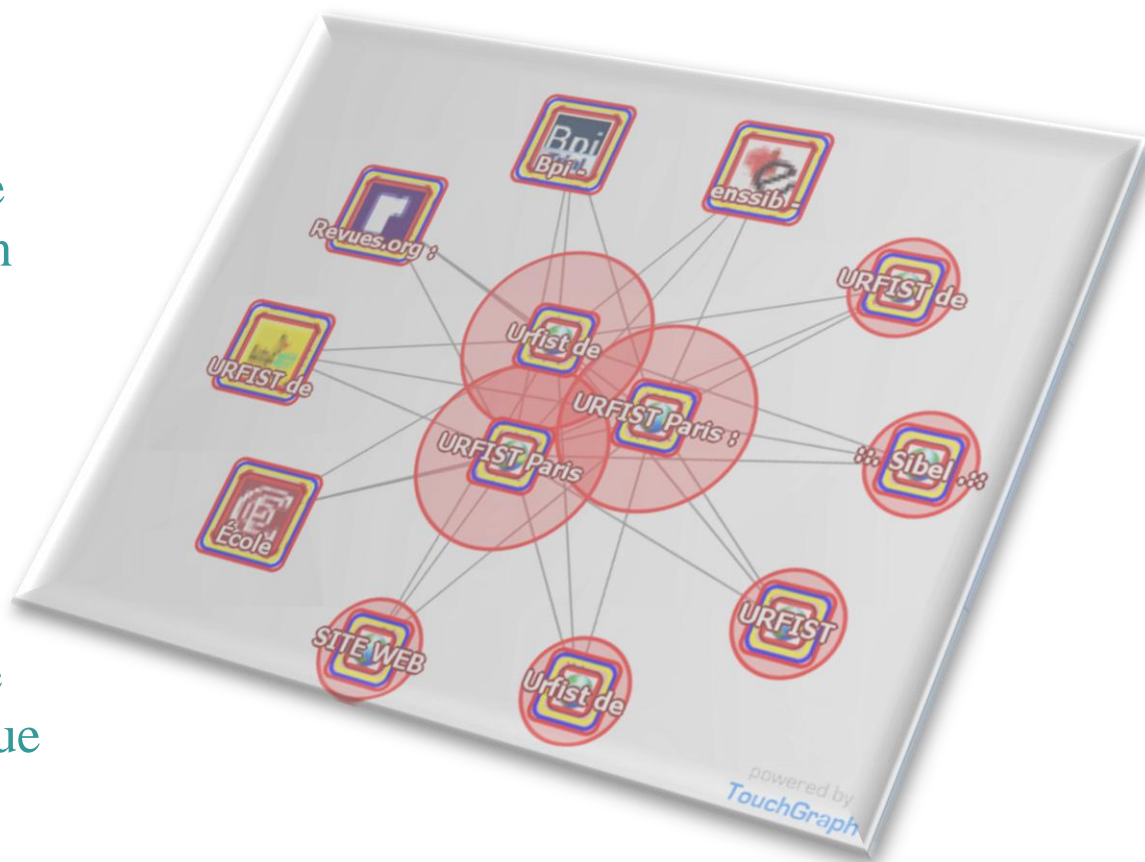


Cartographie de l'information

Objectifs de la journée

- Comprendre ce que recouvre la cartographie de l'information
- Savoir situer un outil dans une famille logicielle
- Evaluer l'apport d'un module de représentation cartographique



Cartographie de l'information

Introduction

- Présupposés : atouts supposés de la représentation visuelle
- Définitions : carte et cartographie
- Ancrage historique : scientométrie et Leximappe
- Problématique de la cartographie de l'information
- Ancrage disciplinaire : sémiologie graphique, sciences cognitives, gestion des connaissances

Composantes des applications cartographiques

1- Composante calculatoire

- Data Mining
- Text Mining
- Data Clustering
- Web Mining

2- Composante visuelle

- Éléments de sémiologie graphique (J. Bertin)
- Théorie de la Gestalt

Famille d'outils et contextes applicatifs

1- Famille d'outils

- Typologies
- Aperçu
- Travaux pratiques

2- Contextes applicatifs

- Éléments d'évaluation
- Les perspectives de la cartographie sémantique

Cartographie de l'information- Introduction

1- Présupposés : la **cognition externe** et les capacités visuelles

- *versus* la représentation interne -> raisonnement distribué
- support externe doté de plusieurs fonctions :
 - aide-mémoire
 - outil de communication-> permet d'accéder à des connaissances et des compétences inaccessibles par des représentations internes [Zhang, 1997].
- mécanisme perceptif
- toujours efficace ? Les représentations externes facilitent le travail cognitif seulement si la représentation est bien choisie ...

Cartographie de l'information- Introduction

1- Présupposés : la cognition externe et les **capacités visuelles**

- Traitement graphique des informations : nous sommes capables d'assimiler quasi instantanément et sans effort un grand nombre d'informations représentées graphiquement.
- Avantages de la visualisation [Ware, 2000; Ware, 2005] :
 - manipuler des structures complexes,
 - percevoir l'émergence de propriétés « inconnues »,
 - mettre en évidence des problèmes dans les données, dans leur collecte,
 - percevoir simultanément des propriétés à grande et à petite échelle sur les données,
 - faciliter la formation d'hypothèses sur les données.

Cartographie de l'information- Introduction

2- Définitions : **carte** et cartographie

Définition du Comité français de cartographie

La carte est une représentation géométrique conventionnelle, généralement plane, en positions relatives, de phénomènes concrets ou abstraits, localisables dans l'espace.

Principales propriétés

- Une carte est un document graphique donc visuel, qui obéit à une “sémiologie graphique” (Bertin).
- Une carte représente toujours un espace d'informations :
 - ces informations peuvent être abstraites (comme les données d'une organisation) ou scientifiques (c'est-à-dire issues de mesures physiques comme des données géographiques)
 - elles sont toujours sélectionnées : une carte ne fait apparaître qu'une catégorie d'informations - son thème.
- L'espace cartographié a pour objectif de transmettre lui-même des informations sur : la nature, les relations, la localisation ou l'importance des données ou des phénomènes qui le composent.

Cartographie de l'information- Introduction

2- Définitions : carte et cartographie

La cartographie de données non géographiques ou « visualisation d'information » :

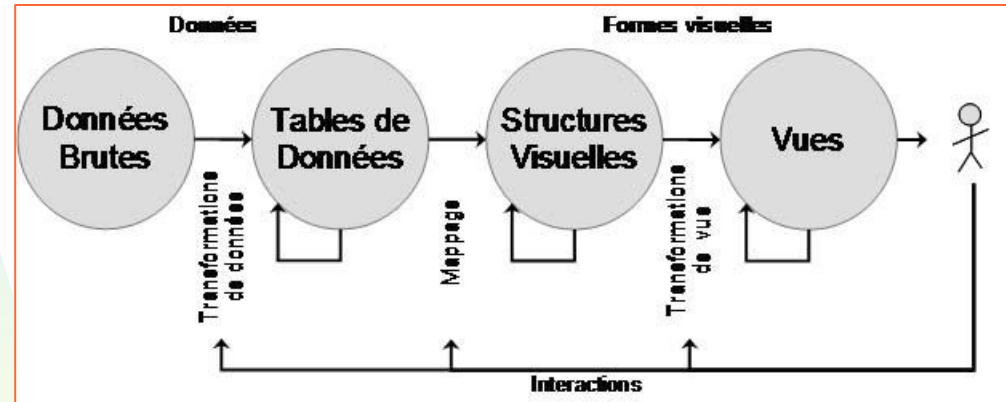
- « La visualisation est l'utilisation de représentations visuelles interactives et informatisées de données pour amplifier la cognition ».
- « visualisation scientifique » : toutes les visualisations de phénomènes physiques. La cartographie géographique entre dans cette classe.
- « visualisation d'informations » : toutes les visualisations de données abstraites (non liés à des phénomènes physiques).
- « La visualisation d'informations est l'utilisation de représentations visuelles interactives et informatisées de *données abstraites* pour amplifier la cognition ».

