



ERIC Lab and digital Humanities

- Numerous new innovative works CS+LSSH
- National and international projects
 - VUE, Armadillo, NORA, SEASR, etc.
- ThatCamp Paris (May 18th/19th 2010)
Manifesto signed by 214 academic people
- Digital humanities:
 - Linguistics, Social Sciences, Humanities
 - Convergence of various communities
 - Based on previous works and collaborations
 - Deeply multi-disciplinary
- Huge issues for both teaching and research



Outline of this introduction

- Opportunities raised by the digital humanities
- Complex data are ubiquitous
- First definition and relation with data mining

Computer Science for interactional corpora

- CLAPI
 - Corpus de Langues Parlées en Interaction
ICAR lab (C. Etienne, C. Plantin, L. Mondada...)
 - IS developed in collaboration with ERIC (F. Bentayeb, S. Loudcher...)
- Example: advertisers work meeting

CORPUS >>> RÉUNION DE TRAVAIL ENTRE PUBLICITAIRES - LYON SAXE			Durée	Lieu
Collecte	le 13-07-2004	>>> AFFICHER	04:00:55	Nb Locuteurs
Présentation				5 principaux
Responsable	Lucas BALTHASAR - Lorenza MONDADA			
2 Enregistrement(s)				
ENREGISTREMENT >>> EXTRAIT LSG 35 : BISCUIT BUSINESS			Lieu	Lyon
Durée	00:15:46	Support		vidéo, anonymisé, de bonne qualité
Recueil	visible			
Genre Détailé	Interactions entre associés d'une entreprise de consulting en contexte privé (domicile)			
Genre Interactionnel	Interactions de travail; Interactions privées			
Collecteur	Lukas BALTHASAR, Lorenza MONDADA			
Présentation	Enregistré le 13 juillet 2004 au domicile de Jean-Baptiste. Enregistrement multisupport. L'extrait se situe entre 01:12:44 à 01:28:29 de la to			
Accès	EN LIBRE ACCÈS			
1 Transcription(s)	No			
Locuteur JEB	Saxo Marcellin, né(e) en 1969-1970. Âgé(e) de 35-45 ans au moment de l'enregistrement; Consultant			
Etudes	Supérieures; Ecole de commerce			
Langue	Matielle --> Français Autre --> Anglais, allemand			
Père du locuteur	Langue --> Français			
Mère du locuteur	Langue --> Français, allemand			
Commentaires	JEB / Jean-Baptiste			



```

174 <token rang="8">on</token>
175 <token rang="9">sait</token>
176 <token generique="que" rang="10">que::</token>
177 </productionVerbale>
178
179 <productionVerbale pseudo="M" rang="5">
180 <space longueur="10" rang="1"/>
181 <chevauchement type="debut" position="Externe" rang="2">[</chevauchement>
182 <token rang="3">oui</token>
183 <space longueur="3" rang="4"/>
184 <chevauchement type="fin" position="Externe" rang="5">]</chevauchement>
185 <space longueur="2" rang="6"/>
186 </productionVerbale>
187
188 <productionVerbale pseudo="C" rang="6">
189 <token rang="1">on</token>
190 <token rang="2">sait</token>
191 <token generique="positionné" rang="3"/>
192 <token rang="4">tous</token>
193 <token generique="général" rang="5"/>
194 <token rang="6">parkings</token>
195 <token rang="7">d'</token>
196 <token generique="attente" rang="8">atTEnte</token>
197 <token rang="9">la/</token>
198 <token rang="10">zone</token>
199 <token rang="11">de/</token>
200 <token generique="déserte" rang="12">deSSERte</token>
201 <token rang="13">l'</token>
202 <token rang="14">accès</token>
203 <token generique="principal" rang="15">principAL</token>
204 <token rang="16">l'</token>
205 <token rang="17">'accès</token>
206 <token elision="" rang="18">d'</token>
207 <token rang="19">service</token>
208 <token type="courte" duree="" rang="20">(.)</pause>
209 <token generique="h" rang="21">.h.:</token>
210 <token rang="22">la/<token>
211 <token rang="23">seule</token>
212 <token rang="24">chose</token>
213 <token type="courte" duree="0.2" rang="25">(0.2)</pause>
214 <token rang="26">qui</token>
215 <token rang="27">serait</token>
216 <token rang="28">à</token>
217 <token rang="29">modifier</token>
218 <token rang="30">pour</token>
219 <token rang="31">l'</token>
220 <token generique="instant" rang="32">instant</token>
221 <token rang="33">.h:</token>
222 <token rang="34">c'</token>
223 <token rang="35">est</token>
224 <token rang="36">on</token>
225 <token rang="37">sait...</token>

