>. ISSN 1201-7302

Cet article est un panorama des pratiques et recherches dans le domaine de l'indexation des documents visuels.

Langages documentaires

[19] Classification décimale de Dewey. In Wikipédia, l'encyclopédie libre [en ligne], mis à jour le 20 juin 2010, [consulté le 20 août 2010]. Accès libre, http://fr.wikipedia.org/wiki/Classification_d%C3%A9cimale_de_Dewey

Cet article clair fournit un extrait de la classification Dewey, utilisé ici comme illustration du fonctionnement des classifications énumératives.

[20] BAYLE Frédéric. Conception d'un thésaurus iconographique dans le cadre du développement d'une agence de photographies créatives sur Internet : le cas de

l'agence photos VOZ'IMAGE. 2008. 141 p. Mémoire, Sciences et techniques de l'information, INTD, 2008

La première partie de ce mémoire est un panorama des particularités du document image et de son traitement.

[21] BEDEL Thierry. Organisation d'un fonds photographique : la classification à facettes, l'exemple de la photothèque de l'Unicef. 1988. 99 p. Mémoire, Sciences et techniques de l'information, INTD, 1988

Ce travail montre quelles solutions une classification à facettes peut apporter aux problématiques d'un fonds spécialisé. L'évaluation du niveau de description à adopter par rapport au contenu des photos et à leur utilisation est également intéressante.

[22] BNF (BIBLIOTHEQUE NATIONALE DE FRANCE). Guide d'indexation Rameau. In [Site de la BNF] [en ligne], Version 1.2009 Paris, 2009, [consulté le 16 août 2010]. Accès libre,

http://guiderameau.bnf.fr/html/rameau_0077.html#d11e2393

Le guide d'indexation Rameau permet de se familiariser avec le fonctionnement assez complexe de ce langage d'indexation synthétique et d'être ainsi en mesure d'établir des rapprochements avec une classification à facettes.

[23] BOUTIN Eric. La recherche d'information sur internet au prisme de la théorie des facettes. 2008. 255 p. Habilitation à diriger des recherches, Sciences de l'information et de la communication, Université du Sud Toulon-Var, 2008

Ce travail montre que le développement des navigations à facettes sur le web est l'occasion pour les professionnels de l'information-documentation d'étudier la place que peuvent y prendre les théories documentaires du même nom.

[24] HUDON Michèle. Le thésaurus : conception, élaboration, gestion. Québec, Editions ASTED, 1994. 220 p. Clé en main. ISBN 2-921548-14-3

Cet ouvrage à visée pratique donne des clefs pour la construction d'un thésaurus. Sa partie consacrée aux facettes, bien que courte, permet d'avoir une vision claire de leurs utilisations possibles.

[25] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. Information and documentation : thesauri and interoperability with other vocabularies, part 1, thesauri for information retrieval. Extrait de la version préparatoire de la norme ISO/WD 25964-1, transmis par l'INTD aux élèves. Genève, ISO, 6 juin 2008.

Ce texte récent sert ici de référence pour la définition des relations génériques partitives dans les thésaurus.

[26] MANIEZ Jacques. Actualité des langages documentaires : fondements théoriques de la recherche d'information. Paris, ADBS Editions, 2002. 396 p. Sciences de l'information. ISBN 2-84365-060-7

Les chapitres sur les classifications, les thésaurus et les facettes permettent de bien distinguer d'une part les caractéristiques distinctives de ces deux langages documentaires et d'autre part les implications, pour chacun, d'une structuration à facettes.

[27] MANIEZ Jacques. Des classifications aux thésaurus : du bon usage des facettes. Documentaliste - Sciences de l'information, 1999, vol. 36, n° 4-5, p. 249-262. ISSN 0012-4508

Cet article fait la lumière sur les différentes acceptions et applications du mot « facette », si fréquemment rencontré. A la suite de cette analyse, l'auteur formule des définitions et fait des propositions en vue d'un bon usage du terme et du procédé.