Source : Tricot (2006)

Cartographie de l'information- Introduction

La cartographie

La cartographie de données abstraites est assimilable à un processus qui permet de passer de données à une carte où les données peuvent décrire des connaissances.



Cartographie de l'information- Introduction

3- Ancrage historique : scientométrie et Leximappe

Ancrage

Début des années 1960 / Solla Price / revue *Scientometrics* créée en 1978

Objectif

Trouver des lois régissant la production scientifique (cf. création par Garfield de l'ISI (Institute for Scientific information), banque de données *Citation Index*).

Méthode

Appliquer des méthodes statistiques aux données secondaires des publications (mesures bibliométriques)

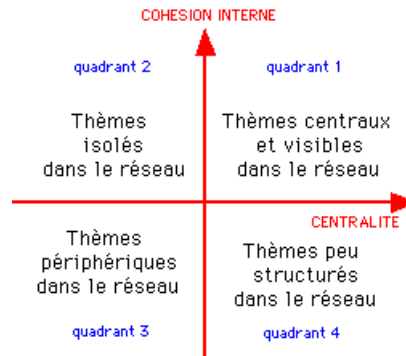
- unidimensionnelles : mesures statistiques simples (nombre d'articles ou brevets produits par un laboratoire, nombre de recherches menées en collaboration, etc), comptage des citations (indicateur de l'impact d'un document sur l'ensemble de la communauté scientifique);
- bi-dimensionnelles : permettent d'établir les liens pouvant exister entre deux publications, par exemple lorsqu'elles citent toutes les deux les mêmes auteurs, les mêmes brevets, etc. (méthode de co-citations ou de co-occurrence de citations) ou qu'elles recourent au même lexique (méthode des mots associés).

Résultat

Des cartes (outils graphiques associés).

Cartographie de l'information- Introduction

Leximappe (développé conjointement par le Centre de Documentation Scientifique et Technique du CNRS, aujourd'hui INIST-CNRS et le Centre de Sociologie et Innovation de l'Ecole Nationale Supérieure des Mines de Paris).



Systeme d'organisation de corpus documentaire, fondé sur la méthode des mots associés :

- Identifier les mots les plus fortement associés entre eux -> thèmes de recherche
- Construction d'agrégats (classes) indexant un sous-ensemble du corpus initial, représentant un pôle d'intérêt du domaine étudié.
- Position des pôles dans un **diagramme "stratégique"** suivant deux variables : la *centralité* souligne le pouvoir de structuration (un agrégat est structurant dans un domaine quand il est relié à d'autres pôles) ; la *densité* reflète la cohérence interne du pôle (plus l'association entre descripteurs est forte, plus le pôle est dense).

Cartographie de l'information- Introduction

Méthode des mots associés

Principe

Si deux documents sont proches parce qu'ils sont indexés par des mots-clés similaires, alors deux mots-clés figurant ensemble dans un grand nombre de documents seront considérés comme proches.

Indice d'équivalence : égal à la cooccurrence au carré des mots-clés i et j , divisée par le produit de leurs fréquences respectives.

$$E_{ij} = C_{ij}^2 / (C_i \times C_j)$$

L'indice d'équivalence évalue la force du lien : lorsqu'il vaut 0, les mots ne sont jamais présents ensemble ; s'il vaut 1, ils apparaissent toujours ensemble, on dit qu'ils sont équivalents.

Classification Ascendante Hiérarchique (CAH), dit du simple lien (« single link clustering »)

- Une classe est constituée de mots associés les uns aux autres par des associations internes.
- Les classes peuvent également avoir des relations entre elles par le biais d'associations externes entre des paires de mots appartenant à deux classes distinctes.
- L'étape de classification des mots-clés achevée, les documents les contenant peuvent finalement être affectés aux classes.
- Les classes sont ensuite positionnées sur un plan cartésien selon les valeurs de leurs « densité » et « centralité », constituant ainsi une « carte » :
 - la densité d'une classe, rapportée sur l'axe des ordonnées, est exprimée par la valeur moyenne des associations internes entre mots-clés formant la classe ;
 - la centralité d'une classe, rapportée sur l'axe des abscisses, est exprimée par la valeur moyenne des associations externes entre les mots qui la constituent et les mots d'autres classes.

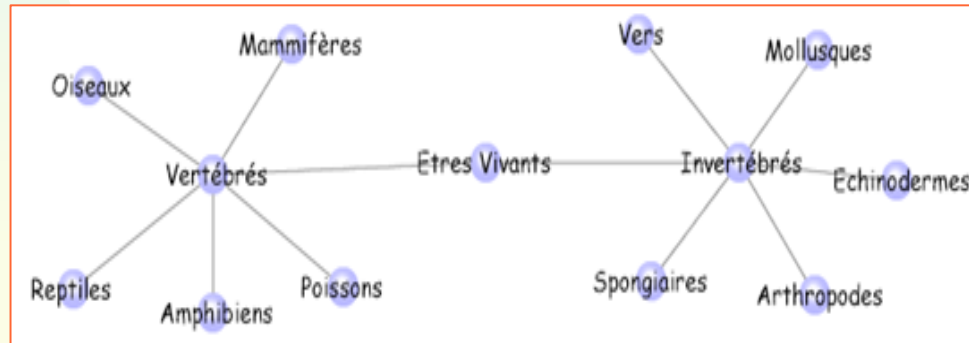
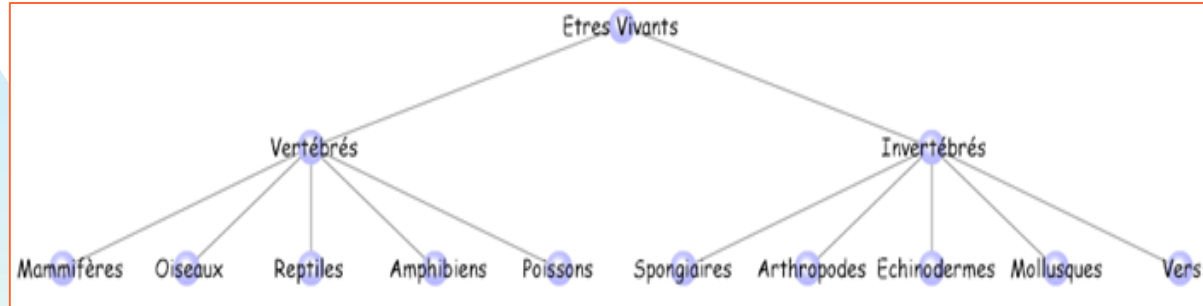
4- Problématique de la cartographie de l'information

- Les données abstraites ne possèdent pas de représentation graphique intrinsèque
 - Pour Card, Mackinlay et Shneiderman, tant que les informations et les données que l'on souhaite représenter sont dérivées de données physiques, alors elles possèdent une représentation graphique intrinsèque.
 - Cas d'une carte géographique qui représente des villes : la localisation physique des villes sur le globe terrestre fournit la position relative des points correspondants aux villes sur la carte. Les représentations graphiques sont alors « à l'image » des phénomènes observés.
- Pour les données abstraites, la difficulté est de leur trouver une représentation et plus particulièrement une répartition dans l'espace de la carte
 - Cas d'une cartographie de concepts : la difficulté n'est pas de trouver une représentation pour chaque concept (par exemple un rectangle ou un cercle) mais bien de les répartir dans l'espace de la carte sachant que la répartition aura un impact important sur la signification perçue.

