Madalina Mitran, Guillaume Cabanac, Mohand Boughanem. Indexation de photos géoréférencées à l'aide du web participatif. Dans : INFormatique des Organisations et Systemes d'Information et de Decision (INFORSID 2011), Lille, France, 24/05/11-26/05/11, Association INFORSID, p. 401-415, mai 2011.

Indexation de photos géoréférencées à l'aide du web participatif

Mădălina Mitran — Guillaume Cabanac — Mohand Boughanem

Université de Toulouse, IRIT UMR 5505 CNRS
118 route de Narbonne, F-31062 Toulouse cedex 9

{mitran, cabanac, boughanem}@irit.fr

Catégorie chercheur

RÉSUMÉ. *La démocratisation des appareils photo numériques et l'intégration de capteurs numériques dans les téléphones mobiles permettent à chacun de prendre de nombreuses photos. Or, des plateformes de partage de photos, telles que Panoramio et Flickr, offrent la possibilité de les stocker, de les étiqueter avec des* tags *et de les partager. Ainsi, plus de 4 millions de photos sont publiées sur Flickr chaque mois. Dans cet article, nous nous intéressons au processus d'indexation de ces photos mis en œuvre par un moteur de recherche. Nous proposons d'indexer ces photos sans considérer leur contenu, mais en exploitant conjointement des métadonnées spatiales, temporelles et thématiques liées aux annotations (*tags*) ajoutées par les internautes. Nous définissons également un cadre d'évaluation pour mesurer la qualité des termes d'indexation en les confrontant au résultat d'une indexation manuelle.*

ABSTRACT. *The democratization of digital cameras and the integration of digital sensors in mobile phones allow everyone to take huge amounts of photos. Despite this, photo sharing platforms (e.g., Panoramio and Flickr) allow individuals to store, to tag, and to share photos. Indeed, more than 4 million pictures are published on Flickr every month. In this paper, we tackle the indexing process for these photos prior to retrieval, as supported by a search engine. We propose to index photos without considering their contents. Instead, we jointly consider spatial, temporal, and topical metadata related to social tags contributed by web users. We also define an evaluation framework to assess the effectiveness of the proposed indexing process. This intends to compare the produced indexing terms with those resulting from manual indexing.*

MOTS-CLÉS : *Photos, indexation, métadonnées, tags, localisation, géoréférence*

KEYWORDS: *Photos, indexing, metadata, tags, localization, georeference*

1. Introduction

La démocratisation des appareils photo numériques permet à chacun de prendre de nombreuses photos à faible coût. De plus, l'intégration de capteurs numériques dans les téléphones mobiles — selon The Economist (2010), il y en aurait actuellement plus d'un milliard de par le monde — ne fait qu'accentuer ce phénomène. Par conséquent, nos disques durs regorgent de photos numériques, que nous pouvons également partager sur des plateformes web telles que Flickr[1] ou Panoramio[2].

Or, sur le disque dur d'un ordinateur, les photos peuvent être organisées d'autant de façons qu'il existe d'individus. Par exemple, on peut toutes les classer dans un unique répertoire, ou par thème (plage, montagne, travail, etc.), ou par date, ou par lieu, ou par appareil photo quand on en possède plusieurs, ou par photographe sur un ordinateur familial, etc. Il n'existe vraisemblablement pas d'organisation consensuelle, ce qui complexifie le classement et la recherche ultérieure pour exploiter ces photos.

Avec Flickr et l'apparition des plateformes de partage de photos en ligne, une telle « désorganisation » n'était pas envisageable. Au lieu d'imposer un classement en répertoires et sous-répertoires, Flickr permet aux usagers d'étiqueter leurs photos à l'aide de *tags*, c'est-à-dire un mot ou une expression qu'il choisit librement (Hammond *et al.*, 2005; Macgregor *et al.*, 2006). Par exemple, une photo d'une écluse sur le Canal du Midi pourrait être étiquetée par les cinq tags « `vélo écluse "Canal du Midi" Béziers péniche` ». N'importe quel internaute peut par la suite accéder à cette photo à partir d'un de ses *tags* (par navigation ou requête) qui sont présentés sous la forme de liens hypertextes.

Dans ce contexte, l'organisation et la recherche de photos ont tôt été qualifiées de *« big challenge »* (Tešić, 2005). Face au volume croissant de photos disponibles, à la fois sur les ordinateurs des individus ou sur le web, deux questions se posent. La première : à l'image du mode de recherche proposé par les moteurs de recherche de documents, tels que Google, *comment peut-on offrir un service efficace de recherche de photos à base de requêtes textuelles ?* Cette question en suscite une seconde : *comment pouvons-nous indexer les photos pour offrir un service efficace de recherche ?*

L'article est organisé comme suit. Dans la section 2, nous présentons le contexte de l'indexation de photos et les problématiques associées. En particulier, l'indexation basée sur l'analyse du contenu des photos est difficile à mettre en œuvre et obtient des résultats peu satisfaisants. Dans la section 3, nous présentons notre contribution : un *processus d'indexation de photos géoréférencées*, basé notamment sur des métadonnées fournies par des internautes. Nous définissons dans la section 4 un cadre d'évaluation visant à comparer la qualité des termes d'indexation produits automatiquement avec ceux résultant d'une indexation manuelle. Enfin, nous concluons cet article dans la section 5 où nous détaillons également les perspectives à ce travail.

---

1. `http://flickr.com` est un site généraliste de partage de photos.
2. `http://panoramio.com` est un site spécialisé dans le partage de photos géoréférencées (comprenant une métadonnée de localisation, telles que des coordonnées GPS).

