

Evaluation

Gilles Godiveau

M2 Pro. médiation culturelle, patrimoine et numérique
de l'université Paris X sous la direction de Mme Bernadette Dufrêne
N° étudiant : 32017459
ID projet : 5b64112c-721b-11e2-a374-00145ea4a2be

«Le web devient audiovisuel»

Séminaire de muséologie numérique
Le projet dans Ligne de Temps

01**Précisions**

Points qui ne seront pas abordés, news et sport. Pour en savoir plus sur ces points voir le site web de la BBC.

02**Rappel**

La BBC diffuse depuis 1922, des services télévisuels depuis 1930 et le web depuis [1994](#). A ce titre le design n'était pas une priorité technologique. Néanmoins il est intéressant de constater que quel que soit le support la mémoire collective reste un des fondements culturels de notre société malgré la qualité de la mise en forme.

03**Support aux programmes**

Dès le début le concept du «support» (site web complémentaire d'une émission télévisuelle) aux programmes est imaginé, jusqu'à aujourd'hui un budget était alloué aux programmes pour créer un site. Beaucoup d'argent pour les sujets connus, très peu ou pas de support pour les niches. Sur 60 services et 1000 à 1500 sujets par jours il était difficile d'apporter un support à tous jusqu'à assez récemment. Néanmoins le fond d'archives de la BBC demeure exceptionnel donc son exploitation, de part son statut de nouveau média mérite une attention particulière. En effet la modularité, l'automatisation, la variabilité ainsi que le transcoding seront au cœur du débat qui va suivre.

04**Support de programme automatisé “slash programs”**

Pour résoudre ce problème d'automatisation, BBC programme se veut un agrégateur d'informations broadcast, base de données d'ayants droits avec un identifiant persistant se voulant permanent. Cela permet d'apporter un minimum d'information pour le support. Les éditeurs peuvent alors s'investir et alimenter ces data. Le système s'externalise et permet d'enrichir plus amplement le support. Exemple est donné sur une série par l'ajout de .rdf sur le chemin faisant apparaître les ontologies structurées. Un schéma est publié pour comprendre ces données avec TV Anytime. Il peut alors décrire l'échange des données purement techniques et la capacité qu'ont désormais les publics de devenir réalisateur après avoir été spectateur.

05**La publication des données**

Les données diffusées tentent d'être le plus inclusives possible en variante de RDF et XML ou Microdata, JSON, XML, etc. Ces données sont destinées aux moteurs de recherche du web, l'avantage pour ces moteurs est de contextualiser fortement le programme par l'ajout de données comme une photo d'un film ou d'autres données à inventer. Principalement utilisé par des développeurs externe, tel Fan Hub qui agrège des Tweets. C'est aussi très utile pour échanger les données dans le système complexe de la BBC.

06**Application à d'autres domaines**

Ces principes énoncés précédemment s'appliquent à d'autres domaines. Le web devient une forme de système de gestion/agrégation de contenu. Par exemple, le page de Tom Waits inclut

des informations provenant de la BBC ainsi que de sources externes (Wikipedia, MusicBrainz, etc...) sans backoffice. Les informations sont donc modifiables directement sur les réseaux sociaux. Cet écosystème s'alimente presque automatiquement, émetteur et récepteur participent à l'élaboration d'un objet qui en produit plusieurs. Un autre avantage est l'instantanéité. Un flux de tweets intégrés annoncera le décès de M. Jackson en temps réel alors qu'un agent mettra peut être longtemps avant faire une MAJ. Principalement mettre en relation des formes narratives et l'exploitation des bases de données. Deux réactions essentielles face au monde.

07

La navigation horizontale

BBC Wildlife Finder est un exemple de navigation horizontale. Il y a beaucoup de domaines différents, pour ne pas rompre le "flow" des données il faut créer des liens entre tous ces différents domaines de manière à ne jamais isoler l'utilisateur dans un domaine particulier. Pour exemple, il existe des clips qui viennent des programmes de la BBC qui sont taggés par l'espèce sur DBpedia.org, avec des liens "same as" sur bbc.co.uk. Il y a donc encore une connexion indirecte entre objets agrégeant des contenus hétérogènes (faune, cuisine, etc...). Contrairement à la navigation verticale, comme le livre, ces nouvelles interfaces culturelles présentent les informations de manière horizontale en permettant un accès à l'information multi-directionnel. La surface d'une page devenant quasiment infinie, la rupture entre hyper-liens est devenue imperceptible et ce dû au fait d'une agrégation au design maîtrisé.

