

ANALYSE HAUT-NIVEAU DE DONNÉES HÉTÉROGÈNES ET MASSIVES : DE L'ENRICHISSEMENT VERS LA CONCEPTUALISATION

DOCTORIALES EARIA 2014

CONTEXTE SOCIÉTAL

- LA COMMUNICATION EST ESSENTIELLE CHEZ LES HOMMES
 - STOCKAGE
 - TRANSMISSION DE CONNAISSANCES
- DE PLUS EN PLUS DE DONNÉES SONT PRODUITES CHAQUE ANNÉE
 - 4,4 Z_o EN 2013 (1 Z_o = 1 MILLIARD T_o)
 - 44 Z_o EN 2020
- NOUVELLE PROBLÉMATIQUE : TROUVER L'INFORMATION PERTINENTE



CONTEXTE DE RECHERCHE

- COMMENT UTILISER LES DONNÉES DISPONIBLES POUR EXTRAIRE DE LA CONNAISSANCE ?
- OBJECTIFS :
 - COMPRENDRE L'ORGANISATION DES DONNÉES ET LEURS LIENS AU SEIN DE DOCUMENTS
 - APPRÉHENDER LE SENS PROFOND DE CES DONNÉES
- DIFFICULTÉS :
 - RIGIDITÉ DE LA MACHINE, MANQUE D'ABSTRACTION
 - INTERPRÉTATION DES DONNÉES (DIFFICILE MÊME POUR DES HUMAINS)



APPLICATION



- FERMETURE DU TRIBUNAL PÉNAL INTERNATIONAL POUR L'EX-YOUGOSLAVIE

- ≈ 240 000 DOCUMENTS DEPUIS LES ANNÉES 1990

- VIDÉOS (.ASF, .WMV, .MPG)
- AUDIO (.MP3)
- IMAGES (.TIF, .JPG, .GIF, .PDF)
- AUTRES (.DOC, .PPT, .RTF)

- ASSOCIATION OF CONCENTRATION CAMPS - DETAINEES IN BOSNIA AND HERZEGOVINA

- VIDEOS
- IMAGES
- ≈ 20 000 DOSSIERS D'ANCIENS PRISONNIERS



PROBLÉMATIQUES

- MOTEUR DE RECHERCHE ACTUEL PEU PERFORMANT
 - RECHERCHE PAR DATE, AFFAIRE, TYPE DE DOCUMENTS, CONTENU TEXTUEL
- ENRICHISSEMENT DES DONNÉES
 - ANNOTATIONS SÉMANTIQUES
 - LIENS ENTRE DOCUMENTS
- CONCEPTUALISATION
 - CRÉATION DE CARTES CONCEPTUELLES PERSONNALISABLES SELON LE CONTEXTE
 - JUDICIAIRE, HISTORIQUE, DEVOIR DE MÉMOIRE, ...
 - CRÉATION DE CHEMINS ENTRE DOCUMENTS
 - ...



MERCI DE VOTRE ATTENTION