L'enjeu de la cartographie de données abstraites est de déterminer le choix des représentations pour les visualiser

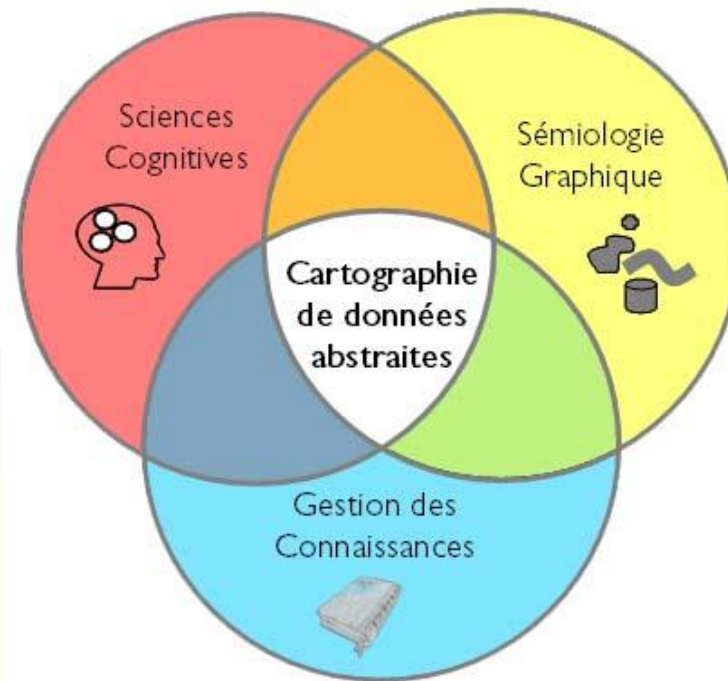
Cartographie de l'information- Introduction

4- Problématique de la cartographie de l'information



Cartographie de l'information- Introduction

5- Ancrage disciplinaire : sémiologie graphique, sciences cognitives, gestion des connaissances (statistiques et intelligence artificielle)



1- Composante calculatoire / **Data Mining**

Terminologie

Forage de données -> fouille de données -> exploration de données -> extraction de connaissances à partir de données (ECD)

Définitions

- Scientifique

Processus non-trivial d'identification de structures inconnues, valides et potentiellement exploitables dans les bases de données (Fayyad,1996).

- Méthodologique

Algorithmes et méthodes destinés à l'exploration et l'analyse de grandes bases de données informatiques, en vue de détecter dans ces données des règles, des associations, des tendances **inconnues** (non fixées *a priori*), des structures particulières restituant de façon concise l'essentiel de l'information utile pour l'aide à la décision.

1- Composante calculatoire / Data Mining

Spécificités

Se distingue de l'analyse de données et de la statistique par les points suivants :

- ne nécessite pas d'hypothèse de départ : c'est des données elles-mêmes que se dégageront les corrélations intéressantes (exemples historiques des tickets de caisse)
- les connaissances extraites par le Data Mining ont vocation à être intégrées dans un schéma organisationnel
- les données traitées sont issues des systèmes de stockage en place dans l'organisation

-> Le data mining fait passer d'analyses confirmatoires à des analyses exploratoires.

Techniques

- **Les techniques descriptives** : visent à mettre en évidence des informations présentes mais cachées par le volume des données : cas des recherches d'associations -> il n'y a pas de variable « cible » à prédire.
- **Les techniques prédictives** : visent à extrapoler de nouvelles informations à partir des informations présentes (cas du scoring) -> il y a une variable « cible » à prédire.

Cartographie de l'information- Composantes des applications cartographiques

1- Composante calculatoire / Data Mining

Méthodes	Définition	Usages	Techniques (liste indicative)
Non supervisées	Sans variable cible	Pour dégager d'un ensemble d'individus des groupes homogènes Ex. : typologie	<ul style="list-style-type: none">• Techniques à base de réseau de neurones• Classification ascendante hiérarchique• Recherche d'associations
Supervisées	Avec variable cible	Pour expliquer et/ou prévoir un ou plusieurs phénomènes observables et effectivement mesurés. Ex. : filtrage collaboratif, classifications	<ul style="list-style-type: none">• Techniques à base d'arbres de décision• Techniques à base de réseau de neurones
Réduction de données	Utilisée en amont des autres méthodes	Permettent de réduire un ensemble de données volumineux à un ensemble de taille plus réduite, épuré de ce que l'on considérera comme de l'information non pertinente ou non signifiante, comme du bruit.	<ul style="list-style-type: none">• Techniques d'analyse factorielle, type AFC

1- Composante calculatoire / **Data Mining : exemples**

Classification : Affecter un objet à une classe en fonction de ses caractéristiques A_1, \dots, A_n

Exemples

Déterminer si un message est un mail de SPAM ou non (2 classes)

Affecter une page web dans une des catégories thématiques d'un annuaire (multi-classes)

Méthode

Si pas de théorie pour définir la classe en fonction de A_1, \dots, A_n alors on étudie un ensemble d'exemples pour lesquels on connaît A_1, \dots, A_n et la classe associée et on construit un modèle.

$$\text{Classe} = f(A_1, \dots, A_n)$$

Techniques :

Analyse discriminante

Arbre de classification

Réseaux de neurones

1- Composante calculatoire / Réseau de neurones

Principe

- Un neurone est une unité de calcul élémentaire : combine des entrées x_1, \dots, x_n en une sortie o .
- Les entrées n'ont pas toutes la même importance : à chaque entrée x_i est associé un poids (ou coefficient synaptique) w_i .
- L'unité calcule d'abord l'activité d'entrée (somme pondérée des entrées)

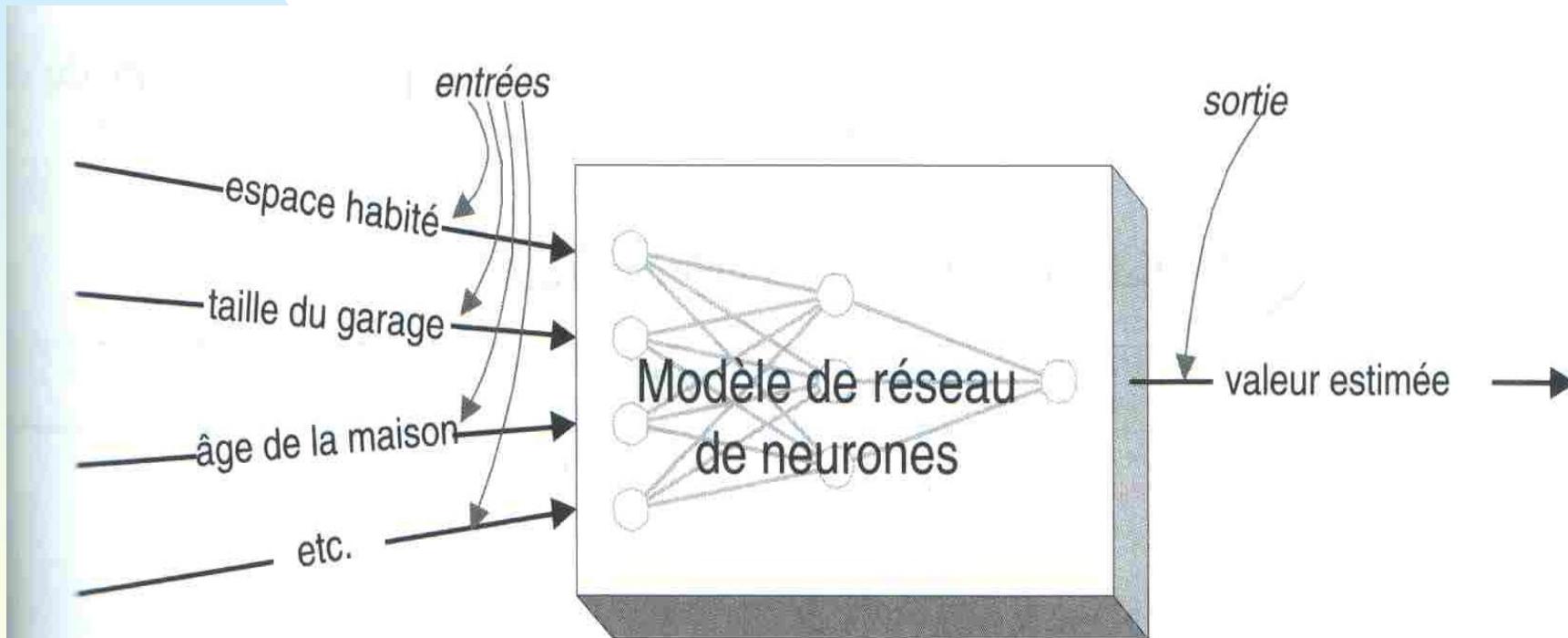
Coefficients synaptiques

- Sont calculés non *a priori* mais sur la base des valeurs disponibles : principe fondamental de l'apprentissage.
- Apprendre, c'est calculer les valeurs des coefficients synaptiques en fonction des exemples disponibles.
- Algorithme « d'entraînement » : modifie les poids synaptiques en fonction d'un jeu de données présentée en entrée du réseau.

