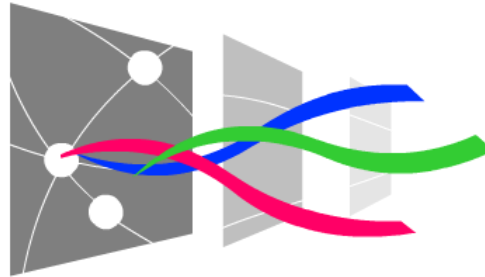


Periplus



Articuler éditorialisation algorithmique et humaine

Nicolas Sauret (IRI)
nicolas.sauret@iri.centrepompidou.fr



Bertrand Delezoide (CEA-List)
bertrand.delezoide@cea.fr



Periplus

Plateforme multi-terminaux de navigation pour des contenus journalistiques multimédias

- Projet ANR 24 mois
- 5 partenaires :
 - Recherche en STIC CEA, Alcatel-L. Bell Labs
 - Design et datavisualisation IRI
 - Usage CNRS-Lutin
 - Industriel Diotasoftware
 - Média Mediapart



Objectifs

Search, navigation et éditorialisation des archives multimédias de Médiapart

Articuler une approche algorithmique et une approche éditoriale

→ Datamining et datavisualisation au service de la valorisation des archives

Technologies mises en oeuvre

- Indexation multimédia manuelle, automatique, collective (sur données non-structurées)
 - Recherche de documents hétérogènes & multimedia
 - Classement conceptuel (clustering)
 - Génération automatique de résumés multimédia
- Datavisualisation
 - Annotations fines
 - Interactions gestuelles

Design Mediapart *by Periplus*

5 années archivées

HEADER

MEDIAPART *by Periplus*

Rechercher **SEARCH**

CONNECTEZ-VOUS

Identifiant

Mot de passe oublié ?

OK

INTERNATIONAL FRANCE ECONOMIE CULTURE CLUSTER X

DOSSIERS DOCUMENTAIRES

STREAMGRAPH

2008 2009 2010 2011 11/2011 01/2012 2012

Hadopi 17 octobre 61 Rentrée littéraire hiver 2012 BD Affaire de Karachi Affaire Bettencourt Syrie

AFFICHAGE

Treemap Liste

FILTRE

De : 01/11/2011 à : 01/01/2011

FILTRES

International France Economie Culture

DOSSIERS ASSOCIÉS

Affaire Bettencourt

Richard Descoings

Crise États-Unis

Jeunesse Ump

Cinema

Philosophie

Henry Jenkins

Derniers Meetings

Peintures..

Syrie

Informations utiles

- Foire Aux Questions
- Conditions Générales d'Utilisation
- Mentions légales
- Le projet Periplus

Raccourcis

- Année 2008, 2009, 2010, 2011, 2012
- International, France, Economie, Culture
- Les dossiers les plus populaires

Partenaires

iri

cea

Alcatel-Lucent

FOOTER

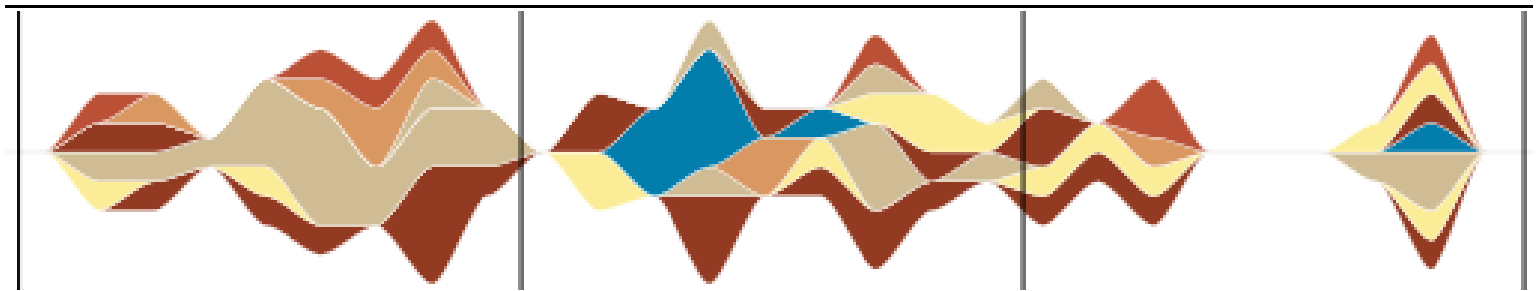
Le Lorem Ipsum est simplement du faux texte employé dans la composition et la mise en page avant impression. Le Lorem Ipsum est