08

La classification des données

Toutes les données doivent être classifiées d'une manière ou une autre en utilisant des sources de données ouvertes avec des URI's comme des catégories. Les classer permet de lier ces différents domaines de les contextualiser et d'agréger ces contenus autour de sujets. Ces objets extraits d'une base de données constituent des ensembles d'informations qui ont la même importance relative. Sans commencement ni fin il importe de façon cruciale de les organiser en séquences thématiques. La base de données et le récit doivent nécessairement se rejoindre sur l'autel des humanités digitales.

09

Comment y intégrer les archives de la BBC ?

La BBC dispose d'une mémoire très étendue et hétérogène là encore. L'enjeu est d'intégrer cette mémoire à la contextualisation des supports aux programmes récents de la chaîne. Contextualiser histoire et culture moderne avec des contenus disparates demande une réflexion d'ordre philosophique et technologique.

10

Des programmes classés manuellement

Jusqu'à maintenant la classification est faite manuellement par un émetteur qui classe ces programmes par espèces, recettes, etc... Les archives de la BBC choisissent des programmes qui peuvent être réutilisés dans des autres programmes, ils utilisent des descriptions très très très fines et détaillées, facilitant sa réutilisation dans d'autres programmes. Une autre source est le segment d'archive, des bouts qui se focalisent sur un programme ou des sujets en particulier. Mais il reste énormément de contenu dans l'archive, comment le prendre en compte ? Le projet de R&D

ABC-IP (automatic broadcast content interlinking) cherche à automatiser ce processus de classification pour désenclaver les archives historiques et les connecter à du contenu récent. Et inversement orienter le récepteur vers du contenu d'archive et valoriser cette mémoire. Archive -> Sujet -> Archive et Sujet -> Archive -> Sujet.

11

Aux programmes classés automatiquement : KIWI

Un bon exemple est celui de BBC World Service qui a récemment numérisé tous les programmes pré-enregistrés : trois années d'écoute continue couvrant soixante ans d'événements divers. Les métadonnées sont minimalistes ou éronnés mais le contenu est de haute qualité en lossless 48k pour un tout d'environ 500TB de données. Kiwi est développé pour tenter de classer automatiquement. Il commence par essayer de faire une transcription via un programme opensource qui s'appelle CMU Sphinx, qui rend 50% d'erreurs sur les mots. Ils ont donc créé un algorithme pour classer les transcriptions et donc les trier pour jeter le surplus en utilisant les URIs d'un certain nombre de sources de données. A cette effet ils utilisent Amazon Web Services pour distribuer ces grandes quantités de données en peu de temps. Cette API classe automatiquement 50.000 programmes d'environ une heure. Par exemple pour un programme sur le 11 septembre des erreurs sont constatées, l'algorithme crée les mêmes catégories pour un programme que des catégories ajoutées manuellement dans environ 20% des cas, les 80% restants sont soit plus détaillés que la personne qui tag manuellement et ne lit que le synopsis, soit se trompe complètement voire même de manière offensante. Cela permet de poser la question de laisser le public corriger et ainsi contribuer à l'amélioration participative des contenus.

12

Question : sphynx et la reconnaissance vocale

Bertrand Sajus, Ministère de la culture : sphynx c'est de la reconnaissance vocale ?

Oui.

13

Question : chronologie et réactions

Question :

50% des tags sont à jeter et peuvent être offensants : depuis combien de temps est-ce que c'est proposé et quelles sont les réactions globales ?

Réponse :

Il faut prendre en compte que c'est un jeune projet (de six mois) et c'est la deuxième fois qu'on en parle publiquement. Les objectifs sont donc maintenant d'étudier la faisabilité d'un système et les rapports avec les utilisateurs. Un premier prototype va être lancé avec un groupe d'utilisateurs de World Service cet été pour voir si cela vaut le coup d'aller plus loin ou pas.