[28] MENON Bruno. Les langages documentaires : un panorama, quelques remarques critiques. Documentaliste-Sciences de l'information [en ligne]. [consulté le 25 juin 2010], 2007, vol 44, n° 1, p. 18-28. Accès payant, http://www.cairn.info/article.php?ID_REVUE=DOCSI&ID_NUMPUBLIE=DOCSI_441&ID_ARTICLE=DOCSI_441_0018. ISSN 0012-4508

Cet article revient sur les distinctions fondamentales entre les grands langages documentaires et s'intéresse aux outils émergents tels que les ontologies.

[29] MENON Bruno. L'évolution des langages documentaires. Mai 2010. 77 diapositives. Cours, Sciences et techniques de l'information, INTD

Ce cours, dispensé cette année à l'INTD aux élèves du Titre 1, permet de distinguer clairement les différents types de langages documentaires et leurs applications tout en mettant en lumière leurs évolutions actuelles.

[30] VICKERY Brian C. La classification à facettes : guide pour la construction et l'utilisation de schémas spéciaux. Paule Salvan, traductrice. Paris, Gauthier - Villars, 1963. 65 p. Documentation et information

Cet ouvrage propose une méthode de travail pragmatique d'un grand intérêt pour qui doit traiter un fonds spécialisé. Les facettes n'y sont pas conçues comme des catégories universelles définies a priori mais comme les différents angles d'approche nécessaires à la

description complète d'une classe de sujets donnée. Elles sont le fruit d'une démarche de généralisation des termes rencontrés dans la littérature du secteur.

Navigation à facettes

[31] BROUGHTON Vanda. The need for a faceted classification as the basis of all information retrieval. Aslib Proceedings [en ligne]. [consulté le 20 juin 2010], 2006, vol. 58, n° 1/2, p. 49-72. Accès payant, <https://cassiope.cnam.fr/http/www.emeraldinsight.com/Insight/viewPDF.jsp?contentType=Article&Filename=html/Output/Published/EmeraldFullTextArticle/Pdf/2760580105.pdf> . DOI 10.1108/00012530610648671, ISSN 0001-253X

Cet article écrit par une théoricienne britannique donne une vue d'ensemble de l'utilisation actuelle des facettes sur le web, en particulier sur les sites marchands, et des recherches menées sur le sujet. A cette occasion, des distinctions sont faites entre les conceptions britannique et américaine des facettes.

[32] DENTON William. How to make a faceted classification and put it on the web. In Miskatonic university press [en ligne], Toronto (Canada), mis en ligne en novembre 2003, mis à jour en février 2007, [consulté le 15 juin 2010]. Accès libre, <http://www.miskatonic.org/library/facet-web-howto.html>

Cet article est publié sur le site personnel d'un cyberdocumentaliste canadien. A visée purement pratique, il s'adresse aussi bien aux webmasters de sites commerciaux qu'aux professionnels de l'information-documentation. L'auteur reprend les points importants de la théorie des facettes et s'y réfère pour expliquer comment construire de façon orthodoxe une navigation à facettes sur n'importe quel type de site.

[33] LA BARRE Kathryn. Faceted navigation and browsing features in new OPACs: robust support for scholarly information seeking ?. Knowledge Organization, 2007, vol. 34, n° 2, p. 78-90. ISSN 0943-7444

Cet article américain présente une étude à venir, dont le but est d'évaluer la pertinence des nouvelles interfaces des OPAC pour la recherche académique. Il donne des informations sur le développement actuel de ces OPAC.

[34] MAISONNEUVE Marc, TOUITOU Cécile. Une nouvelle famille d'Opac : navigation à facettes et nuages de mots. Bulletin des bibliothèques de France [en ligne]. [consulté le 25 juin 2010], 2007, n° 6, p. 12-19. Accès libre, <http://bbf.enssib.fr/consulter/bbf-2007-06-0012-002>. ISSN 0006-2006

Les auteurs examinent les particularités des nouveaux OPAC, principalement en Amérique du Nord, et établissent des comparaisons entre leurs interfaces. Il est intéressant d'observer les différents types de facettes utilisées dans chacune.

[35] WILSON Travis. Facetmap. [en ligne]. [Californie], 2003, mis à jour en 2009 [consulté le 15 août 2010]. <Accès libre, <http://facetmap.com/>>

Ce site fournit des outils et des conseils pour la mise en place d'une navigation à facettes. La consultation de ce site permet de comprendre comment sont envisagées ces interfaces dans le monde non documentaire et de quels outils les webmasters disposent pour les mettre en œuvre.