© Mediapart © Periplus | Mentions légales | Charte éditoriale | CGV | Mediapart.fr

Streamgraph & treemap

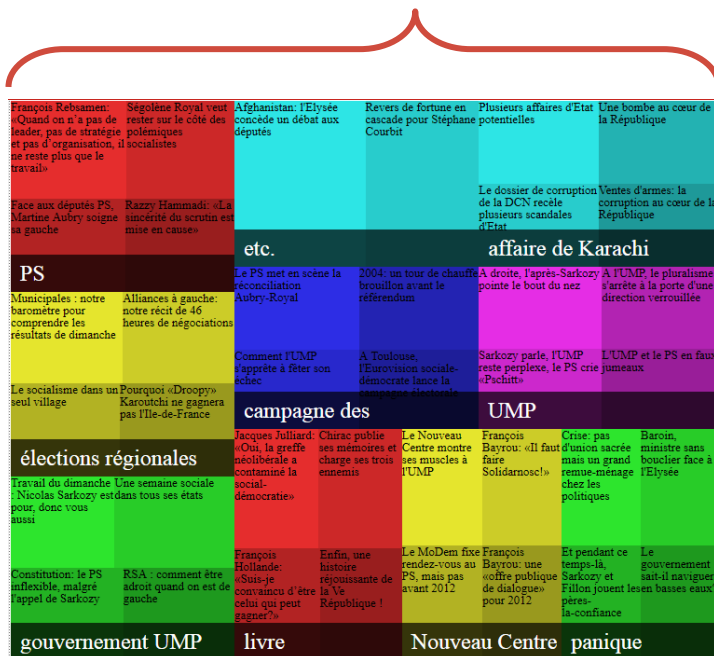
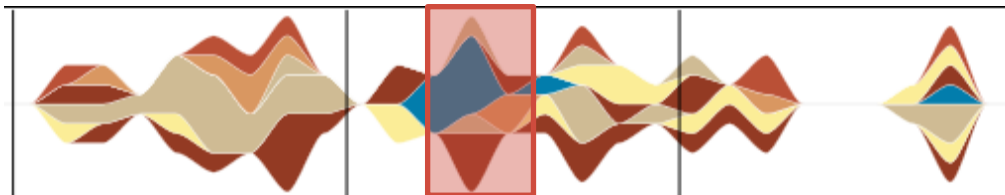
Streamgraph : représentation quantitative de l'ensemble des clusters dans le temps

→ volume d'articles d'un cluster dans le temps



Streamgraph & treemap

Treemap : représentation éditoriale d'une sélection temporelle → résumé de clusters