14

Question : paramétrage des tags

Question :

Peut on paramétrer pour ne conserver que les tags les plus probables ?

Réponse :

Sphinx ne fait pas de tags. Il donne une transcription automatique et linéaire. C'est un autre algorithme qui prend cette transcription et fait une liste de tags avec la probabilité que ce tag peut être associé à ce programme. On peut donc facilement limiter le nombre de tags pour améliorer la confiance dans les tags extraits par le système.

15

Question : comparaison ontologique

Question :

Est-ce que ce qui est extrait par Sphinx peut être comparé ontologiquement pour les re-filtrer ?

Réponse :

Nous avons un modèle qui utilise en gros, le graphe de données de DBpedia, il utilise la connectivité entre différentes sources pour pousser dans un sens ou un autre pour désambiguïser. Exemples le mot clé "white house" peut être chanson ou un édifice.

16

Question : comparaison ontologique et contribution

Question :

La BBC apprécie et utilise les contributions des utilisateurs sur le site. Au-delà de la contribution des sources comme Wikipédia, que devient ce mis à contribution pour les interfaces en général ?

Réponse :

Ils veulent vraiment expérimenter. C'est un travail monstrueux de vérifier les tags manuellement, la question se pose d'ouvrir cette vérification à la communauté très présente. C'est très intéressant pour le World Service car il existe une communauté très forte d'utilisateurs, les archives pourraient y être ouvertes globalement. Le prototype pour cet été s'adresse à eux, il sera ouvert d'abord à un public assez restreint pour plus tard s'ouvrir à tous.

17

Question : partenariats

Question :

Existe-il des partenariats avec le chapitre français de Wikipédia et l'anglais de Wikimedia avec des plateformes d'enrichissement, ou est-ce que ces choses sont à venir ?

Réponse :

Ce n'est pas mis en place à l'instant. Il existe quelques relations avec Wikimedia, mais rien de formel. Un nouveau projet, Wikidata, suscite beaucoup d'intérêt pour des organisations média (financé par Google...), comme le New York Times. Ils se posent la question si cette plateforme peut devenir un outil idéal à utiliser.

18

Question : tag INA

Question :

Par rapport à Voxlead, est-ce que le filtrage en amont d'occurrence est-il possible, par exemple un mot-clé qui a été très peu utilisé peut être obtenu malgré sa pertinence ? Comment faire la

balance entre trop filtré et pas assez ? Garder les mots-clés qui ont une occurrence faible mais qui ont une pertinence forte ?

Réponse :

Vous avez mis le doigt sur les trois facteurs de l'algorithme pour associer un tag particulier. Combien de fois ce tag est apparu ? À quel point ce terme est-il spécifique à ce document ? À quel point ce terme est près d'un modèle du sujet principale du programme.

Constatation :

la différence culturelle est difficile à surmonter pour appliquer ce modèle en France. Malgré que ce soit techniquement possible.

19

Question : contenu externe contrôlé

Question :

Quel est le statut du contenu sur le site, qui peut être modifié très rapidement et qui n'est pas contrôlé par la BBC ? Un site raccordé tombe que faites vous ?

Réponse :

Les sources de données ouvertes peuvent toujours disparaître. Oui cela arrive c'est toujours un compromis. Il faut avoir des mesures derrière pour éviter, en hébergeant sa propre version par exemple ou en répliquant le site. Chaque édit de Wikipédia à propos d'un artiste qui a été dans un programme va apparaître sur le site malgré quelques filtres. Les données reflètent le plus possible l'actualité. Par contre, les éditeurs et les artistes regardent leurs pages régulièrement pour vérifier qu'il n'y trouve rien d'horrible.

Question :

y'a t'il un cadre juridique plus souple en Grande Bretagne ?

Réponse :

Oui, Wikimedia à des soucis. Nous mettons une phrase qui explique l'origine des contenus et qui demande aux utilisateurs de modifier Wikipédia eux-mêmes s'ils voient quelque chose de faux. Nous renvoyons tout à la source qui est de fait responsable. Des notifications existe aussi pour avertir d'un changement de contenu. De plus il faut informer des URIs qui ne sont plus pérenne.