2. Contexte de l'indexation de photos : approches de la littérature

Dans le contexte de la recherche d'information (RI), nous nous intéressons dans cet article à la recherche de photos numériques (soit un type particulier de document électronique). Notre travail se focalise sur l'indexation des photos par leur transformation préalable en documents textuels, pour ensuite permettre une recherche classique (recherche de documents à partir d'une requête). De façon générale, l'indexation d'images et de photos a fait l'objet de travaux que nous présentons dans la section suivante.

2.1. *Indexation de photos : (con)textuelle, manuelle ou basée sur leur contenu*

La littérature présente quatre approches principales concernant l'indexation des photos. Nous les synthétisons ci-dessous.

1) *L'indexation contextuelle* peut s'appuyer sur des données recueillies lors de la création de la photo. Par exemple, Monaghan *et al.* (2006) analysent les signaux Bluetooth émis par les appareils électroniques mobiles pour identifier les membres du réseau social de l'usager présents sur le cliché ou à proximité. De plus, la localisation (par coordonnées GPS) de l'individu qui prend la photo est exploitée dans (Fan *et al.*, 2010) pour extraire une description du lieu de la prise de vue à partir de ressources textuelles. Ces éléments de contexte sont également exploités dans (Viana *et al.*, 2008a; Viana *et al.*, 2008b; Viana *et al.*, 2009) en complément d'autres données (date et heure de la prise de vue, météo, propriétés du dispositif photographique, notamment). Par ailleurs, le contexte d'utilisation peut également être exploité. Par exemple, Egyed-Zsigmond *et al.* (2007) analysent les traces de navigation d'usagers au sein d'une collection de photos pour suggérer des mots-clés en fonction des photos consultées.

2) *L'indexation manuelle* est réalisée par des personnes qui décrivent les photos en les annotant avec des termes, éventuellement issus de vocabulaires contrôlés. Ce travail est notamment réalisé par des documentalistes, en général sur des collections spécifiques et de faible taille, telles que des photos de paysages pour des agences de presse ou de voyage.

3) *L'indexation textuelle* ou *text-based image retrieval* (Gong *et al.*, 2006; Deschacht *et al.*, 2007; Torjmen *et al.*, 2010) est basée sur l'analyse du voisinage textuel de la photo, tel que le titre du document qui la contient ou le paragraphe qui l'entoure dans le document. Cette technique est notamment utilisée par le moteur de recherche Google images[3] : en réponse à une requête constituée de mots-clés, le moteur de recherche restitue les photos contenues dans les pages qu'il juge pertinentes.

4) *L'indexation sur le contenu* ou *content-based image retrieval* est basée sur l'extraction de caractéristiques visuelles : la couleur, la forme ou la texture qui sont extraites de la photo. Ces caractéristiques permettent ensuite de calculer une similarité entre les photos (Halawani *et al.*, 2006). Pour une description détaillée de cette approche, nous renvoyons le lecteur aux travaux de Smeulders *et al.* (2000) présentant une

3. `http://images.google.fr` restitue des images correspondant à la requête d'un utilisateur.

synthèse de 201 références sur ce sujet. Notons que l'identification de personnes présentes sur un cliché a fait l'objet de travaux récents (O'Hare *et al.*, 2009).

La problématique de l'indexation des photos pour une recherche ultérieure est d'actualité. L'analyse de la littérature nous a permis d'identifier les limites des approches que nous avons présentées dans cette section. Nous détaillons ces limites dans la section suivante, ainsi que les nouveaux axes de recherche récemment développés et notre positionnement par rapport à ces travaux.

2.2. *Limites liées aux quatre approches d'indexation de photos*

Les quatre approches de la littérature présentées dans la section 2.1 souffrent de limites que nous discutons ci-dessous.

1) *L'indexation contextuelle* impose des contraintes techniques (émetteurs Bluetooth, notamment) ou d'usage (nécessité de disposer de traces de navigation, par exemple) sur la production et l'utilisation des photos. Or, les millions de photos issues des plateformes de partage ne satisfont pas ces contraintes. Aussi, les techniques d'indexation proposées sont restreintes à des cas particuliers et sont inopérantes dans le cas général.

2) Bien que produisant des termes d'indexation de bonne qualité, la difficulté de mise en œuvre de *l'indexation manuelle* est un frein à son utilisation. Cette tâche est laborieuse, difficile à réaliser pour de grands volumes de photos (des millions sur les plateformes de partage en ligne) et sujette à l'interprétation des annotateurs. Le coût de ce type d'indexation représente une limite supplémentaire à sa généralisation (Layne, 1994). Toutefois, sur des plateformes de partage telles que Panoramio, les individus publient leurs photos et les annotent eux-même à l'aide de *tags* qu'ils choisissent librement. Aussi, des travaux ont étudié la faisabilité d'utiliser ces *tags* à l'instar d'une indexation manuelle. L'exploitation de la production des internautes à des fins d'indexation est connue sous le nom de *crowdsourcing* (Alonso *et al.*, 2008), signifiant « production de la foule ». Rorissa (2010) montre que les *tags* diffèrent des termes qui auraient été choisis par les documentalistes. Cette observation n'est cependant pas surprenante, car deux personnes différentes indexent un objet avec seulement 20 % de termes en commun (Furnas *et al.*, 1987).