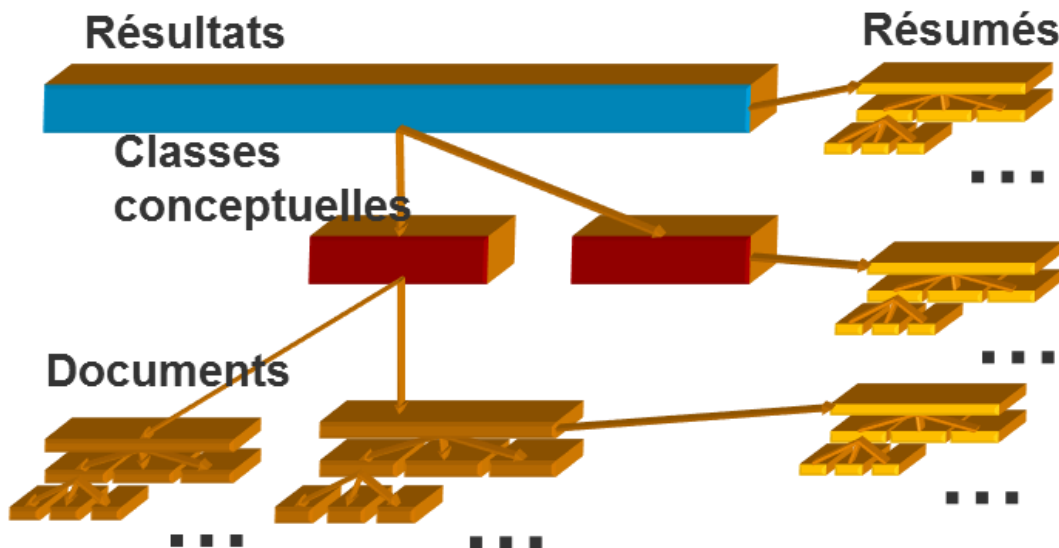


Modèle d'information

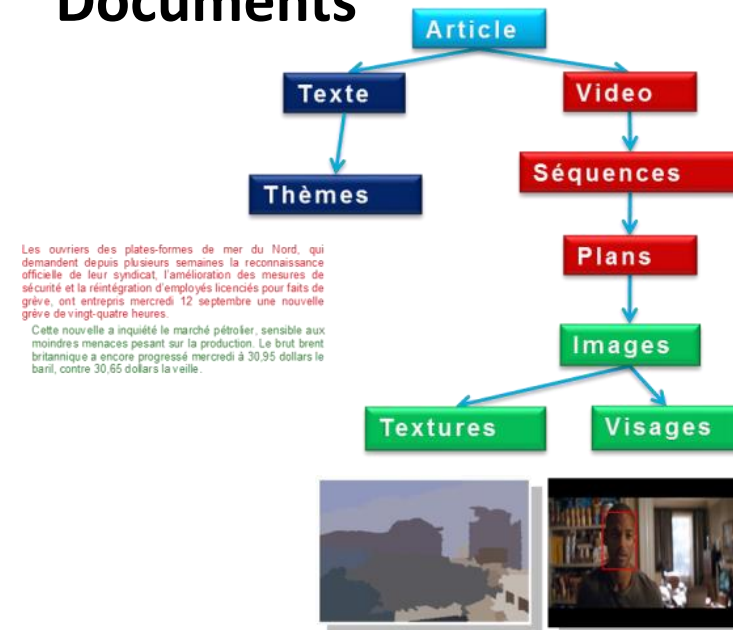
Modèle de **dossier documentaire** à trois niveaux structurels hiérarchiques : **résultats** d'une recherche, **classes** conceptuelles, **articles**.

A chaque niveau un **résumé** est synthétisé

Dossiers documentaires



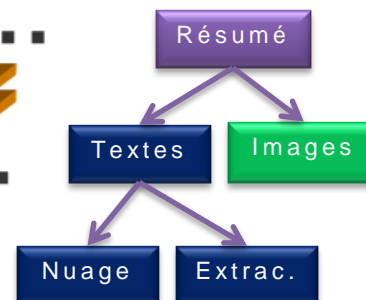
Documents



Les ouvriers des plates-formes de mer du Nord, qui demandent depuis plusieurs semaines la reconnaissance officielle de leur syndicat, l'amélioration des mesures de sécurité et la réintégration d'employés licenciés pour faits de grève, ont entrepris mercredi 12 septembre une nouvelle grève de vingt-quatre heures.

Cette nouvelle a inquiété le marché pétrolier, sensible aux moindres menaces pesant sur la production. Le brut brut britannique a encore progressé mercredi à 30,95 dollars le baril, contre 30,65 dollars la veille.

Résumés



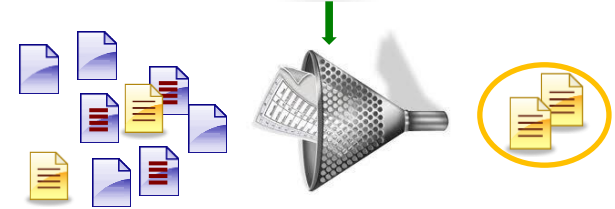
Traitement multimédia

Moteur de recherche : Retrouver un ensemble de documents à l'aide d'une requête.

Clustering : Regroupement automatique de documents fondé sur leurs similarités sémantiques.

Synthèse de documents : Produire une vision courte du contenu d'un document.

Enjeux : Multimedia, Crosslinguisme, Rapidité, Volume.



Clustering



Nuage



Template

Incident:	Green
Incident:	Location
Incident:	Type
Perpetrator:	Individual ID
Perpetrator:	Organization ID
Perpetrator:	Org. Confidence
Physical Target:	Description
Physical Target:	Effect
Human Target:	Name
Human Target:	Description
Human Target:	Effect

Résumé Résumé



Indexation multimédia

Objectif : Extraire une description du contenu et de la structure d'un document multimédia.