20

Question : utiliser DBpédia

Arthur Lutz, Logilab

Question : Comment fonctionne l'utilisation de DBpédia en détail ?

Réponse :

La BBC utilise DBpédia normal, c'est-à-dire celui qui est générer de manière périodique, DBpédia Live donne les données en temps réel mais est moins stable. DBpédia Lite peut être utilisé pour donner des URIs vraiment permanentes.

21

Question : Kiwi API

Arthur Lutz, Logilab

Question :

Comment est-ce que Kiwi API va être publié, en tant que service, en tant que logiciel ?

Réponse :

Avec RDM SIM nous sommes en train de clarifier la licence dans laquelle ils vont le publier dans les prochaines semaines.

22

Adresser et annoter les fragments de contenus multimédias

Présentation d'une technologie du W3C, une billet de blog classique présente une vidéo en citant les passages intéressants, c'est compliqué. Un lien pointant directement sur une portion de vidéo serait parfait. Youtube propose maintenant cette option de partage. C'est cela le principe du media fragments.

23

Qu'est-ce que c'est q'un fragment de média ?

C'est le fait d'adresser, dans une URI, une séquence temporelle ou une région spatiale d'une image. Ainsi que des notions de track ou pistes comme sur un dvd (langues, accès). Un fragment temporel de média a des sections mises en avant, mais sans perdre le contexte de la vidéo dans son intégralité. Dans le temps mais aussi dans l'espace, un fragment spatial de média obscurcit toute l'image sauf la région sélectionnée.

24

Les cas d'utilisation

Les "Use Cases" sont des fragments de médias nombreux : les mettre en favoris ou de les partager, les annoter pour une partie d'un document, faire des mash-ups (attention aux droits d'auteurs), et aussi de sauver de la bande passante en donnant les instructions nécessaires au serveur. Va devenir dans trois semaine une recommandation officielle. Cette spécification est donc stable.

25

Les requirements

Quatre dimensions : le fragment temporel pour isoler un segment et le connecter dans la vidéo avec un point de départ et un point d'arrivée. Le fragment spatial pour réduire à un rectangle : une bande in box. Le fragment de piste (track) sélectionner la piste pour les formats de vidéo qui contiennent plusieurs pistes, html 5 reconnait ces pistes et peut les adresser. Enfin named fragments un raccourci pour parler d'une séquence temporelle en la nommant pour la retrouver comme un chapitre de DVD.

26

Les URIs

Il existe deux composants avec lesquels on peut agir : la partie query modélisé par un "?" dans l'URI et la partie fragment avec un "#" dans l'URI. Ce sont deux choses différentes mais la syntaxe est la même, elles peuvent être combinées pour créer une nouvelle ressource avec ces signes.

Les fragments sont une ressource secondaire qui a un contexte intégral dans la vidéo permettant d'accéder au contenu à partir de n'importe quel endroit. Cacheable. Les queries sont différentes ; nous créons une nouvelle ressource qui est dérivée de celle qui précède mais on a perdu la relation c'est un nouvel objet sans lien avec l'original. Non Cacheable.

27

Media fragments resolution

Pour le query, le fichier vidéo est géré du côté serveur qui va générer et envoyer un nouveau fichier vidéo au client. Pour l'utilisation d'un fragment, il existe des cas différents. En particulier le client intelligent (user agent) va repérer qu'il y a un fragment et décider de demander seulement une partie. On ne remet pas en cause ce qui est établi sur le web.

28

Influence des formats de vidéos

Tous les formats de vidéo ne sont pas égaux, streamable ou pas, découposable ou pas, un constat qui a influé sur ce qu'on peut faire ou ne pas faire avec les vidéos. Un format moderne streamable MPEG4, OGG, etc... qui peut être joué par partie est l'idéal pour le media fragments. Les séquences temporelles et pistes peuvent être isolées, les spatiales non, donc des besoins différents en bande passante. Le discours client serveurs entre en relation seulement avec temporel et track.

29

Utiliser media fragments dès maintenant

Oui à condition d'un navigateur récent, firefox 9 minimum. Exemple donné avec une vidéo jouée dans un fragment x à y. Attention aux skins qui ne mettent pas le fragment en avant dans la timeline.