3) *L'indexation textuelle* présente deux limites. Premièrement, des photos sont postées sans être pour autant incorporées dans des documents textuels. C'est notamment le cas de plus de quatre millions de photos géoréférencées mises en ligne chaque mois sur Flickr. Bien que potentiellement intéressantes, ces photos ne peuvent pas être indexées avec leur contexte d'apparition (texte englobant) car il n'existe pas. Deuxièmement, la recherche par appariement de chaînes de caractères entre termes entourant les photo et termes de la requête, sans résolution des entités spatiales exprimées dans la requête ni analyse des métadonnées de géoréférencement des photos, peut mener à des résultats non pertinents. Par exemple, la requête « *Chutes du Niagara* » peut restituer des photos de l'*hôtel Niagara*, alors que celui-ci est situé en France.

4) La limite principale de *l'indexation basée contenu* concerne son manque de spécificité. Par exemple, des photos de la plage principale d'Acapulco pourront être indexées par « `mer`, `plage`, `sable`, `soleil` » sans qu'il y ait pour autant moyen de reconnaître automatiquement le paysage d'Acapulco. Ces photos ne pourraient donc pas être restituées à l'utilisateur posant la requête « *plage Acapulco* ». Notons cependant que des solutions ont été proposées à ce problème, notamment en ayant recours à l'indexation automatique des photos en utilisant : les métadonnées de géoréférencement, l'estampille temporelle correspondant à la capture de la photo ainsi que ses caractéristiques visuelles pour faciliter toute recherche ultérieure (Viana *et al.*, 2008a; Viana *et al.*, 2008b; Viana *et al.*, 2009; Lee *et al.*, 2010).

Pour résumer, l'indexation de photos se heurte à plusieurs écueils : grand volume de photos publiées en ligne sans voisinage textuel (texte englobant), coût trop élevé de l'indexation manuelle par des documentalistes, manque de spécificité de l'approche basée sur l'analyse du contenu des photos. Pour contrebalancer ces limites, nous nous appuyons sur des métadonnées additionnelles que les usagers associent aux photos sur les plateformes de partage. Ces aspects, que nous détaillons dans la section suivante, n'ont pas été abordés dans la littérature à notre connaissance.

3. Contribution : indexation des photos étiquetées et géoréférencés

Notre proposition pour l'indexation de photos repose sur l'analyse de deux types de métadonnées liées aux photos :

– inspirés par Rorissa (2010) et selon le principe du *crowdsourcing* (Alonso *et al.*, 2008), nous exploitons les métadonnées de description (*tags*) librement associées aux photos par les différents utilisateurs des plateformes de partage. Nous nous appuyons ici sur la théorie de *l'intelligence collective* de Surowiecki (2005) : face à un problème donné, l'agrégation des solutions d'un *grand nombre* de personnes *indépendantes* aboutit à une solution meilleure que la solution de la personne la plus avisée du groupe. Or, les plateforme de partage offrent des points de vue descriptifs variés au sujet des photos publiées. Un nombre élevé de *tags* sur une photo, provenant de plusieurs individus, permet de recouper les points de vue. La redondance de *tags* observée indique un consensus dans la description de la photo. Ainsi, parmi tous les *tags* associés à la photo d'Acapulco, nous faisons l'hypothèse que les plus descriptifs (tels que `Acapulco`, `plage`, `baie`, `Mexique`) pourront être identifiés car employés par plusieurs utilisateurs ;

– inspirés par les travaux de (Viana *et al.*, 2008a; Viana *et al.*, 2008b; Viana *et al.*, 2009; Lee *et al.*, 2010), nous exploitons les métadonnées de géoréférencement (coordonnées GPS) intégrées aux photos postées sur les plateformes de partage. Elles permettent de désambiguïser le lieu de capture de chaque photo, évitant ainsi le manque de spécificité des approches basées sur le contenu. Ainsi, la localisation du photographe ayant pris une photo de la plage d'Acapulco est désormais connue et exploitable par le moteur de recherche : il était à Acapulco. Nous exploitons aussi l'estampille temporelle de la photo car elle permet d'identifier des *tags* pertinents : ceux qui ont été associés à d'autres photos prises dans la même fenêtre spatio-temporelle.

À notre connaissance, l'exploitation conjointe des métadonnées de géoréférencement, d'estampilles temporelles et de métadonnées de description via le web participatif n'a pas été proposée dans la littérature. Le processus d'indexation que nous définissons dans les sections suivantes est constitué des trois étapes illustrées dans la Figure 1.

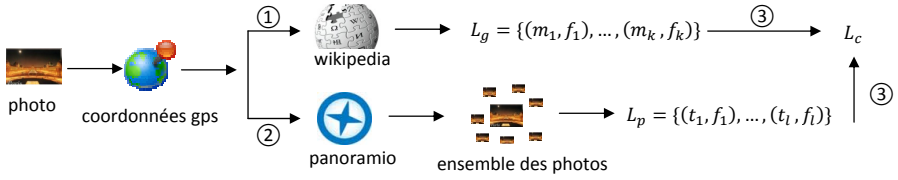


Figure 1 – Indexation en trois étapes d'une photo étiquetée et géoréférencée.

L'étape ① exploite les coordonnées de géoréférencement de la photo. L'étape ② exploite les *tags* et estampilles temporelles associées aux photos sur les plateformes de partage. Enfin, l'étape ③ combine les résultats des étapes précédentes. Nous détaillons ces trois étapes dans les sections suivantes.