Généraliser

L'intérêt des réseaux de neurones réside dans leur capacité à généraliser à partir du jeu de test.

Illustration du réseau de neurones



Un modèle de réseau de neurones calcule la valeur estimée (la sortie) à partir des entrées. Le calcul est un processus complexe qui produit les valeurs estimées sans que l'on doive comprendre son fonctionnement.

1- Composante calculatoire / **Text Mining**

Définition

Procédé consistant à synthétiser (classer, structurer, résumer, ...) les textes en analysant les relations, les patterns et les règles entre unités textuelles (mots, groupes, phrases, documents).

Du Data Mining au Text Mining

Ajout de la lexicométrie

Comme en Data Mining, on trouve en Text Mining :

- des algorithmes descriptifs : recherche des thèmes abordés dans un ensemble (corpus) de documents, sans connaître à l'avance ces thèmes;
- des algorithmes prédictifs : recherche des règles permettant d'affecter automatiquement un document à un thème, parmi plusieurs thèmes prédéfinis.

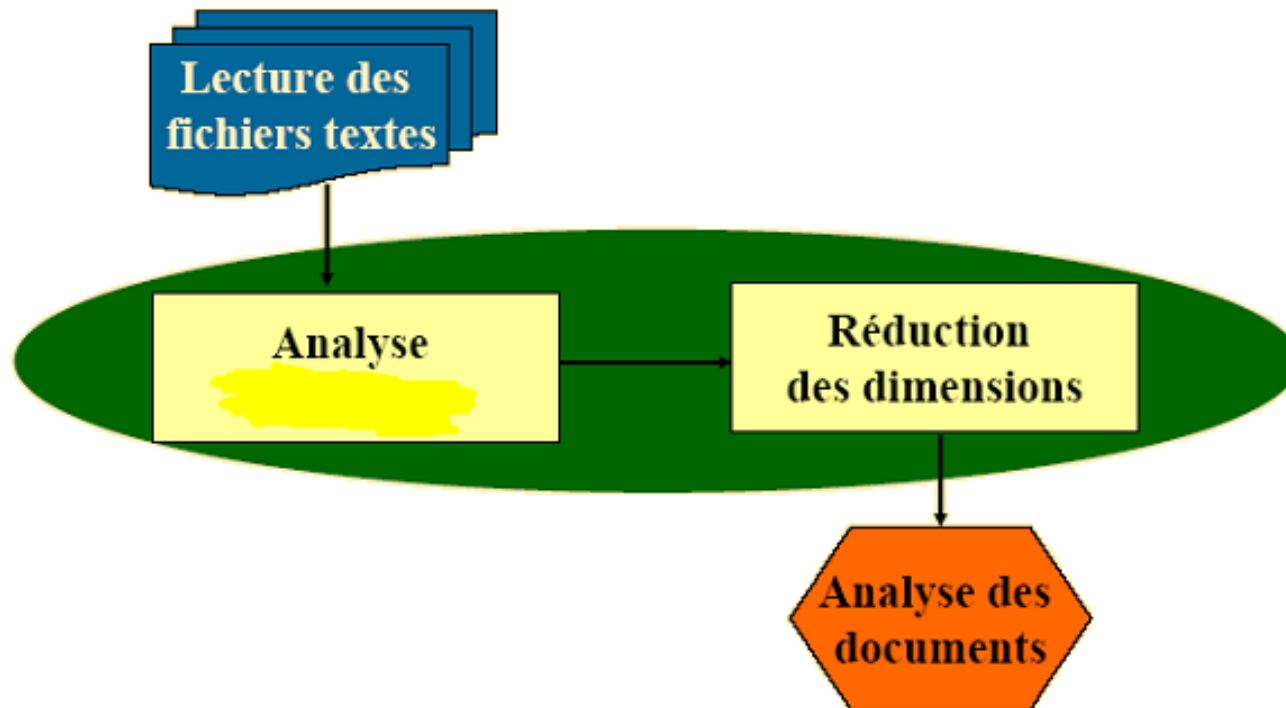
Application des techniques de Data mining :

- Individus = documents
- Caractères des individus = thèmes/termes des documents

Text Mining / Data Mining

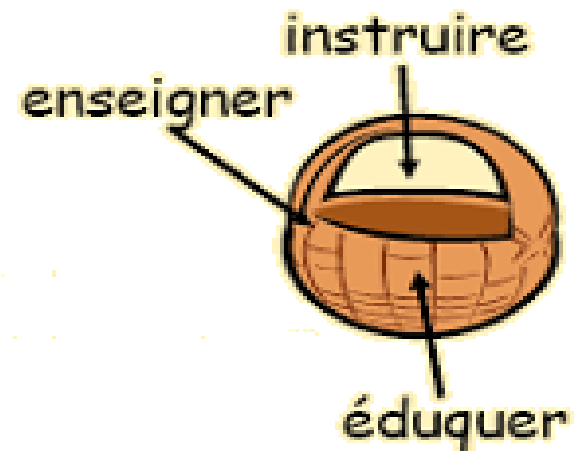
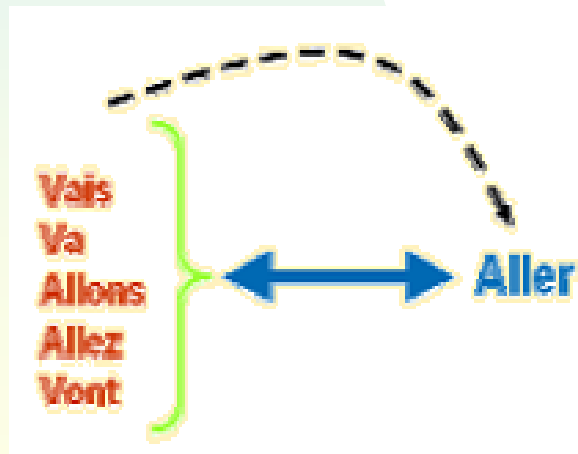
	Data Mining	Text Mining
Objet	numérique & catégorique	textuel
Structure	structuré	non-structuré
Représentation	simple	complexe
Dimension	milliers	milliers
Maturité	Implémentation vaste dès 1994	Implémentation vaste dès 2000

Vue simplifiée du Text Mining



Analyse lexicométrique

- Eliminer les mots vides
- Découper les textes en unités
- Lemmatiser
- Réduire le nombre de termes à traiter



Analyse statistique

Table de fréquences d'apparition des termes

	Mot 1	Mot 2	Mot 3	Mot 4
Doc 1	1	2	0	3
Doc 2	5	1	1	5
Doc 3	2	0	0	2

Term frequency (TF)

- Un terme qui apparaît plusieurs fois dans un document est plus important qu'un terme qui apparaît une seule fois
- w_{ij} = Nombre d'occurrences du terme t_i dans le document d_j
- TF_{ij} = Fréquence du terme t_i dans le document d_j

$$TF_{ij} = \frac{w_{ij}}{|d_j|}$$

Inverse document frequency (IDF)