30

Sauvegarde de la bande passante : Sysnote non.

Le player Sysnote parse bien le fragment YouTube mais ne sauvegarde pas la bande passante.

31

Sauvegarde de la bande passante : Ninsuna oui.

Le player Ninsuna sauvegarde la bande passante et travaille avec la séquence temporelle et spatiale. De plus il illumine le fragment et avec les coordonnées spatiales on peut appliquer une div opaque et faire un recadrage d'une partie de la vidéo.

32

Examiner le trafic http

En examinant le trafic nous observons une sauvegarde de la bande passante.

Techniquement je décroche mais je constate une relation client/serveur entre la demande et la réponse en http.

33

Exemple : demander pour un fragment temporel correctement

Qui ne marche pas : un développeur envoie une requête, le serveur comprend Ninsuna et retourne une réponse 206 complète permettant de mettre en cache un byte range. Pour que cela fonctionne : la fonction "Include setup" permet de résoudre ce problème avec une réponse correcte ça marche.

34

Exemple : demander pour un fragment spatial

La spécification laisse le choix au client, rien n'est envoyé au serveur. Tout ce passe coté client et applique ses préférences.

35

Conclusion Ninsuna

NinSuna est une solution complète client/serveur qui permet de découper la vidéo de se servir des bouts et d'afficher des séquences temporelles et zones spatial. Il peut être utilisé comme proxy.

36

Générer des fragments avec Yuma.min.js

Yuma.min.js est un prototype qui à été créé dans le cadre d'Europeana pour mettre en liaison des bibliothèques et des institutions du patrimoine culturel. Il propose d'annoter spatialement et temporellement des documents multimédias avec Javascript en générant des médias fragments.

37

Produire des fragments audio-descripteurs avec ACAV (Annotation Collaborative pour l'Accessibilité Vidéo)

Le projet ACAV est un autre producteur des fragments de média qui essaie de rendre accessible les vidéos de DailyMotion aux malvoyants et aux malentendants. Le principe de base est de fournir des annotations est de rendre dans des players enrichis, en plage braille ou speech. On peut aussi ajouter des données sémantiques. Sphinx et HTK sont des outils complémentaires utilisables. Les retombés sont moyennes. Mais ce sont bien des média fragments liés et catégorisés.

38

Pour les média fragments, quel est le niveau de support pour les CDNs ?

YouTube et Dailymotion ont prévu de le gérer nativement. La principale raison de le faire c'est la sauvegarde de la bande passante, et ils espèrent donc que les autres CDNs dont Akamai et Squeed, le gère aussi.

39

Est-ce que ce format est le même que celui utilisé pour faire des sous-titres ?

Ce modèle de données est spécial, le sous-titre c'est un mélange de la partie de données (la voix) et la partie de la présentation (comment la présenter), et il faut dissocier ces deux choses. Mais les sous-titres sont possibles. De plus on peut importer et exporter des fichiers SRT, dans ACAV c'est abstrait. On peut exporter aussi en télétexte. Le modèle de données permet de supporter tout ça.

40

L'utilisateur peut-il partager des média fragments simplement ?

Nous avons voulu une syntaxe accessible. On peut imaginer des players qui font des liens qui peuvent être copiés, et partagés.

41

Est-ce que les télévisions deviennent Web ?

LinkedTV est un projet européen où les média fragments jouent un rôle premier. l'idée de base est d'effectuer de l'analyse sur ces contenus. Par exemple, l'émission Antiques Roadshow, amener des objets qui peuvent être liés avec les informations sur leurs histoires ou chronologies. Il peuvent être analysés, broadcastés et devenir des média fragments.

42

Vous utilisez HTK, mais aussi d'autres que vous avez créés ? Est-ce que DailyMotion veut transcrire tout son catalogue automatiquement ou est-ce qu'il existe un autre procédé ?