3.1. *Étape ① : exploitation des coordonnées de géoréférencement*

À partir d'une photo p, nous extrayons les coordonnées GPS associées (latitude et longitude) à partir de ses métadonnées (Tešić, 2005), notamment intégrées dans le format EXIF (EXchangeable Image File) en en-tête de fichier. Notons que les pages Wikipédia décrivant des lieux contiennent également leurs coordonnées GPS. Par conséquent, nous pouvons identifier les pages Wikipédia dédiées aux lieux les plus proches de p — un seuil δ_1 de distance maximale entre le lieu de la page Wikipédia et p ou un seuil ν_1 limitant le nombre de pages Wikipédia à considérer sont à fixer expérimentalement.

Par la suite, nous mettons en œuvre des techniques classiques de recherche d'information (RI) pour extraire de cette page les termes les plus significatifs. Ce processus décrit en détails dans (Manning *et al.*, 2008, chap. 6) est constitué de quatre phases : segmentation du texte en mots en fonction de délimiteurs (espaces, ponctuations, etc.), élimination des mots-outils qui sont vides de sens (pronoms et déterminants, notamment), normalisation des mots pour obtenir des termes (élimination des marques d'accord et de conjugaison en ayant recours à la lemmatisation, par exemple). Les termes sont ensuite pondérés pour renforcer ceux qui apparaissent beaucoup dans la page Wikipédia traitée, tout en étant peu employés globalement dans les autres pages ; les termes possédant ces deux caractéristiques sont discriminants de la page traitée. Cette phase s'appuyant sur des statistiques de fréquence et de rareté fait référence aux concepts de TF *(term frequency)* et d'IDF *(inverse document frequency)* définis par Spärck Jones (1972). Enfin, nous introduisons une pondération complémentaire pour favoriser les termes présents en début de la page Wikipédia, zone généralement dédiée au résumé du sujet traité dans la page.

Nous faisons l'hypothèse que les termes extraits de cette façon sont descriptifs de la photo en cours d'indexation. La liste $L_g = [(m_1, f_1), \ldots, (m_k, f_k)]$ est ainsi produite pour la photo indexée p, où m_i représente un terme et f_i la fréquence de m_i dans la page Wikipédia indexée, pondérée par la localisation de m_i au sein de la page.

Par exemple, considérons le cas d'une photo prise lors de la cérémonie d'investiture de Barack Obama au parc du *National Mall* de Washington, D.C., le 20 janvier 2009. La liste L_g contiendra les termes suivants, ordonnées par poids décroissant : `washington`, `lincoln`, `mall`, `monument`, `united`, `national`, `states`, `capital`, `american`...

3.2. *Étape ② : exploitation des* tags *et estampilles temporelles*

À partir d'une photo p, nous extrayons ses coordonnées GPS et son estampille temporelle (date et heure de la prise de vue). Ensuite, nous interrogeons une plateforme de partage de photos (par exemple, Panoramio) pour obtenir les photos $P = \{p_1, \ldots, p_q\}$ prises dans la même fenêtre spatio-temporelle que p, ainsi que les *tags* associés à chacune d'entre-elles. Un seuil δ_2 de distance maximale entre p et p_i, un seuil τ limitant la durée écoulée entre la création de p et de p_i ou un seuil ν_2 limitant le nombre de photos à considérer sont à fixer expérimentalement. Notons que les photos et les *tags* ont pu être postés par des personnes différentes, exploitant ainsi un aspect de l'intelligence collective (Surowiecki, 2005).

Nous faisons l'hypothèse que les *tags* extraits de cette façon sont représentatifs du point de vue des personnes qui ont pris des photos quasi au même moment et au même endroit que p. La liste $L_p = [(t_1, f_1), \ldots, (t_l, f_l)]$ est ainsi produite pour la photo indexée p, où t_i représente un *tag* et f_i représente le nombre de fois que t_i a été utilisé pour annoter une photo issue de P.

Dans l'exemple précédent de la photo prise à Washington, la liste L_p contiendrait les *tags* suivants, ordonnées par nombre d'utilisation décroissant : `obama`, `president`, `2009-01-20`, `hope`, `barack`, `yes_we_can`, `US`, `american`, `mall`, `black`, `washington`... Notons l'importance de la dimension temporelle dans cette étape : une photo prise au même endroit le 28 août 1963 pendant le discours *I have a dream* de Martin Luther King aurait été associée avec des *tags* bien différents.

3.3. *Étape ③ : combinaison des résultats liés aux géoréférences et aux tags*

Les deux étapes précédentes aboutissent à deux listes qui nous paraissent complémentaires pour décrire une photo à indexer. Aussi, dans cette troisième étape, nous combinons les résultats précédents en une seule liste.

Au préalable, notons que les valeurs f_i des deux listes appartiennent à des domaines de définition de $\mathbb{R}$ dissemblables. C'est pourquoi nous recourons à une fonction de normalisation $\mathrm{norm} : \mathbb{R}^k \times \mathbb{R} \to [0\,;1]$ pour transformer chaque valeur $v_i \in \mathbb{R}$ d'une liste L de k valeurs vers l'intervalle $[0\,;1]$. Nous pouvons employer, par exemple,

l'instanciation $\mathrm{norm}_{\mathrm{Lee}}$ de norm que Lee (1997) a expérimenté en RI dans le cadre de la combinaison de scores dissemblables de documents issus de moteurs de recherche :

$$\mathrm{norm}_{\mathrm{Lee}}(L, v) = \frac{v - \min(L)}{\max(L) - \min(L)}. \quad [1]$$