```

Computer Science for historical data

- SyMoGIH
 - Système Modulaire de Gestion de l'Information Historique
LARHRA, pôle méthode (F. Beretta, P. Vernus...)
 - IS développé en collaboration avec ERIC (J. Darmont, O. Boussaïd...)
- Exemple: base de données « cartes postales »

ID: 22

Titre : Deux petites filles en pied l'une portant un panier

Support: Carton Fin

Taille: Photo-carte de Visite

Nature: Noir et Blanc

Legende Verso: Ethel and Grace

Photographe(s) :

1:Nom: WADE G

Thématique(s) :

Cadrage --> En pied

Genre et âges de la vie --> Enfants

Photographe ID: 10891

Nom: WADE Prénom: G

Sexe: Homme Pays: Angleterre

Technique: Plaque Sèche

Activité Principale: Photographe de studio

Stock: Oui

Date début activité: 1880

Computer Science and Journalism

- Collaborative work with researchers in communication and information science
 - Centre M. Weber (J.C. Soulage...), Univ. of Parana (Brazil)
- Urgent need of extracting useful information from huge data repositories on the Web, in particular for developing modern research engines
- In particular:
 - « How to surface the best comments, videos and pictures from a variety of sources in real time and then how to verify them ? »
 - « How to quickly surface the best comments and work out which ones are worth investigating further ? »
 - « How to identify quickly the key influencers on any particular story, so they can get inside information or interview them for their news outlets ? »



5 septembre 2016

LE HUFFINGTON POST

EN ASSOCIATION AVEC LE GROUPE *Le Monde*

FR

À LA UNE POLITIQUE ÉCONOMIE INTERNATIONAL CULTURE MÉDIAS PEOPLE LE BON LIEN C'EST DEMAIN C'EST LA VIE LE HUFFPLAY

Primaire de la droite • Crise au pouvoir • La preuve en images • Réussir autrement • Alimentation • Sexualité • Déconnecter pour respirer • Ça marche • Vie de bureau

US

EDITION **US**

THE HUFFINGTON POST
INFORM • INSPIRE • ENTERTAIN • EMPOWER

NEWS POLITICS ENTERTAINMENT WELLNESS WHAT'S WORKING VOICES VIDEO ALL SECTIONS

5 setembro 2016

HUFFPOST BRASIL

ASSOCIADO A Abril

BR

Edição: BR ▾

HOME PAÍS MULHERES VOZES DA RUA LGBT RIO 2016 MUNDO EQUILÍBRO LIVROS E HQS VIRAL TEM JEITO!

Família • Homem Moderno • Animais • Tech • Meio Ambiente • Ciência • Saúde • Educação • Arte • Diversão • Comportamento • Esportes • LGBTfobia • 100% Carioca

sign in subscribe search

dating more International - theguardian

UK world sport football opinion culture business lifestyle fashion environment tech travel

home

headlines

Now 4°C 12:00 15:00 18:00 21:00 9°C 13°C 9°C 6°C Lyon

Climate change / February breaks global temperature records by 'shocking' amount 

Warnings of climate emergency after surface temperatures 1.35C warmer than average temperature for the month