Ils ont fait leur propre modèle sur un investissement de 2.000€. Ils vont réfléchir sur la façon de le publier et le partager. Pour la question de la transcription, il est aujourd'hui possible de la faire plus vite que le temps réel. L'idée n'est pas de transcrire tout le catalogue de DailyMotion. Grâce à l'onglet «accessibilité» les utilisateurs peuvent indiquer des préférences de handicap et se lier avec la communauté des annotateurs qui veulent prendre le temps de corriger les transcriptions automatiques, ajouter des informations supplémentaires. Une fois les annotations validées elles sont consultables et utilisables en sous-titres, close caption, etc. L'utilisateur "Officiel" parmi motion maker et lambda aura lui, le droit de créer et annoter.

43

Introduction à la présentation et aux projets de l'IRI

Par Vincent Puig

Nous allons ouvrir le capot de ces outils avec Raphaël Velt, mettre en pratique ces technologies avec les utilisateurs de façon pragmatique. Nous collaborons avec le CEA et d'autres laboratoires. Ces travaux sont articulés avec la question de l'annotation et les méthodes automatiques. Cela a commencé avec Ligne de Temps en collaboration avec l'ANR CinéLab pour aboutir à cette présentation. Nous parlons plus de digital studies que humanities, étudier les pratiques et la modification qu'elle peut avoir sur les disciplines.

44

Cinq enjeux brièvement présenté

1) L'économie de la contribution et la figure de l'amateur

Phénomène majeur depuis 5/6 ans

2) L'organologie numérique

Etudier des dispositifs techniques au cœur d'une pratique

3) Les métadonnées et la folksonomie

Notamment à l'intérieur de vidéo mais plus largement comment les croisées

4) Le travail collaboratif et la publication contributive

La science progresse grâce à la contribution, il faut apprendre à ne plus travailler seul

5) La sensori-motoricité

Le rôle du corps dans le numérique

45

L'organologie numérique

Présentation des outils. La question des outils pour une partition musicale qui mélange les notions de fragments temporels et de fragments spatiaux pour annoter une note ou une mesure.

46

Le projet Ligne de Temps

Ligne de temps est un annotateur qui va spatialiser le temps du film de en le décomposant avec un algorithme automatique selon les besoins de l'utilisateur. Les fragments sont ici symbolisés par des segments hors champs. Ce principe se matérialise avec un lien menant à une vue, cette vue est décrite et possède une identité avec une url. Nous espérons une standardisation rapide. Ces vues sont graphiques, textuelles et modulables, elles peuvent aussi utiliser un player extérieur Youtube ou DailyMotion et produisent un "regard signé"

47

Regard signé

Le film "regard signé" n'est pas tout à fait un nouveau film, parce que l'original n'est pas modifié, mais il est ré-assemblé d'une autre manière en ajoutant un commentaire et à ce tire produit un nouveau regard signé. Une autre forme d'authoring. Les méta données ne sont pas bien sûr oubliées mais nativement collaboratives. Ligne de Temps est plutôt un back office qui produit des méta données, un outil d'écriture sur l'image qui est publié ensuite par l'outil Metadata Player. En cela il se rapproche de la notion du montage cinématographique propre au travail artistique augmenté de la notion de jeux.

48

Metadata Player

C'est un outil qui permet la publication du travail dans un format plus simple.

Les médias de Ligne de Temps sont tous sur une plateforme. C'est un annotateur lourd.

L'annotateur léger est Metadata Player qui permet d'agréger de la vidéo en la pointant directement et de la synchroniser avec ses annotations ainsi qu'avec Polemic Tweet. Il propose enfin un nuage de mots clés extrait des tweets.

Ceci est comparable au passage du cinéma à la vidéo.

49

L'intérêt de cette plateforme

Elle permet de rechercher dans les vidéos. L'avantage de ce système d'annotation est de rajouter sans avoir besoin d'algorithmes de la recherche intra vidéo. Exemple avec "music" ou est elle faite et dans quel média est il utilisé et surtout dans quelle endroit à l'intérieur de la vidéo avec éclairage du segment concerné. La base de données devenant presque un système autonome.

50

Exemple de publication Metadata Player et Ligne de temps dans le domaine du cinéma

Travail de Christian Licop de l'institut Mines télécom a fait un travail de réflexion sur l'image pour analyser les relations d'autorité dans le film "Entre les murs". Tout les commentaires ont été re-synchronisés pour analyser le film en croisant les regard signés. Les endroits sans commentaires sont relatifs à deux scènes où la relation professeur élèves n'existe pas. Ce sujet révèle l'intérêt particulier des uns et des autres.