Puis, nous identifions le langage d'indexation $\mathcal{L}$ d'une photo p en réalisant l'union des termes (issus de l'étape ①) et des *tags* (issus de l'étape ②) trouvés pour p. Notons que l'union ensembliste élimine les doublons. Enfin, à chaque terme d'indexation de $\mathcal{L}$ est associé le résultat d'une fonction d'agrégation appliquée aux valeurs f_i (éventuellement pondérées pour moduler l'importance d'une liste par rapport à l'autre) provenant des deux listes. Le but ici est d'attribuer un poids d'autant plus élevé qu'un terme est « fort » dans les deux listes. La moyenne arithmétique est une telle fonction d'agrégation ; elle est employée dans [2], où une valeur f_i est obtenue par la fonction $\mathrm{val}(L, t)$ appliquée à une liste L et un terme t, qui retourne 0 si $t \notin L$ ou bien la valeur f_i associée à t dans L, dans le cas contraire. Avec cette fonction d'agrégation spécifique, nous obtenons la liste :

$$L_c = \left[\left(t_1, \frac{\mathrm{val}(L_g, t_1) + \mathrm{val}(L_p, t_1)}{2} \right), \ldots, \left(t_{|\mathcal{L}|}, \frac{\mathrm{val}(L_g, t_{|\mathcal{L}|}) + \mathrm{val}(L_p, t_{|\mathcal{L}|})}{2} \right) \right]. \quad [2]$$

Nous avons fait l'hypothèse que notre approche d'indexation en trois étapes fournit des termes d'indexation descriptifs et objectifs (notion de consensus dans les *tags* inspirée par la théorie de Surowiecki (2005) sur l'intelligence collective). La qualité de l'indexation repose sur la sélection judicieuse des termes et *tags* considérés. En effet, il faut ignorer les termes de Wikipédia qui ne sont pas descriptifs et également ignorer les *tags* de Panoramio qui sont personnels (tels que `super` et `maman`). Ces derniers sont, par construction, relégués en queue de la liste L_c triée par poids décroissant.

Afin d'évaluer la pertinence de notre approche d'indexation de photos, nous proposons dans la section suivante un cadre d'évaluation permettant de comparer ses résultats avec les résultats d'une indexation manuelle.

4. Proposition d'un cadre d'évaluation de la qualité du processus d'indexation

Nous définissons un cadre d'évaluation pour mesurer la qualité d'un processus d'indexation de photos. Pour une photo donnée, nous mesurons la capacité de ce dernier a fournir 1) des termes descriptifs qui sont 2) classés par spécificité décroissante. À cet effet, nous adaptons le paradigme Cranfield (Cleverdon, 1962; Voorhees, 2002) d'évaluation de la RI, qui permet d'évaluer $\mathcal{M} : D, r \mapsto [d_1, \ldots, d_n]$, un moteur de recherche $\mathcal{M}$ qui restitue, pour une requête r donnée, une liste de documents $d_i \in D$ (issus d'un corpus documentaire D) ordonnés par pertinence décroissante. Dans notre contexte, il s'agit d'évaluer le processus d'indexation $\mathcal{I} : T, p \mapsto [t_1, \ldots, t_n]$ qui

restitue, pour une photo p donnée, une liste de termes d'indexation $t_i \in T$ (issus d'un ensemble T de termes) ordonnés par spécificité décroissante.

Nous présentions en section 4.1 les notions de RI relatives à l'évaluation grâce à des collections des tests. Puis, nous détaillons la constitution d'une collection de test adaptée au cas des photos géoréférencées en section 4.2. Enfin, nous exposons son utilisation pour évaluer la pertinence d'un processus d'indexation de photos en section 4.3.

4.1. *Évaluation à l'aide de collections de test en RI textuelle*

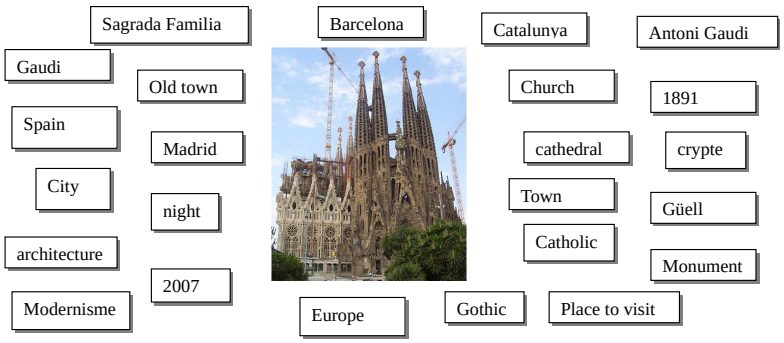
En RI, une collection de test permet d'évaluer des moteurs de recherche (Sanderson, 2010), notamment lors d'une campagne d'évaluation telle que TREC[4] (Voorhees *et al.*, 2005). Elle comprend 1) un corpus documentaire, 2) des besoins en information exprimés sous forme de n requêtes, ainsi que 3) les jugements de pertinence associés : la connaissance des documents pertinents pour chaque requête (les « bonnes réponses »). Un moteur de recherche évalué indexe le corpus au préalable, puis traite les n requêtes pour fournir des listes de réponses (documents ordonnés par pertinence décroissante) qui sont ensuite analysées en fonction des jugements de pertinence. Une mesure statistique compare alors les réponses du moteur de recherche avec les réponses attendues pour produire une valeur décimale représentant la performance qualitative du moteur de recherche. Notons que Buckley *et al.* (2000) montrent qu'il faut au moins $n = 25$ requêtes pour obtenir des conclusions cohérentes du point de vue de la statistique, la valeur $n = 50$ étant retenue à TREC.