- Un terme qui apparaît dans peu de documents est un meilleur discriminant qu'un terme qui apparaît dans tous les documents
 - df_i = nombre de documents contenant le terme t_i
 - d = nombre de documents du corpus
- Inverse document frequency

$$IDF_i = \log \frac{d}{df_i}$$

Pondération TF-IDF

- TF-IDF signifie Term Frequency x Inverse Document Frequency :
 - Proposée par [Salton 1989], mesure l'importance d'un terme dans un document relativement à l'ensemble des documents.

$$w_{i,j} = tf_{i,j} \times \log \left(\frac{N}{df_i} \right)$$

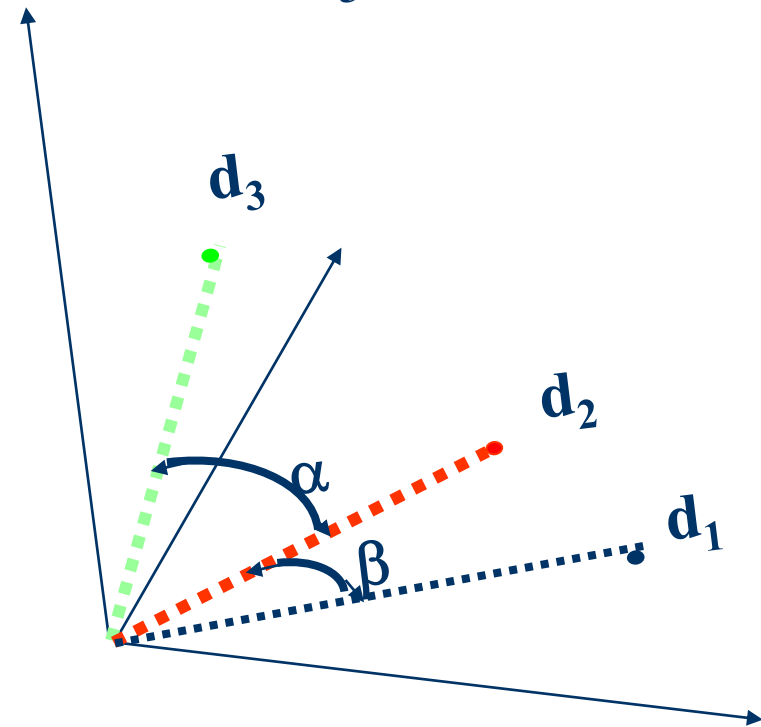
- $tf_{i,j}$ = fréquence du terme i dans le document j
 df_i = nombre de documents contenant le terme i
 $N = d$ = nombre de documents du corpus

Similarité entre documents

- Permet de ranger les documents par pertinence
- Le cosinus de l'angle est souvent utilisé

$$\cos(d1, d2) = \frac{d1^T \cdot d2}{\|d1\| \cdot \|d2\|}$$

- $\alpha > \beta \rightarrow \cos(\alpha) < \cos(\beta)$
- d_2 est plus proche de d_1 que de d_3



1- Composante calculatoire / **Data Clustering** ou partitionnement de données

Définition

Méthode statistique d'analyse des données qui a pour but de regrouper un ensemble de données en différents paquets homogènes : chaque sous-ensemble partage des caractéristiques communes, qui correspondent le plus souvent à des critères de proximité que l'on définit en introduisant des mesures de distance.

Principe

Pour obtenir un bon partitionnement, il convient de :

- minimiser l'inertie intra-classe pour obtenir des grappes (*cluster* en anglais) les plus homogènes possibles.
- maximiser l'inertie inter-classe afin d'obtenir des sous-ensembles bien différenciés.

Composition

Un cluster est composé de :

- une liste de mots-clés,
- une liste d'associations internes,
- une liste d'associations externes,
- une étiquette,
- une liste de documents affectés après la classification.

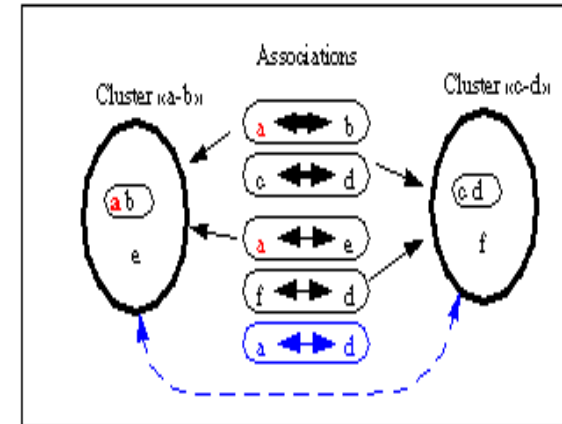
1- Composante calculatoire / DataClustering

List of Key-Words

[6] Milou
[6] Tintin
[3] Afrique
[7] montagne
[5] mouton
[5] berger
[3] touriste
[3] gangster

Internal Relationships

[6] Milou - Tintin
[3] Afrique - Tintin
[3] Afrique - Milou
[2] Tintin - montagne



- L'association la plus forte permet de créer le premier cluster : ici «a-b»).
- Une nouvelle association peut donner lieu à un «enrichissement» de cluster : «a-e» vient enrichir le cluster «a-b» qui contient déjà «a») ou à une création de nouveau cluster («c-d» par exemple).

1- Composante calculatoire / Web Mining

Définitions

Web mining : Data mining appliqué aux données de navigation sur le web

Objectifs du web mining

1. Optimiser la navigation dans un site web (confort des internautes),
2. Augmenter le nombre de pages consultées et l'impact des liens et des bannières publicitaires
3. Déceler les centres d'intérêt, et donc les attentes, des internautes visitant le site.

Types d'Analyses

- Statistique descriptive
 - « 70% des internautes ont consulté 3 pages ou moins »
 - « 40% des internautes accèdent au site sans passer par la page d'accueil »
- Détection des règles d'association
 - « 20% des internautes visitant la page A visitent la page B dans la même session »
- Segmentation des internautes
 - Selon les sites de provenance, les pages d'entrée, le nombre de pages consultées, les fichiers téléchargés, les pages de sortie, etc.

1- Composante calculatoire / Web Social Data Mining

Hypothèse documentaire

Web structuré par des « localités thématiques » définies par une forte corrélation entre contenu des documents et proximité hypertexte, dessinant comme un territoire.

Problématique de la « traçabilité » du social

- Le web se présente comme une *matrice technique distribuée* façonnée par des *usages* dont le réseau *garde la mémoire*.
- Le seul fait d'utiliser une technologie numérique contribue à produire des *traces* et, à grande échelle, elles constituent un réservoir immense de données disponibles.

Principes : hubs et autorités (Kleinberg et al.)

- Autorités : ensemble des pages contenant les meilleures sources d'information pour un thème donné.
- Hubs : pages pointant sur les autorités.
- Le calcul : ressemble au calcul d'importance d'un article par nombre de citations (les liens du Web vus comme des recommandations).

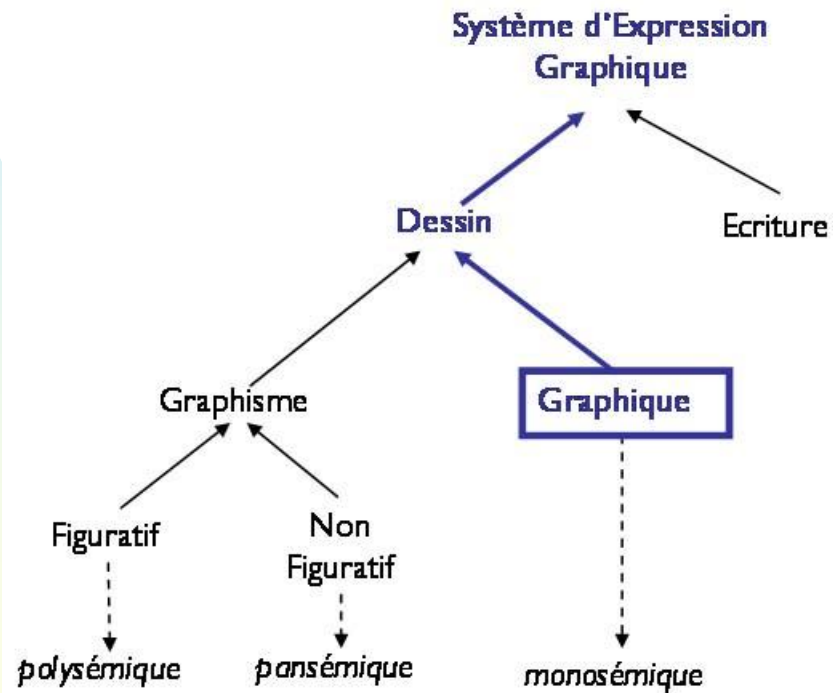
Applications

Sujets controversés où le seul lexique n'est pas discriminant.