51

Projet ANR : Eulalie

Un projet ANR qui s'appelle Eulalie sur les nouvelles formes d'écriture collaborative de fiction. Avec Orange et TelFrance. Il existe plusieurs prototypes de formats de publications, comme les vidéo-livres ou le projet de France Culture ou Arte. Aujourd'hui, avec Eulalie c'est une fiction de série télé sur le web qui s'appelle Anarchy où vous êtes projeté dans un monde limité financièrement. Les annotations influencent le déroulement du scénario.

52

Les fonctionnalités de Polemic Tweet

Les fonctionnalités de bases sont toujours présentes comme autoriser son compte Twitter à dialoguer avec Polemic Tweet. Maintenant on peut facilement voir les conversations et les polémiques en zommant/dézoignant et en se situant facilement dans les conversations. Il apparaît que les tweets négatifs repoussent les gens.

53

Le nuage des mots-clés

On peut séparer les mots extraits des contenus tweetés avec les couleurs dans lesquelles ils ont été prononcés. Une expérimentation est menée avec les mots-clés pouvant aussi être suggérés à l'avance.

54

Bubble-T

Développé dans un concours de visualisation de données lancé par Google pendant les élections qui utilise les données de Twitter. Nous avons voulu nous démarquer en étant ludique, on peut voir l'évolution des citations des politiciens en temps court ou long, avec des mots clés singuliers. Avec Google News nous voyons les hommes politiques par période et par mot clé avec son évolution dans le temps. Cela se rapproche du métier de data journalisme.

55

L'évolution de ces outils, la contribution

Quelle editorialisation de ces outils et comment le web sémantique peut aider au travers de projets futurs, comme le portail des arts ou Hda. L'idée générique est d'ouvrir les silos de données et d'appuyer sur la contribution.

56

Exemple avec des étudiants de l'Ecole du Louvre

Des étudiants de l'Ecole du Louvre sont évalués sur leurs annotations du présent séminaire dans Ligne de temps et notés sur ce travail. Les segments de leurs annotations deviennent

potentiellement des références bibliographiques pour enrichir les réseaux sociaux comme Wikipédia avec des contenus textes et vidéos grâce aussi au devenir de Media fragments.

57

La plateforme HdA Lab

Un plateforme qui utilise les données de l'HdA de la plateforme du ministère de la Culture avec 350 contributeurs institutionnels pour thématiser des cours. La création d'un back office permet de sémantiser les tags et proposer des outils de navigation plus évolués et modulaires créés par l'IRI pour d'autres projets. Avec dbpédia nous pouvons apporter 400.000 entrées de Sémanticpédia.

58

Questions

Questions :

Quels comptes Twitter utilisez vous ? Est-ce que Polemic Tweet est la bonne manière de visualiser un nombre très conséquent de données et alors comment les Tweets nombreux peuvent ils être bien représentés ?

Réponses :

Des comptes publics. L'interface change en fonction du nombre des Tweets et le système doit être testé dans des cas différents ce sont des travaux en cours. Il y a un changement d'échelle en fonction de la granularité mais oui cela pourrait montrer de limites. Un test sur une émission avec des niveaux élevés est à prévoir d'urgence.

59

Est-ce que le code source de Metadata Player est disponible ?

Réponse :

Il a vocation à l'être, en termes de licence, il n'existe pas d'obstacle, mais il reste un peu plus de travail à faire en termes de documentation et d'organisation.

Cette question anodine reflète bien les préoccupations d'une personne concernée et désireuse d'appliquer dès demain tous les principes évoqués ci-dessus. De grandes ambitions limitées au seuil technologique, la disponibilité de ces données et son implication socio-sémantique demeurant l'attrait principal de l'humanisme digital il est regrettable qu'il s'arrête à un obstacle technique qui sera résolu d'ici quelques années. Néanmoins l'IRI contribue à réduire par son expérimentation et ses partenariats l'écart entre science et application de la vie de notre époque en mesurant le recul historique nécessaire.

60

Remerciements et annonces Kiwi, développeur traque, etc...

Merci de votre attention.
Gilles Godiveau