4.2. *Collection de test pour évaluer l'indexation de photos géoréférencées*

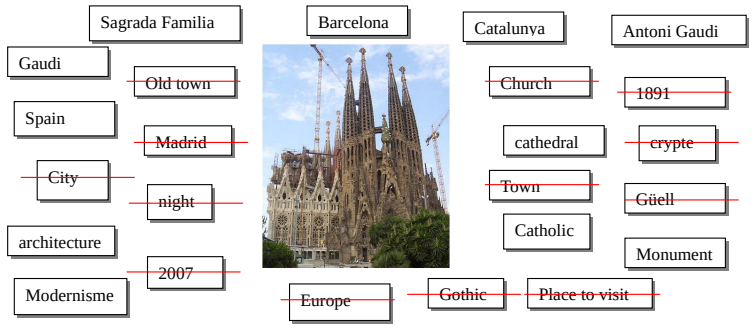
Pour évaluer le processus d'indexation de photos proposé, il n'existe pas à notre connaissance de collection de test appropriée (comprenant photos géoréférencées et *tags*). Aussi, nous proposons d'en constituer une sur le modèle fourni par les travaux de RI (Voorhees, 2002). Cette collection comprend : 1) un ensemble de termes d'indexation candidats, 2) un ensemble de n photos géoréférencées à indexer, ainsi que 3) les jugements de pertinence associés : les termes pertinents pour décrire chaque photo.

En terme de mise en œuvre, nous constituons la collection de test de la manière suivante. Nous rassemblons n photos pour lesquelles nous extrayons l'ensemble T. Ce dernier est composé : a) des termes des pages Wikipédia dont les géoréférences sont les plus proches de la géoréférence de la photo (selon les seuils δ_1 et ν_1 déterminés empiriquement), ainsi que b) des *tags* des photos (publiées sur Panoramio, par exemple) prises à proximité et au même moment (selon les seuils δ_2, τ et ν_2 déterminés empiriquement).

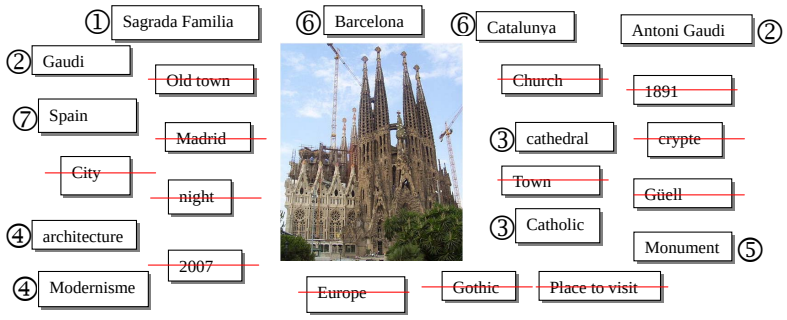
4. Text REtrieval Conference, cf. `http://trec.nist.gov`.



(a) Photo à indexer et termes candidats.



(b) Élimination des termes jugés non descriptifs par l'annotateur.



(c) Classement des termes descriptifs par l'annotateur.

Figure 2 – Constitution des jugements de pertinence pour une photo (a) par identification (b) et classement (c) manuels des termes descriptifs. Exemple de la *Sagrada Família* de Barcelone en Espagne, basilique conçue par Antoni Gaudí.

Par ailleurs, les jugements de pertinence sont graduels : un terme est plus ou moins adapté pour décrire une image de façon spécifique. Nous les recueillons en sollicitant plusieurs assesseurs qui procèdent comme illustré sur la Figure 2, en trois étapes. Premièrement, l'assesseur prend connaissance d'une photo et des termes d'indexation proposés, issus de T (Figure 2a). Deuxièmement, l'assesseur rejette les termes qu'il juge non descriptifs en les barrant (Figure 2b). Troisièmement, l'assesseur ordonne les termes restants de 1 (terme jugé le plus spécifique et descriptif) à $m > 1$. Des termes peuvent être *ex aequo* (Figure 2c). La valeur m est laissée à la discrétion de l'assesseur afin de ne pas le contraindre dans sa tâche.

Inspirés par le *crowdsourcing* (Alonso *et al.*, 2008), nous sollicitons plusieurs annotateurs car cette tâche est subjective. En recoupant les jugements produits, nous souhaitons obtenir des jugements de pertinence plus représentatifs qu'avec un seul assesseur. Le coefficient $\kappa \in [-1\,;1]$ de Fleiss (1971) permet alors de mesurer le degré d'accord entre les annotateurs (Artstein *et al.*, 2008), pour éventuellement réaliser des analyses catégorielles. Nous utilisons la méthode de vote CombMNZ proposée par Fox *et al.* (1993) afin de combiner les jugements individuels en un seul jugement où un terme est d'autant plus pertinent qu'il a été identifié comme tel par un grand nombre d'assesseurs. Notons que la pertinence de cette méthode a été montrée par Lee (1997) dans le cadre de la combinaison de listes de résultats de recherche à TREC.

4.3. *Évaluation de la pertinence du processus d'indexation*

Dans le cadre de la RI, plusieurs mesures statistiques ont été proposées pour évaluer un moteur de recherche (Buckley *et al.*, 2005; Sanderson, 2010). En particulier, lorsque les jugements de pertinence sont graduels (un document étant jugé plus ou moins pertinent pour une requête) la mesure NDCG $\in [0\,;1]$ signifiant *Normalized Discounted Cumulative Gain* (Järvelin *et al.*, 2002) est d'autant plus élevée que les documents les plus pertinents sont restitués en tête de la liste produite par le moteur de recherche. Cette mesure est calculée[5] pour chacune des n requêtes, puis la moyenne de ces valeurs NDCG permet de caractériser la performance qualitative globale du moteur de recherche. La qualité du moteur est d'autant plus avérée que ce score global est élevé. C'est ainsi que plusieurs moteurs de recherche peuvent être comparés en fonction de leurs scores globaux.