2- Composante visuelle

- Éléments de sémiologie graphique (J. Bertin)
- Théorie de la Gestalt

2- Composante visuelle / Eléments de sémiologie graphique (J. Bertin)



2- Composante visuelle / Eléments de sémiologie graphique (J. Bertin)

Type de dessin		Signification des signes	Description	Exemple
Graphisme	Image non-figurative	Pansémique	Le système s'ouvre à toute signification.	Tableau d'art
	Image figurative	Polysémique	Le système a pour objectif de définir un concept ou une idée mais les interprétations peuvent diverger.	Photographie aérienne.
Graphique		Monosémique	Transcriptions de relations entre des concepts préalablement définis.	Un organigramme

2- Composante visuelle / Éléments de sémiologie graphique (J. Bertin)

« La graphique est un système de signes [structures visuelles] qui permet de **transcrire** les relations de différence, d'ordre ou de proportionnalité **existant** entre des données qualitatives ou quantitatives » [Bonin, 1997].

« La carte est une image graphique qui *transcrit* les relations de *proportionnalité* : ordre et différence, par des variations proportionnelles **ordonnées** et **différentielles** ».

2- Composante visuelle / Eléments de sémiologie graphique (J. Bertin)

LA CARTE est une **IMAGE GRAPHIQUE** qui transcrit :

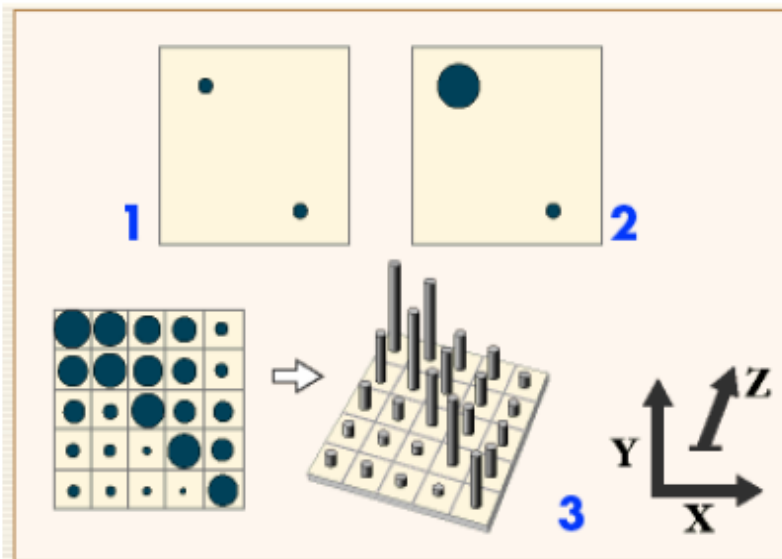
les relations de	proportionnalité ordre différence	par des variations	proportionnelles ordonnées différentielles
------------------	---	--------------------	--

LE SYSTÈME GRAPHIQUE est constitué par :

Types d'implantation	POINT LIGNE ZONE			
<u>Variations visuelles</u>	X, Y TAILLE VALEUR GRAIN COULEUR ORIENTATION FORME			
Propriétés des variables visuelles	proportionnelle ordonnée sélective associative dissociative	permettant de transcrire	la proportionnalité l'ordre la différence	des données

Atelier de cartographie de Sciences Po (Paris, France), octobre 2007
<http://cartographie.sciences-po.fr>

2- Composante visuelle / Eléments de sémiologie graphique (J. Bertin)



Les trois dimensions de l'image instantanée

1. Dans le plan, une tache peut être en haut ou en bas, à droite ou à gauche. La perception construit deux dimensions indépendantes **X** et **Y**, séparées par la perpendicularité.
2. Une variation d'énergie lumineuse construit en **Z** une 3e dimension indépendante de X et Y.

L'image, forme significative perçue instantanément se crée sur trois dimensions indépendantes **x, y, z** (3).

Elle peut donc transcrire les relations entre trois ensembles indépendants.

La variation en Z de l'énergie lumineuse, sur un support papier est fournie par la variation de la taille ou de la valeur des taches. **Taille** et **valeur** des taches sont, avec le plan **X, Y**, les **variables visuelles de l'image**.

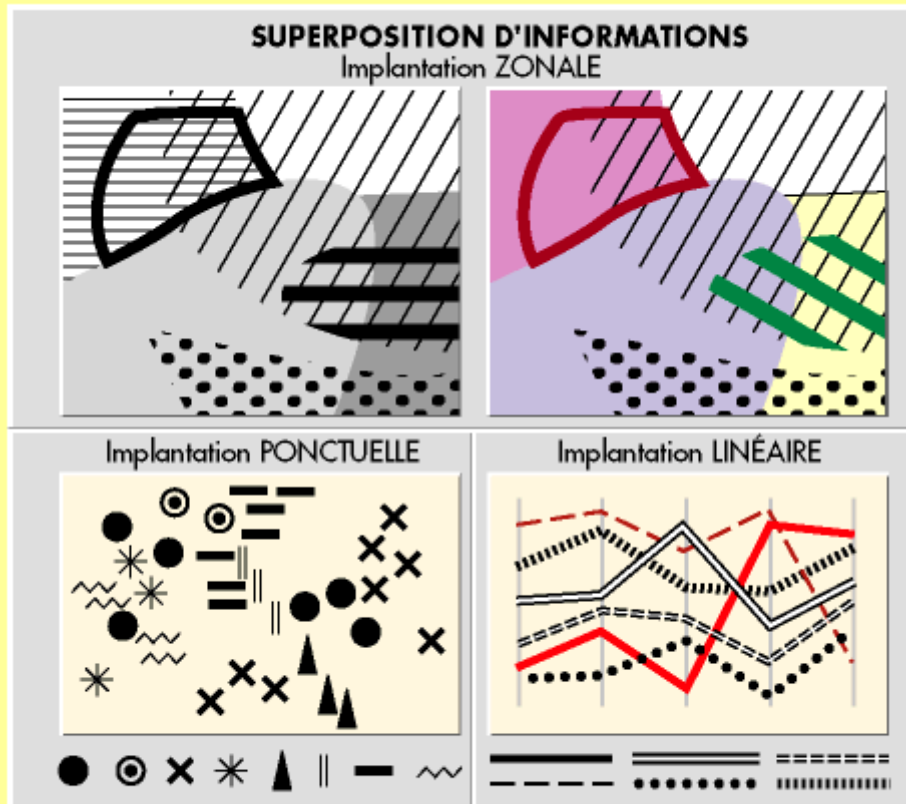
2- Composante visuelle / Eléments de sémiologie graphique (J. Bertin)

LE SYSTÈME GRAPHIQUE

LE SYSTÈME GRAPHIQUE considère 8 variables visuelles qui peuvent s'appliquer à 3 implantations. Elles possèdent 5 propriétés permettant de transcrire 3 relations entre les données.