Great Barrier Reef Severe coral bleaching worsens 

1,716

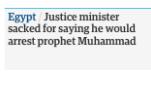
Japan US sailor arrested in Okinawa on suspicion of rape 

German elections Anti-refugee AfD party makes dramatic gains 

463

US elections 2016 Clinton and Sanders attack 'pathological liar' Trump 

1,489

Ivory Coast Gunmen open fire on tourist resort, killing 16 

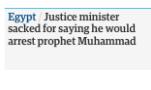
489

Brazil More than a million protest over 'horror' government 

303

Turkey President vows to defeat terror as bomb kills 34 in Ankara 

United Arab Emirates Plane reported missing in Yemen 

Egypt Justice minister sacked for saying he would arrest prophet Muhammad 

489

+ More headlines

highlights

100 Best Nonfiction Books of All Time / No. 7: *The Right* 

Outline of this introduction

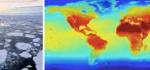
- Opportunities raised by the digital humanities
- Complex data are ubiquitous
- First definition and relation with data mining

Google climate change

Tous Actualités Images Vidéos Livres Plus Outils de recherche

Environ 142 000 000 résultats (0,29 secondes)

Images correspondant à climate change Signaler des images inappropriées

Plus d'images pour climate change

NASA: Climate Change and Global Warming climate.nasa.gov/ ▾ Traduire cette page Vital Signs of the Planet: Global Climate Change and Global Warming. Current news and data streams about global warming and climate change from NASA. Evidence - Scientific consensus - Causes - Effects

Climate change - Wikipedia, the free encyclopedia https://en.wikipedia.org/wiki/Climate_change ▾ Traduire cette page Climate change is a change in the statistical distribution of weather patterns when that change lasts for an extended period of time (i.e., decades to millions of ... Global warming - Scientific opinion on climate ... - Effects of global warming on ...

What is Climate Change? What Causes Global Warming? www.takepart.com/flashcards/what-is-climate-change ▾ Traduire cette page Climate change, also called global warming, refers to the rise in average surface temperatures on Earth. An overwhelming scientific consensus maintains that ...

Home | Climate Change | US EPA https://www3.epa.gov/climatechange/ ▾ Traduire cette page Official government site provides comprehensive information on the issue of climate change and global warming including climate change science, U.S. climate ...

News

Barack Obama a conclu une année d'âpre bataille politique. Le président américain a promulgué, mardi 23 mars, la réforme de l'assurance-maladie, dans la prestigieuse "East Room" de la Maison Blanche après que le texte a été adopté par le Congrès, dimanche.

Le texte va garantir une couverture maladie à 32 millions d'Américains qui en sont dépourvus. La réforme, d'un coût de 940 milliards de dollars sur dix ans, est jugée trop coûteuse par les républicains, unanimes dans leur rejet d'une loi qu'ils assimilent à une prise de contrôle du système de santé par l'Etat fédéral et à un gouffre budgétaire. Les républicains entendent faire de ce sujet l'enjeu de la campagne pour les élections de mi-mandat en novembre, qui renouveleront l'ensemble de la Chambre des représentants et un tiers du Sénat, tous deux contrôlés par les démocrates.

Short texts

Objective + subjective

Weak structure

"LES GAUCHISTES CONTRÔLENT VOTRE SYSTÈME DE SANTÉ"

Dès jeudi, Barack Obama doit faire un premier déplacement dans l'Iowa, où il s'agira de convaincre les Américains du bien-fondé d'un arsenal législatif devenu impopulaire, mais aussi d'éviter de laisser les républicains occuper le