[36] YEE Ka-Ping, SWEARINGEN Kirsten, LI Kevin, HEARST Marti. Faceted metadata for image search and browsing. Conference on human factors in computing systems, CHI letters [en ligne]. [consulté le 15 juin 2010], 5-10 avril 2003, vol. 5, n° 1, p. 401-408. Accès libre, <http://bailando.sims.berkeley.edu/papers/flamenco-chi03.pdf>. DOI 10.1145/642611.642681

Cet article décrit les modalités et les résultats d'un test utilisateur mené dans le cadre du projet Flamenco. Cette interface à facettes novatrice y est mise en concurrence avec une interface classique. Les résultats indiquent une préférence pour la première.

Sites internet utilisés en exemple

Banques d'images

[37] CNRS (CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE). La Banque d'images du CNRS. [en ligne]. Paris, [consulté le 10 juin 2010]. <Accès libre, <http://phototheque.cnrs.fr/>>

Les mots clefs de cette photothèque illustrent le résultat d'une indexation post-coordonnée.

[38] HEARST Marti. The Flamenco search interface project. [en ligne]. Berkeley (Californie), Université de Berkeley, mis en ligne en 2006, mis à jour en 2008 [consulté le 15 août 2010]. <Accès libre, <http://flamenco.berkeley.edu/>>

Flamenco est une interface de recherche innovante développée par l'université de Berkeley. Elle est basée sur une navigation à facettes et n'est appliquée pour l'instant qu'à l'image. Des démonstrations sont proposées, notamment à partir du fonds du musée des Beaux-Arts de San Francisco et d'une base photographique de Prix Nobel. Flamenco se caractérise par la possibilité d'explorer les facettes dès la première page, sans nécessairement commencer par une recherche en langage libre.

OPAC

[39] NASHVILLE PUBLIC LIBRARY. [Site de la bibliothèque de Nashville]. [en ligne]. Nashville, 2006, mis à jour en 2010 [consulté le 15 août 2010]. <Accès libre, <http://www.library.nashville.org/>>

Cette bibliothèque dispose d'un OPAC de nouvelle génération, dont l'interface comporte une navigation à facettes. Les facettes portent presque exclusivement sur la description du document en tant que support et non sur son sujet.

[40] QUEENS LIBRARY. [Site de la bibliothèque du Queens, à New-York]. [en ligne]. New-York, [consulté le 15 août 2010]. <Accès libre, <http://www.queenslibrary.org/>>

Cette bibliothèque dispose également d'un OPAC de nouvelle génération proposant une navigation à facettes. Ici, les facettes portent autant sur le document en tant que support que sur son sujet.

[41] PAPY Fabrice. Visual... Catalog [de l'Université d'Artois]. [en ligne]. Lille, 2007, mis à jour le 23 septembre 2009 [consulté le 16 août 2010]. <Accès libre, <http://visualcatalog.univ-artois.fr/vc2/>>

Conçu par une équipe de recherche de Paris VIII, le Visual catalog est également un OPAC de nouvelle génération. La navigation est toutefois originale : elle repose sur les indices de la classification Dewey, mis en regard des vedettes-matières Rameau décoordonnées.

Site marchand

[42] E-BAY.FR. [en ligne]. France, E-Bay inc., 1995, mis à jour en 2010 [consulté le 10 août 2010]. <Accès libre, <http://annonces.ebay.fr/>>

Ce grand site de commerce en ligne propose d'affiner les recherches de produits par facettes. Il est en cela une bonne illustration du fonctionnement actuel de nombreux sites marchands.

Thésaurus

[43] GETTY INSTITUTE. Art and architecture thesaurus : about the AAT. In The Getty [en ligne], Los Angeles, The J. Paul Getty trust, mis à jour le 13 juillet 2010, [consulté le 30 juillet 2010]. Accès libre, http://www.getty.edu/research/conducting_research/vocabularies/aat/about.html

Ce thesaurus américain est organisé à facettes dès le premier niveau. Ce mode de structuration, ainsi que son domaine d'application, en font un exemple intéressant.