8 VARIABLES VISUELLES	3 IMPLANTATIONS	5 PROPRIÉTÉS	3 RELATIONS
X et Y (les deux dimensions du plan) TAILLE VALEUR GRAIN COULEUR ORIENTATION FORME	point P ligne L zone Z	QUANTITATIVE ORDONNÉE SÉLECTIVE DISSOCIATIVE (visibilité variable) ASSOCIATIVE (visibilité constante)	PROPORTIONNALITÉ ORDRE RESSEMBLANCE

2- Composante visuelle / Eléments de sémiologie graphique (J. Bertin)



Des zones contenant des informations différentes peuvent se superposer en utilisant des variations de **taille, grain, couleur, orientation... associées et très contrastées.**

Chaque zone doit garder son autonomie visuelle et ne pas être confondue avec les autres : **la sélectivité doit être maximale.**

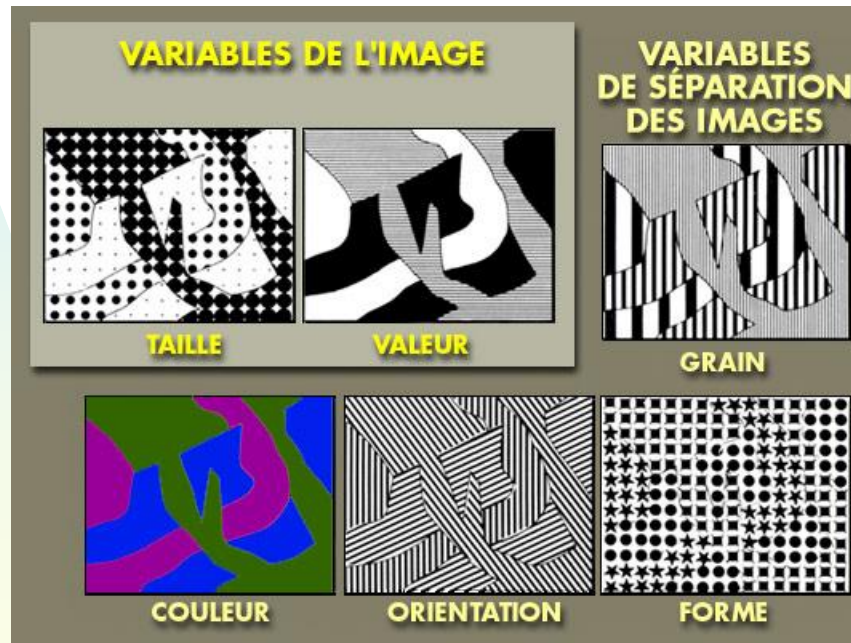
On obtient ainsi un effet de transparence.

Le choix de variables visuelles adaptées permet de créer des hiérarchies ou des relations de différence / ressemblance entre les éléments représentés.

2- Composante visuelle / Éléments de sémiologie graphique (J. Bertin)

Variables visuelles (« variables rétiniennes »)

Variation des propriétés graphiques des structures visuelles



2- Composante visuelle / Éléments de sémiologie graphique (J. Bertin)

Variables visuelles

		LES VARIABLES DE L'IMAGE				PROPRIÉTÉS							
		POINTS		LIGNES		ZONES							
XY	2 DIMENSIONS DU PLAN	x	x	x	/	~	/	14-15-9	18-2	Q	O	≠	≡
Z	TAILLE	█	█	█	/	~	/	16-21-2	1-21-15	Q	O	≠	≡
	VALEUR	█	█	█	/	~	/	14-15-1	1-2-9	O	≠	≡	
		LES VARIABLES DE SÉPARATION DES IMAGES											
	GRAIN	█	█	█	/	~	/	14-15-1	1-2-9	O	≠	≡	
	COULEUR	█	█	█	/	~	/	14-15-1	1-2-9	≠	≡		
	ORIENTATION	█	█	█	/	~	/	14-15-1	1-2-9	≠	≡		
	FORME	█	█	█	/	~	/	14-15-1	1-2-9	≠	≡		

D'après : Jacques Bertin

LES VARIABLES DE L'IMAGE construisent des figures sur un fond. Elles sont sélectives.

LES VARIABLES DE SÉPARATION construisent des images sans relief. Elles donnent à chaque caractère la même visibilité. Elles sont associatives.

Sélectivité : permet de sélectionner un caractère et de faire abstraction visuelle de tout le reste.

Associativité : Visibilité constante.

Dissociativité : Visibilité variable.

Proportionnalité : Q

Ordre : O

Associativité : ≡

Dissociativité : ≠

Sélectivité : ≡

Cartographie de l'information- Composantes des applications cartographiques

2- Composante visuelle / Eléments de sémiologie graphique (J. Bertin)

TYPE DE DONNÉES	RELATION ENTRE LES DONNÉES	TRANSCRIPTION GRAPHIQUE	VARIABLES VISUELLES
Quantités absolues	proportionnalité	points lignes	taille
			taille
Quantités relatives	ordre	zones	valeur
			valeur
Données qualitatives	ordre / hiérarchie	points lignes zones	valeur
			taille, valeur, grain, couleur, orientation et forme associées
	différence / ressemblance	points lignes zones	taille, valeur, grain, couleur, orientation et forme associées
			taille, valeur, grain, couleur, orientation et forme associées

VARIABLES VISUELLES	EXEMPLES
taille	
valeur (clarté)	
grain	
couleur	
orientation	
forme (associée aux autres variables)	

2- Composante visuelle / Théorie de la Gestalt (1890)

Principe

Le contexte est très important dans la perception visuelle [Guillaume].

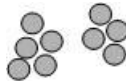


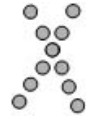

Les structures visuelles ne sont pas indépendantes les unes des autres et le contexte est important pour la signification qui leur est attribuée.

En graphique

La perception visuelle s'attache à reconnaître des **modèles** dans un ensemble de structures visuelles.

De cette théorie résultent des **lois** correspondant à des catégories d'interactions possibles entre les structures visuelles.

Cartographie de l'information- Composantes des applications cartographiques

Lois	Effets	Illustrations
<i>Prégnance</i>	Une image est facile à comprendre si sa structure est simple et inversement.	
<i>Proximité</i>	Deux composants qui sont proches ont tendance à être perçus comme un seul composant.	
<i>Similarité</i>	Les composants similaires sont perçus comme s'ils étaient regroupés.	
<i>Fermeture</i>	Les contours proches sont perçus comme unifiés.	
<i>Continuité</i>	Des éléments voisins sont perçus groupés lorsqu'ils possèdent potentiellement un trait qui les relie.	
<i>Symétrie</i>	Des éléments sont perçus comme un élément global lorsqu'ils forment une symétrie.	
<i>Trajectoire identique</i>	Des éléments qui se déplacent avec la même trajectoire semblent groupés.	
<i>Familiarité</i>	Des éléments sont plus facilement groupables si le groupe est familier ou significatif.	