Blogs, forums, chats

Sujet	Auteur	Envoyée le
■ Victoire idéologique de la gauche ?	The Reporter	02/04/2009 12:27
■ Re: Et la réponse fut	Flora	06/04/2009 10:04
■ Re: Et la réponse fut	The Reporter	06/04/2009 10:22
■ Re: Victoire idéologique de la gauche ?	The Reporter	06/04/2009 09:44
■ Re: Victoire idéologique de la gauche ?	Fertois	03/04/2009 07:17
■ Re: Victoire idéologique de la gauche ?	The Reporter	03/04/2009 11:42
■ Victoire de la prostration	dambredaine-la	03/04/2009 11:51
■ Re: Victoire de la prostration	The Reporter	03/04/2009 12:58
■ Les mecs pollis	dambredaine-la	02/04/2009 23:46
■ Re: Victoire idéologique de la gauche ?	Tesla	02/04/2009 14:14
■ Re: Victoire idéologique de la gauche ?	frecitizen	02/04/2009 17:30
■ Re: Victoire idéologique de la gauche ?	Tesla	02/04/2009 17:41
■ Re: Victoire idéologique de la gauche ?	frecitizen	02/04/2009 23:39
■ Re: Victoire idéologique de la gauche ?	Tesla	03/04/2009 09:13

Visitor has requested translation from language: French

Stephen says:

Hello Marie. My name is Stephen how can I help you?

(Bonjour Marie. Mon nom est Stephen comment ose je vous aide ?)

Marie On www.whoson.com says:

I need price [J'ai besoin de prix]

Stephen says:

Link Sent: http://www.whoson.com/installableorder.aspx

Stephen says:

Ok. For prices please go to http://www.whoson.com/installableorder.aspx

(Ok. Pour des prix veuillez vont à http://www.whoson.com/installableorder.aspx)

Toutes nos boutiques

Amazon.fr Ventes Flash · Meilleures ventes · Offres reconditionnées · Nos idées cadeaux · Services Amazon · Amazon Assistant

Star Wars : Battlefront - édition limitée · Commentaires client

Commentaires client

★★★★★ 59
3,2 sur 5 étoiles

5 étoiles 17
4 étoiles 14
3 étoiles 7
2 étoiles 8
1 étoile 13

Evaluez cet article Écrire un commentaire

Hidden for obvious reasons

Meilleur commentaire positif

Voir les 31 commentaires positifs >

★★★★★ Pas parfait mais un Star Wars
Par Client d'Amazon le 21 décembre 2015

Le titre pourrait être plus riche en terme de contenu, surtout en solo qui fait seulement guise d'introduction aux bases, mais l'immersion est tellement réussie que les fans de l'univers Star Wars seront conquis.

L'ambiance sonore et visuelle est magistrale, et incarner un stormtrooper en pleine bataille d'Endor ou sur Hoth est un réel plaisir !

A éviter si vous ne jouez pas en ligne.

Meilleur commentaire critique

Voir les 28 commentaires critiques >

7 sur 7 personnes ont trouvé cela utile

★★★★★ Déçu
Par julien [REDACTED] le 6 décembre 2015

Pas de campagne, juste un multi joueur qui se rattrape par de super graphisme mais ça ne suffit pas... Et bien évidemment le reste sera en DLC ce qui fera grimper le jeu environ 130€ (édition deluxe) donc pas pour moi...

Scientific articles

Un cadre pour la représentation et l'analyse de débats sur le Web

Anna Stavrianou*, Julien Velcin**, Jean-Hugues Chauchat*

ERIC Laboratoire - Université Lumière Lyon 2, Université de Lyon

avenue Pierre Mendès-France 69676 Bron Cedex, France

^{**} julien.vclcin@univ-lyon2.fr

Résumé. Les débats en ligne sont souvent modélisés par des réseaux sociaux d'utilisateurs représentés sous forme de graphes, chaque nœud correspondant à un d'entre eux. Ici, nous proposons un nouveau modèle basé sur un graphe de messages : chaque nœud du graphe correspond à l'un des messages échangés et chaque arc relie un message à celui auquel il répond. Ces arcs peuvent être caractérisés par les mots-clés résumant la relation entre les messages connectés. Cette modélisation permet une meilleure représentation de la dynamique des débats ainsi que l'identification de chaînes de discussions. Nous présentons des résultats obtenus sur deux datasets de débats en ligne et graphes représentant les messages, puis nous analysons la connaissance qu'il peut être extrait à partir de chacun d'eux. Nos expériences sur des débats réels valident le modèle proposé et montrent les informations complémentaires qui sont apportées par le graphe des messages.