[44] INA (INSTITUT NATIONAL DE L'AUDIOVISUEL). Thésaurus [extrait de deux pages transmis par Mme Kattnig le 26 mai 2010].

Ce thesaurus de l'image est construit de la façon suivante : sous chaque champ sémantique, les termes spécifiques sont classés par facettes (personne, infrastructure, action, etc.). C'est une bonne illustration d'un système mixte thème + facette.

[45] SCÉRÉN - CNDP. Thésaurus Motbis. [en ligne]. Edition 2010 (9e édition). Poitiers (Futuroscope), SCÉRÉN - CNDP, 2010 [consulté le 20 août 2010]. <Accès libre, <http://www.thesaurus.motbis.cndp.fr/site/>>

Le thesaurus généraliste de l'Education nationale sert ici de référence pour la dénomination de différents champs sémantiques mentionnés dans les exemples. Bien que cet aspect ne soit pas exploité dans ce travail, notons que, sous chaque thème, les termes spécifiques sont classés sous des facettes « *appropriées au domaine couvert* ».

Arrêt des recherches : mi-août 2010

Annexes

Annexe 1 Thésaurus du STIF

Seules les relations hiérarchiques sont présentées ici.

Événement - Projet

Contexte

Phénomène de pointe

Événement

Conférence de presse
Conseil d'administration
Déménagement
Fête
Inauguration
Lancement de travaux
Mise à l'eau
Remise de prix
Salon
Séminaire
Signature de contrat
Visite

Visite technique
Visite touristique

Vœux

Phase d'événement

Discours
Interview
Repas – cocktail

Projet – concept

Accessibilité
Multimodalité
Stationnement abusif
Communication
Contrat
Contrats de type 2
Création de ligne
Eole
Météor
Intermodalité
Modernisation
Renouvellement
Rénovation
Prolongement de ligne
Qualité de service
Information voyageurs
Information horaire
SIEL
Information tarifaire
Règles d'usage
SIVE
Maintenance
Nettoiemnt

Lieu de prise de vue [Remarque : la liste est complétée au fur et à mesure]

Ile de France

75 - Paris
91 - Essonne
Etampes
Evry

- Palaiseau
- 92 - Hauts-de-Seine
 - Antony
 - Boulogne-Billancourt
 - Nanterre
- 93 - Seine-Saint-Denis
 - Bobigny
 - Le Raincy
 - Saint-Denis
- 94 - Val de Marne
 - Créteil
 - L'Hay-les-Roses
 - Nogent-sur-Marne
- 95 - Val d'Oise
 - Argenteuil
 - Pontoise
 - Saint-Ouen L'Aumône
 - Sarcelles

Mots-clés

Bâtiment

- Bâtiment collectif
- Bâtiment industriel
- Habitation
- Bureaux
- Patrimoine
 - Œuvre d'art

Chantier

Équipement

- Billettique
 - Billetterie automatique
 - Guichet
 - Ligne de péage
 - Valideur
 - Composteur
 - Valideur sans contact
 - Titre de transport
 - Coupon magnétique
 - Passe sans contact
 - Passe Navigo
 - Titre gratuit
- Équipement de circulation
 - Aménagement PMR
 - Ascenseur
 - Bande podotactile
 - Braille
 - Comble-lacune
 - Escalier mécanique
 - Feu sonore
 - Plancher surbaissé
 - Trottoir roulant
 - Bouche de métro
 - Couloir
 - Embarcadère
 - Escalier
 - Feu de circulation routière
 - Bouton d'appel piéton
 - Passage sur voie
 - Passage à niveau
 - Traversée piétonne à niveau
 - Quai
 - Salle d'échanges
 - Signalétique
 - Signal de direction - localisation
 - Plan
 - Plan de ligne
 - Plan du réseau
 - Signal de prescription - indication