Nous appliquons la même démarche dans notre contexte d'évaluation de processus d'indexation de photos géoréférencées. Le processus d'indexation proposé est réalisé pour chacune des n photos. Comme illustré dans le tableau 1, les résultats de chaque liste (L_g, L_c et L_p) permettent de calculer les n valeurs NDGC, qui sont ensuite moyennées pour produire un score global par liste (v_g, v_c et v_p). Nous pouvons alors comparer qualitativement plusieurs processus d'indexation ou leurs variantes. Par exemple, $\frac{v_p}{v_g} - 1$ mesure le pourcentage d'amélioration apportée par la liste L_p en

5. Le programme `trec_eval` peut être utilisé à cet effet. Il est disponible en téléchargement sur le site de TREC : `http://trec.nist.gov/trec_eval`.

Tableau 1 – Comparaison qualitative des listes L_g, L_c et L_p avec NDCG.

Photo	NDCG		
	$ndcg(L_g)$	$ndcg(L_c)$	$ndcg(L_p)$
1	v_g^1	v_p^1	v_c^1
$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	v_g^n	v_p^n	v_c^n
Moyenne	v_g	v_p	v_c

prenant comme référence la performance NDCG de la liste L_g. Enfin, l'amélioration obtenue par une liste comparativement à une autre pourra être validée par des tests statistiques de significativité utilisés en RI à cet effet (Hull, 1993; Sanderson, 2010), tels que le test t pairé bilatéral de Student (1908). En complément, des analyses catégorielles pourront être réalisées, notamment selon le type de photo (extérieur/intérieur, objet/personne, par exemple) ou la popularité du lieu de la prise de vue (densité variable de photos étiquetées à cet endroit).

5. Conclusion et perspectives

La problématique considérée dans cet article concerne l'indexation de photos numériques massivement publiées en ligne, de façon à pouvoir les trouver par des requêtes composées de mots-clés (à l'image des requêtes soumises aux moteurs de recherches). Nous avons proposé un processus d'indexation original basé sur l'exploitation conjointe des métadonnées de géoréférencement, des estampilles temporelles et des métadonnées thématiques (étiquetage social via des *tags*) associées par les internautes sur les plateformes de partage de photos. Puis, dans la perspective de valider qualitativement ce processus, nous avons défini un cadre d'évaluation en établissant un parallèle avec les procédures d'expérimentation appliquées en RI, notamment à TREC (Voorhees *et al.*, 2005).

Notre contribution présente cependant diverses limites. Le processus d'indexation n'est applicable qu'aux photos disposant de métadonnées spatiales, temporelles et thématiques (*tags*). Par ailleurs, il semble adapté pour l'indexation de paysages ou de moments mais est dans l'incapacité d'indexer judicieusement des photos de personnes ou d'objets lorsqu'elles ne sont pas liées à des événements à forte couverture médiatique (cérémonie d'investiture d'un président, par exemple). La mise en œuvre du cadre d'expérimentation proposé est la principale perspective à ce travail. L'évaluation qualitative du processus d'indexation proposé nous permettra d'en juger la pertinence par rapport à une indexation manuelle ou à d'autres processus d'indexation. L'expérimentation offrira également l'opportunité de valider chaque élément du processus et de mesurer les améliorations qui y seront ensuite apportées.

Remerciements

Merci à Jérémie Clos pour son implication dans le développement des prototypes exploités lors de la mise en œuvre du cadre d'évaluation présenté dans cet article.

6. Bibliographie

Alonso O., Rose D. E., Stewart B., "Crowdsourcing for relevance evaluation", *SIGIR Forum*, vol. 42, n° 2, p. 9–15, 2008.

Artstein R., Poesio M., "Inter-Coder Agreement for Computational Linguistics", *Comput. Ling.*, vol. 34, n° 4, p. 555–596, 2008.

Buckley C., Voorhees E. M., "Evaluating Evaluation Measure Stability", *SIGIR'00: Proceedings of the 23rd international ACM SIGIR conference*, ACM, New York, NY, USA, p. 33–40, 2000.

Buckley C., Voorhees E. M., "Retrieval System Evaluation", *in* Voorhees *et al.* (2005), chapitre 3, p. 53–75, 2005.

Cleverdon C. W., Report on the Testing and Analysis of an Investigation Into the Comparative Efficiency of Indexing Systems, ASLIB Cranfield Research Project, Cranfield, UK, 1962.

Deschacht K., Moens M.-F., "Text Analysis for Automatic Image Annotation", *ACL'07: Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, The Association for Computer Linguistics, p. 1000–1007, 2007.

Egyed-Zsigmond E., Iszlai Z., Lajmi S., "PhotoMot, Collaborative Image Management. Interacting with Use Traces", *ICDIM'07: Proceedings of the 2nd International Conference on Digital Information Management*, p. 104–109, 2007.

Fan X., Aker A., Tomko M., Smart P., Sanderson M., Gaizauskas R., "Automatic image captioning from the web for GPS photographs", *MIR'10: Proceedings of the 11th international conference on Multimedia information retrieval*, ACM, New York, NY, USA, p. 445–448, 2010.

Fleiss J. L., "Measuring nominal scale agreement among many raters", *Psychol. Bull.*, vol. 76, n° 5, p. 378–382, 1971.

Fox E. A., Shaw J. A., "Combination of Multiple Searches", *in* D. K. Harman (éd.), *TREC-1: Proceedings of the First Text REtrieval Conference*, NIST, Gaithersburg, MD, USA, p. 243–252, 1993.