1 Introduction

Le développement du Web2.0 a provoqué la création d'un grand nombre de blogs, de forums et de discussions en ligne. L'analyse de ce type de discussions est très intéressante, tant pour des organismes publics que pour l'industrie ou le secteur commercial. Les discussions en ligne contiennent les intérêts et les avis des internautes, des critiques de produits, la présentation d'idées politiques, etc. En conséquence analyser ces données devient une problématique stratégique.

Turney, P. et M. Littman (2003). Measuring praise and criticism: inference of semantic orientation from association. *ACM TOIS* 21(4), 315–346.

Zhang, J., M. Ackerman, et L. Adamic (2007). Expertise networks in online communities: Structure and algorithms. In *WWW '07: Proceedings of the 16th international conference on World Wide Web*, pp. 221–230.

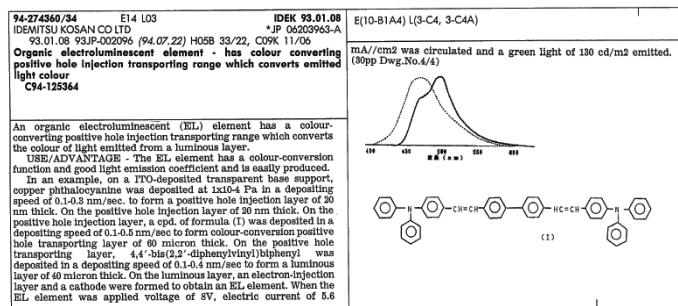
Zhou, D., I. Councill, H. Zha, et C.-L. Giles (2007). Discovering temporal communities from social network documents. In *ICDM: International Conference on Data Mining*, pp. 745–750. IEEE Computer Society.

Summary

The online discussions are often modeled as a social network of users and they are represented by a graph where each node denotes a participant of the discussion. In this paper, we propose a message-based graph representation of the discussion forum. In this model, each node corresponds to a message of the discussion and it has (and often) points out which message the specific node replies to. The edges can be weighted by the keywords that characterize the connected messages. This model allows a better representation of the content of the discussion and facilitates the identification of discussion clusters. We compare the two representations: the original message-based graph and our new message-based graph representation that can be extracted from them. Our experiments on real data validate the proposed framework and show the additional information that can be extracted from a message-based graph.

Patents

Outline of this introduction



- Opportunities raised by the digital humanities
 - Complex data are ubiquitous
 - First definition and relation with data mining

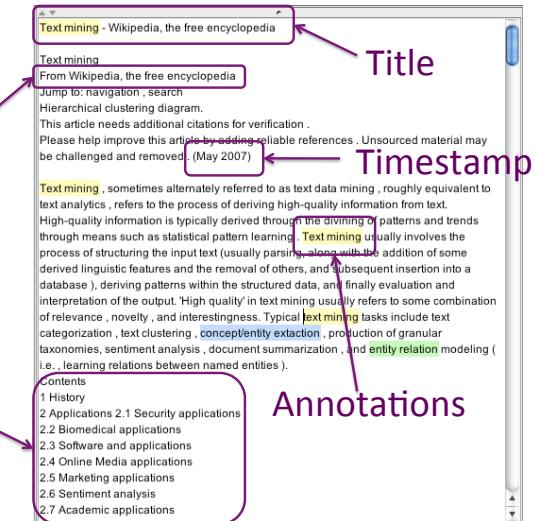
One possible definition

Text mining is...

« finding interesting regularities in large textual datasets and using them for solving a specific task »

Not only
textual
data...

Source



Outline

22

Text mining and more...

Here is a dialogue from the film **2001: A Space Odyssey**: Hal the computer and Frank the astronaut are playing chess.

Frank: Anyway, queen takes pawn.