- Pictogramme
 - Support d'information
 - Affiche
 - Autocollant
 - Ecran
 - Nouveau media
 - Panneau
 - Panneau numérique
- Equipement de confort
 - Abri voyageur
 - Climatisation
 - Eclairage
 - Poubelle
 - Salle d'attente
 - Sanitaires
 - Siège
 - Strapontin
- Equipement de sécurité
 - Alarme
 - Barrière de sécurité
 - Façade de quai
 - Borne d'appel
 - Sécurité incendie
 - Télesurveillance
- Equipement de services
 - Cabine téléphonique
 - Distributeurs de boissons - confiseries
 - Horloge
 - Kiosque à journaux
 - Parcmètre
 - Point information
 - Borne d'information automatique
 - Stand d'exposition
- Equipement technique
 - Equipement de véhicule
 - Cabine de conduite
 - Compartiment
 - Bogie
 - Pantographe
 - Pneu
 - Roue
 - Soufflet
 - Espace technique
 - Atelier de maintenance
 - Centre d'appel voyageurs
 - Poste de commandement

Infrastructure

- Espace vert
 - Coulée verte
- Ouvrage d'art
 - Château d'eau
 - Passage souterrain
 - Pont
 - Tunnel
- Place
- Voie ferrée
 - Electrification ferroviaire
- Voie navigable
- Voirie
 - Route
 - Rue
 - Stationnement [Infrastructure]
 - Aire de stationnement
 - Parking
 - Parking d'échange
 - Zone réservée
 - Couloir de bus
 - Espace piétonnier
 - Passage piéton
 - Piste cyclable

- Trottoir
- Zone de rencontre
- Zone industrielle
- Offre de transport** [Remarque : la liste des lignes est complétée au fur et à mesure]
- Transport de marchandises
- Transport en commun
 - Offre routière
 - Bus
 - BHNS
 - Réseau bus Optile
 - Phébus
 - Stivo
 - Stivo 30
 - Réseau bus RATP
 - Balabus
 - Bus 26
 - Bus 27
 - Bus 38
 - Bus 91
 - Montmartrobus
 - Trans-Val-de-Marne
 - Mobilien
 - Noctilien
 - N24
 - TCSP
 - TZEN
 - TZEN Sénart-Corbeil
- Offre ferroviaire
 - Métro
 - Funiculaire de Montmartre
 - Ligne 1
 - Ligne 2
 - Ligne 3
 - Ligne 3bis
 - Ligne 4
 - Ligne 5
 - Ligne 6
 - Ligne 7
 - Ligne 7bis
 - Ligne 8
 - Ligne 9
 - Ligne 10
 - Ligne 11
 - Ligne 12
 - Ligne 13
 - Ligne 14
 - RER
 - RER A
 - RER B
 - RER C
 - RER D
 - RER E
 - Train
 - Grande ligne
 - Eurostar
 - ICE
 - Thalys
 - Transilien
 - Ligne H
 - Ligne J
 - Ligne K
 - Ligne L
 - Ligne N
 - Ligne P
 - Ligne R
 - Ligne U
 - Tramway
 - T1
 - T2
 - T3

- T4
- T5
- Navette fluviale
 - Voguéo
- Transport à la demande
 - Filéo
 - PAM
- Transport individuel
 - Automobile
 - Autopartage
 - Deux-roues motorisé
 - Mode actif
 - Marche à pied
 - Patin à roulettes
 - Trottinette
 - Vélo
 - Vélo en libre service
- Taxi
- Transport semi-collectif
 - Covoiturage
 - Ramassage d'entreprise
 - Ramassage scolaire
 - Pédibus

Point d'arrêt [Remarque : la liste est complétée au fur et à mesure]

- Aéroport
 - Paris - Charles de Gaulle
 - Paris - Orly
- Arrêt de bus
- Dépôt de véhicules
 - Dépôt des Joncherolles
- Escale
 - Escale Voguéo
- Gare RER
 - Auber
 - Châtelet - Les Halles
 - Hausmann - Saint-Lazare
 - Magenta
- Gare routière
 - Paris - Gallieni
- Gare [train]
 - Gare d'Austerlitz
 - Gare de l'Est
 - Gare de Lyon
 - Gare du Nord
 - Gare Montparnasse
 - Gare Saint-Lazare
 - Gare de Noisy-le-Sec
 - Gare de Provins
- Pôle d'échanges
- Station de métro
 - Havre-Caumartin
 - La Fourche
 - Les Halles
 - Opéra
- Station de tramway
 - Pont du Garigliano