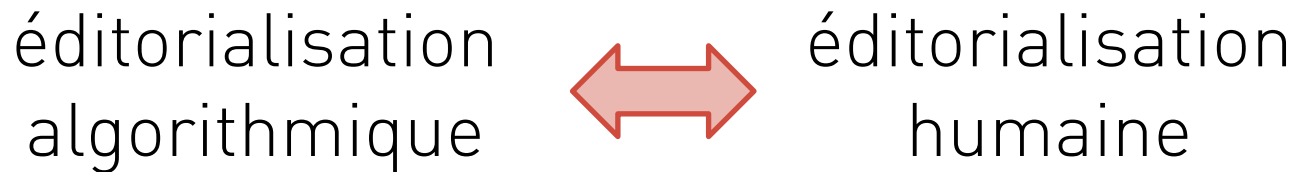
Challenges : Documents non-structurés,
Bigdata 3V: Variété, Vélocité, Volume .

Positionnement : sémantique, connaissance.



Articuler l'homme et la machine...

Du calcul de clusters à leur représentation



Appropriation des résultats de l'algorithme :

- Dossiers documentaires : tri et collecte par les utilisateurs
- Annotation fine : indexation manuelle

Dataviz des **Dossiers de la rédaction**



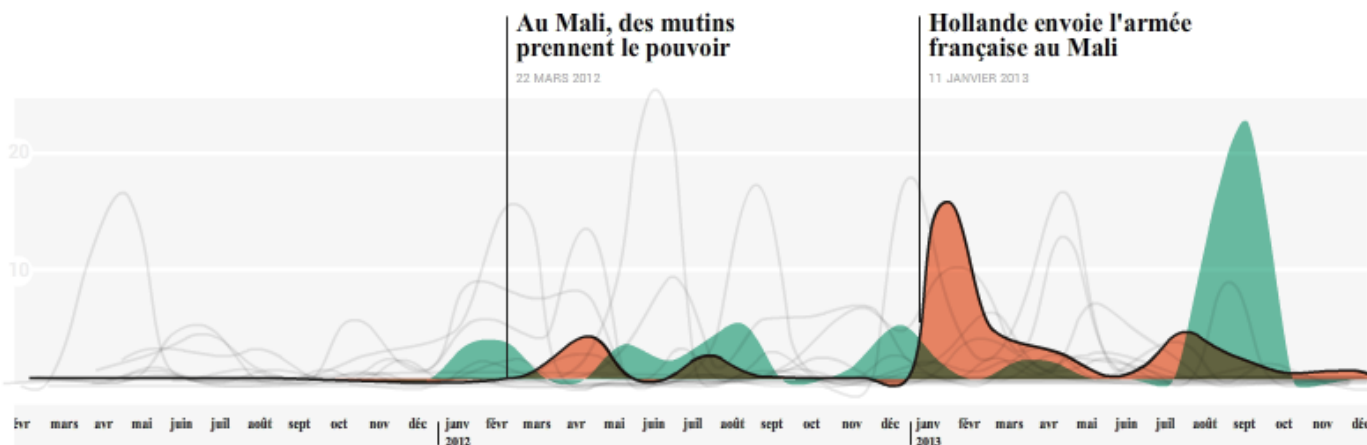
MEDIAPART

MAR.19 NOVEMBRE 2013 – ÉDITION DE LA MI-JOURNÉE

RSS
TWITTER
FACEBOOK
APPLICATIONS
FRENCH LEAKS

LE JOURNAL | INTERNATIONAL | FRANCE | ÉCONOMIE | CULTURE | *ENGLISH* | *ESPAÑOL* | LE CLUB | PARTICIPEZ

TOUS LES DOSSIERS



INTERNATIONAL



Mali, guerre et crises

Retrouvez nos articles, analyses, reportages, entretiens sur ce pays au cœur d'un Sahel en crise.

MIS À JOUR LE 04/11/2013 (82 ARTICLES)

Le compte Cahuzac	Retraites, la réforme 2013	Notre Dame des Landes
Mali, guerre et crises	L'affaire Tapie-Lagarde	Elections allemandes: Merkel-III
Le mariage pour tous	L'affaire Bettencourt	Social : l'état d'urgence
Nicolas Sarkozy, président sortant	Guerre civile en Syrie	L'affaire DSK
Hollande 2012	Législatives 2012	La réforme bancaire