Hal: Bishop takes knight's pawn.

Frank: Rook to King 1.

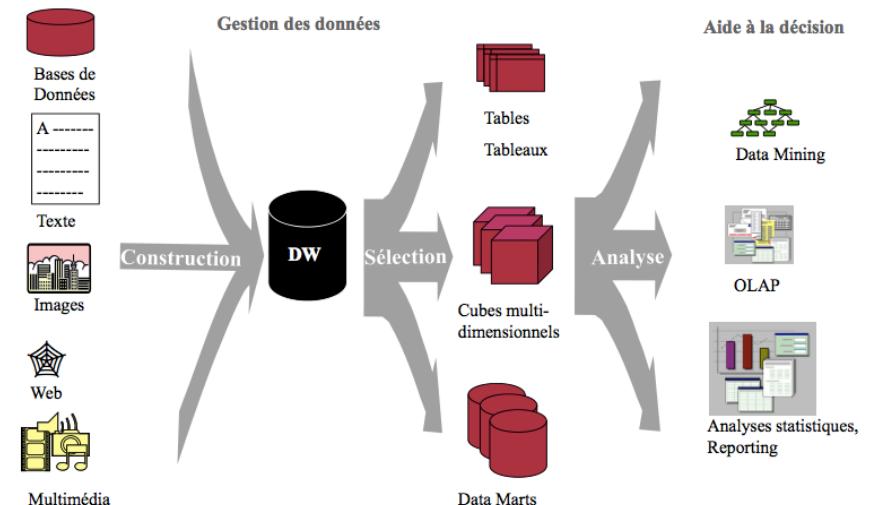
Hal: I'm sorry, Frank. I think you missed it. Queen to Bishop 3, bishop takes queen, knight takes bishop, mate.

Frank: Yeah, looks like you're right. I resign.

Hal: Thank you for an enjoyable game.

Frank: Thank you.

Data mining and business intelligence



23

What about Data Mining?

- First answer:
 - Good modeling entails good data mining: the power of OLAP
 - Datawarehousing does not fit alone all the issues
- Go beyond:
 - Enriching the data: knowledge integration, machine learning
 - Proposing novel and original indexes: clustering
 - **Content analysis:** text mining, web mining, image mining...
- Each point is associated to major challenges in CS!

Challenges in Data Mining

- **Representation of complex data**
 - Attributes and pertinence
 - Multi-view indexing
 - Fixing the curse of dimensionality!
- **Modalities fusion**
 - Different modalities: plain text, images, annotations, etc.
 - Various and heterogeneous sources
- **Semantic enrichment**
 - Integrating domain knowledge (the so-called “ontologies”)
 - Information retrieval, machine learning, etc.
 - From novel information to knowledge: the role of validation

Various disciplines involved

- Statistics, data analysis
- Linguistics
- Artificial Intelligence (AI)
- Information Retrieval (IR)
- Natural Language Processing (NLP)
- Computational linguistics
- Semantic Web
- Machine learning (deep learning ?)

« Some » difficulties

- A very large vocabulary
 - Concepts and instances
 - Underlying relation between words (synonymy, antonymy, meronymy, metonymy etc.)
 - Semantic ambiguity: “He saw the boy with his glasses.” (who has the glasses?)
- Pair-wise comparison
 - Curse of dimensionality
- Different tasks => different ways for representing the textual content

Course schedule

- Introduction
- Applications of text mining
- Basics in text mining:
 - representing and comparing texts
 - basic pre-processing: tokenization, stemming...
- Introducing knowledge into text mining
- Machine learning methods for text mining

Material on the Internet

- NCSU MSA Program Course Module
<http://research.csc.ncsu.edu/ase/courses/analytics/textmining/>
- *Information to Information Retrieval.*
Cambridge University Press [Manning et al., 2008]
<http://nlp.stanford.edu/IR-book/information-retrieval-book.html>
- Tutorial at EDBT'06
rene-witte.net/system/files/IntroductionToTextMining.pdf