Véhicule

- Bateau
 - Catamaran
 - Péniche
- Matériel roulant
 - Matériel ferroviaire
 - Rame de métro
 - MF
 - MF 2000
 - MF 67
 - MF 77
 - MF 88
 - MP

MP 59
 MP 73
 MP 89
 Train – RER
 Locomotive
 BB 17000
 BB 27300
 BB 67400
 Matériel tracté
 RIB-RIO
 VB2N
 Rame automotrice
 AGC
 Francilien
 MI 79
 MI 84
 MI2N Altéo
 MI2N Eole
 MS 61
 Z 20500
 Z 20900
 Z 5300
 Z 5600
 Z 6100
 Z 6400
 Z 8800
 Rame de tramway
 Citadis 302
 Citadis 402
 TFS
 Tramtrain
 Avanto S70
 Translohr
 Matériel routier
 Bus [véhicule]
 Gépébus Oréos 55e
 Gruau Microbus
 Heuliez
 Heuliez Gx 117
 Heuliez Gx 127
 Heuliez Gx 317
 Heuliez Gx 77H
 Irisbus
 Irisbus Agora
 Irisbus Agora L
 Irisbus Agora Line
 Irisbus Agora S
 Irisbus CitéLis [Irisbus]
 Irisbus Citalis 18
 Irisbus CitéLis
 Irisbus Citelis Line
 Man
 Man Lion's City G
 Man Lion's City GI
 Man NI 223
 Mercedes
 Mercedes O 530 Citaro
 Mercedes Sprinter Oa 412 Dk
 Renault
 Renault Agora Gaz (Gnv)
 Renault R312
 Scania
 Scania Omnicity articulé
 Scania Omnicity standard
 Trambus
 Irisbus Créalis néo
 Véhicule utilitaire
 Camion
 Camionnette

Prototype
 Véhicule en construction

Personne

Personne morale

Autorité organisatrice des transports

EMTA

STIF

Sytral

Collectivité territoriale

Entreprise

Agence République

Bombardier

Exploitant

Batobus

Optile

Kéolis

Transdev

Véolia

RATP

SNCF

Transpole

Personne physique

Non professionnel

Scolaire – étudiant

Foule

Voyageur

PMR

Femme enceinte

Handicapé

Handicapé moteur

Usager en fauteuil roulant

Handicapé visuel

Personne âgée

Personne avec jeunes enfants

Personne encombrée

Personne nommée

Huchon Jean-Paul

Menant Christophe

Mougard Sophie

Palmerio Jean

Professionnel

Agent

Agent d'entretien

Agent de police

Agent RATP

Agent SNCF

Agent STIF

Commerçant

Dirigeant

Président

Directeur

Administrateur

Elu

Journaliste

Personnel d'accueil

Prêtre

Technicien

Propriétés

Photo de groupe

Portrait

Vue aérienne

Vue de nuit

Vue extérieure

Vue intérieure

Annexe 2 Proposition de facettes de navigation

Lieu

- Ile-de-France
- Province
- Etranger

Année

- 2010
- 2009...

Contexte

- Evénement extérieur
 - Conférence de presse
 - Inauguration
 - Lancement de travaux
 - Remise de prix
 - Salon
- Evénement interne
 - Conseil d'administration
 - Déménagement
 - Fête
 - Séminaire
 - Signature de contrat
 - Visite
 - Visite technique
 - Visite touristique
 - Vœux
- Quotidien

Personne

- Dirigeant STIF
- Agent STIF
- Partenaire
- Personnalité extérieure
- Voyageur / habitant

Objet

- Véhicule
- Point d'arrêt
- Equipement
 - Accessibilité
 - Billettique
 - Information voyageurs

Sécurité
Technique
Locaux (intérieur)
Ville

Mode

Bus
Métro
Mode actif
Mode individuel motorisé
Navette fluviale
Train / RER
Tramway
TZEN

Remarque : toutes les photos ne sont pas associées à un mode. Il serait donc plus exact de laisser le choix entre la sélection d'un mode, qui donnerait accès aux modes répertoriés, ou de l'absence de mode. Il semble toutefois plus simple de laisser cette dernière option sous-entendue.