Furnas G. W., Landauer T. K., Gomez L. M., Dumais S. T., "The vocabulary problem in human-system communication", *Commun. ACM*, vol. 30, n° 11, p. 964–971, 1987.

Gong Z., U. L. H., Cheang C. W., "Web image indexing by using associated texts", *Knowl. Inf. Syst.*, vol. 10, n° 2, p. 243–264, 2006.

Halawani A., Teynor A., Setia L., Brunner G., Burkhardt H., "Fundamentals and applications of image retrieval: An overview", *Datenbank-Spektrum*, vol. 18, p. 14–23, 2006.

Hammond T., Hannay T., Lund B., Scott J., "Social bookmarking tools (I): A general review", *D-Lib Magazine*, vol. 11, n° 4, p. en ligne, 2005.

Hull D., "Using Statistical Testing in the Evaluation of Retrieval Experiments", *SIGIR'93: Proceedings of the 16th annual international ACM SIGIR conference*, ACM Press, New York, NY, USA, p. 329–338, 1993.

Järvelin K., Kekäläinen J., “Cumulated gain-based evaluation of IR techniques”, *ACM Trans. Inf. Syst.*, vol. 20, n° 4, p. 422–446, 2002.

Layne S. S., “Some issues in the indexing of images”, *J. Am. Soc. Inf. Sci.*, vol. 45, n° 8, p. 583–588, 1994.

Lee J. H., “Analyses of Multiple Evidence Combination”, *SIGIR’97: Proceedings of the 20th annual international ACM SIGIR conference*, ACM Press, New York, NY, USA, p. 267–276, 1997.

Lee Y.-H., Kim B., Kim H.-J., “Photograph Indexing and Retrieval using Combined Geo-information and Visual Features”, *CISIS’10: Proceedings of the 2010 International Conference on Complex, Intelligent and Software Intensive Systems*, IEEE Computer Society, Washington, DC, USA, p. 790–793, 2010.

Macgregor G., McCulloch E., “Collaborative tagging as a knowledge organisation and resource discovery tool”, *Libr. Rev.*, vol. 55, n° 5, p. 291–300, 2006.

Manning C. D., Raghavan P., Schütze H., *Introduction to Information Retrieval*, Cambridge University Press, 2008.

Monaghan F., O’Sullivan D., “Generating Useful Photo Context Metadata for the Semantic Web”, *MDM’06: Proceedings of the 7th International Conference on Mobile Data Management*, IEEE Computer Society, Washington, DC, USA, p. 92–96, 2006.

O’Hare N., Smeaton A. F., “Context-aware person identification in personal photo collections”, *Trans. Multimed.*, vol. 11, n° 2, p. 220–228, 2009.

Rorissa A., “A comparative study of Flickr tags and index terms in a general image collection”, *J. Am. Soc. Inf. Sci. Technol.*, vol. 61, n° 11, p. 2230–2242, 2010.

Sanderson M., “Test Collection Based Evaluation of Information Retrieval Systems”, *Found. Trends Inf. Retr.*, vol. 4, n° 4, p. 247–375, 2010.

Smeulders A. W. M., Worring M., Santini S., Gupta A., Jain R., “Content-based image retrieval at the end of the early years”, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, n° 12, p. 1349–1380, 2000.

Spärck Jones K., “A statistical interpretation of term specificity and its application in retrieval”, *J. Doc.*, vol. 28, n° 1, p. 11–21, 1972.

Student, “The Probable Error of a Mean”, *Biometrika*, vol. 6, n° 1, p. 1–25, 1908.

Surowiecki J., *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations*, Anchor Books, New York, 2005.

Tešić J., “Metadata practices for consumer photos”, *IEEE MultiMedia*, vol. 12, n° 3, p. 86–92, 2005.

The Economist, “Dotty but dashing: Nanotechnology could improve the quality of mobile-phone cameras”, *The Economist*, avril, 2010. `http://www.economist.com/node/15865270`.

Torjmen M., Pinel-Sauvagnat K., Boughanem M., “Using textual and structural context for searching Multimedia Elements”, *Int. J. Bus. Intell. Data Min.*, vol. 5, n° 4, p. 323–352, 2010.

Viana W., Bringel Filho J., Gensel J., Villanova-Oliver M., Martin H., “PhotoMap: from location and time to context-aware photo annotations”, *J. Locat. Based Serv.*, vol. 2, n° 3, p. 211-235, 2008a.

Viana W., Gensel J., Villanova-Oliver M., Martin H., “Indexation sémantique d’images géoréférencées”, *Revue Internationale de Géomatique*, vol. 19, n° 2, p. 169–189, 2009.

Viana W., Hammiche S., Moisuc B., Villanova-Oliver M., Gensel J., Martin H., “Semantic keyword-based retrieval of photos taken with mobile devices”, *MoMM’08: Proceedings of the 6th International Conference on Advances in Mobile Computing and Multimedia*, ACM, New York, NY, USA, p. 192–199, 2008b.

Voorhees E. M., “The Philosophy of Information Retrieval Evaluation”, *in* C. Peters, M. Braschler, J. Gonzalo, M. Kluck (éd.), *CLEF’01: Second Workshop of the Cross-Language Evaluation Forum*, vol. 2406 of *LNCS*, Springer, p. 355–370, 2002.

Voorhees E. M., Harman D. K., *TREC: Experiment and Evaluation in Information Retrieval*, MIT Press, Cambridge, MA, USA, 2005.