2- Composante visuelle / Théorie de la Gestalt (1890)



Cartographie de l'information- Composantes des applications cartographiques

2- Composante visuelle / Conception d'un langage graphique (JB Lamy)

1- Analyser les informations ou connaissances que l'on souhaite représenter

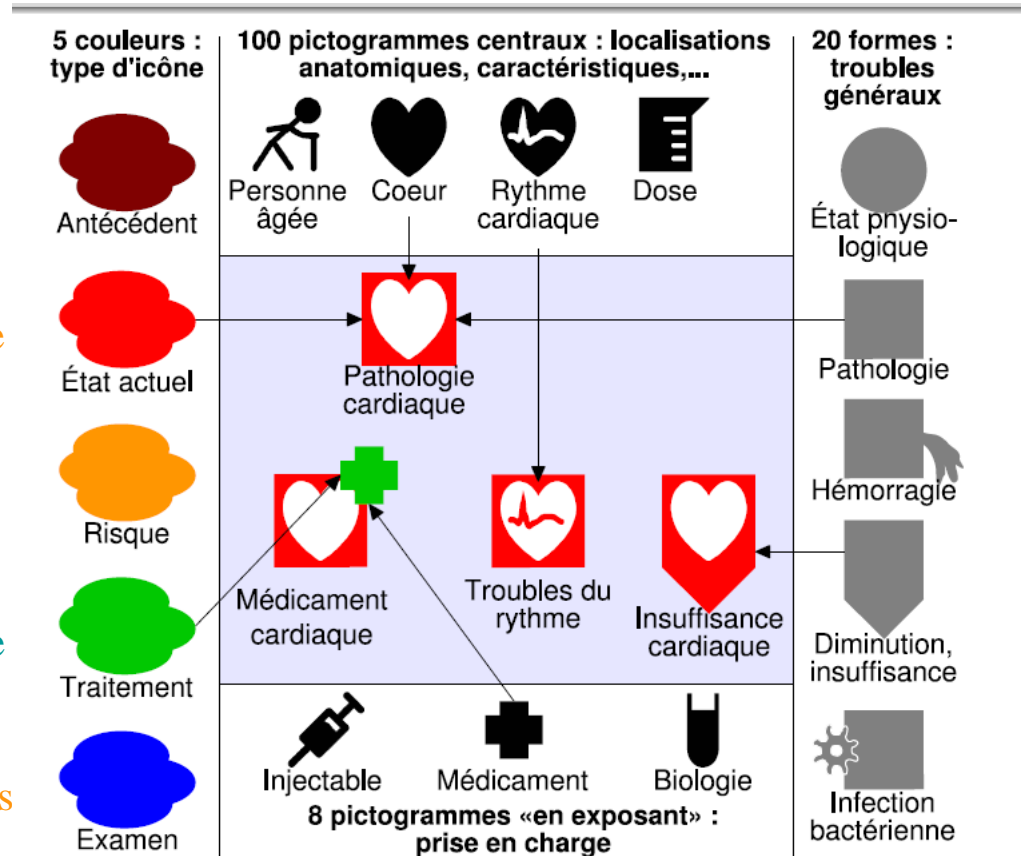
Determiner les attributs qui définissent les termes à représenter

- Par exemple pour représenter des maladies : les attributs pourraient être : la localisation anatomique (cardiaque, rénale,...), la sévérité, etc.

2- Associer à chaque attribut une variable rétinienne pour le représenter

- Par exemple pour les maladies : Associer l'attribut *localisation anatomique* à la variable forme/pictogramme
- Représenter la *sévérité* par la variable couleur

3- Les relations entre les différents termes représentés doivent être explicitées.



Famille d'outils et contextes applicatifs

1- Famille d'outils

- Typologies
- Aperçu
- Travaux pratiques

2- Contextes applicatifs

- Éléments d'évaluation
- Les perspectives de la cartographie sémantique

1- Famille d'outils

- Typologies
- Aperçu
- TP à partir de la sélection : http://delicious.com/mylk02/Stage_Carto
- Et encore : <http://www.visualcomplexity.com/vc/> (plus de 500 projets dans le domaine)

1- Famille d'outils : typologies

1.1- Typologie scientifique : établie selon des critères internes au domaine :

- les trois principaux paradigmes : technique de visualisation, type de données et type d'interaction (Keim DA 2002)
- la nature des données : unidimensionnelles, bidimensionnelles, tridimensionnelles, temporelles, multidimensionnelles, hiérarchiques, réseaux (Shneiderman B 1996)

1.2- Etat de l'art / évaluation - benchmarking : typologie orientée usages (veille et intelligence économique)

Quelques exemples :

- Inist : <http://outils.veille.inist.fr/>
- Esiee : <http://www.aa-esiee.com/event/2005/AG/ClaudeAschenbrenner.ppt>
- Ecole des Mines : <http://www.demoscience.org/resources/>
- Commercial : <http://www.k-praxis.com>
- Commercial : http://www.information-mining.info/state_of_art/
- Universitaire : <http://well-formed-data.net/thesis>

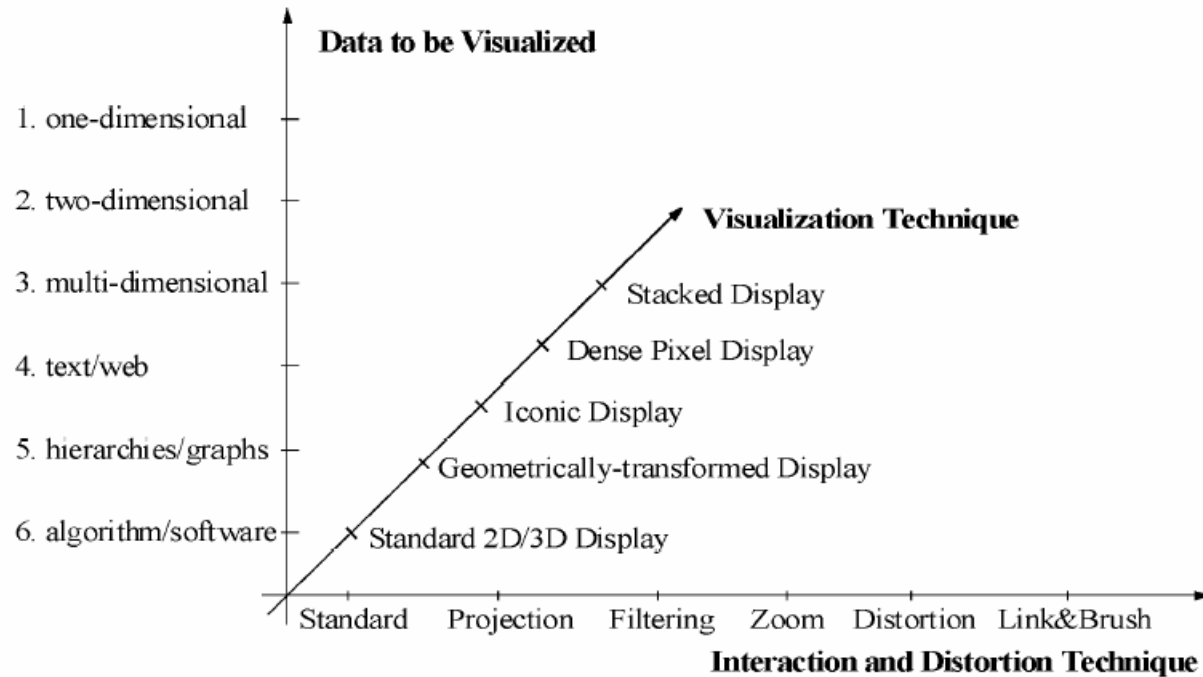
1.3- Typologie cartographique

Quelques exemples de bloggers :

- Claude Aschenbrenner : <http://www.serialmapper.com/archive/2007/01/10/la-pierre-de-rosette-de-la-cartographie-de-l-information.html>, avec son mode d'emploi
- Kartoo : <http://www.mapdream.com/>

Cartographie de l'information- Famille d'outils et contextes applicatifs

1- Famille d'outils : typologie scientifique



Cartographie de l'information- Famille d'outils orientés « bibliothèques »

2- Les outils de navigation et de recherche sur Internet (sélection de Pierre Nobis)

A- Les navigateurs : Nestor / Navicrawler

The screenshot displays a computer desktop with two windows. The left window, titled 'MapInfo: Map has been repaired', shows a map interface with various icons and a search bar. The right window, titled 'Browser', shows a Wikipedia page for 'Albert Lewin'. The browser window includes a search bar, navigation buttons, and a sidebar with 'Navigation' and 'Contribuer' sections. The main content area of the browser displays the article title 'Albert Lewin', a notice about the article being a draft, and a biographical summary.

Navigation

- Accueil
- Portails
- Thématiques
- Index alphabétique
- Un article au hasard
- Contacter Wikipédia

Contribuer

- Aide
- Communauté
- Modifications récentes
- Accueil des nouveaux arrivants
- Faire un don

Rechercher

Consulter

Albert Lewin

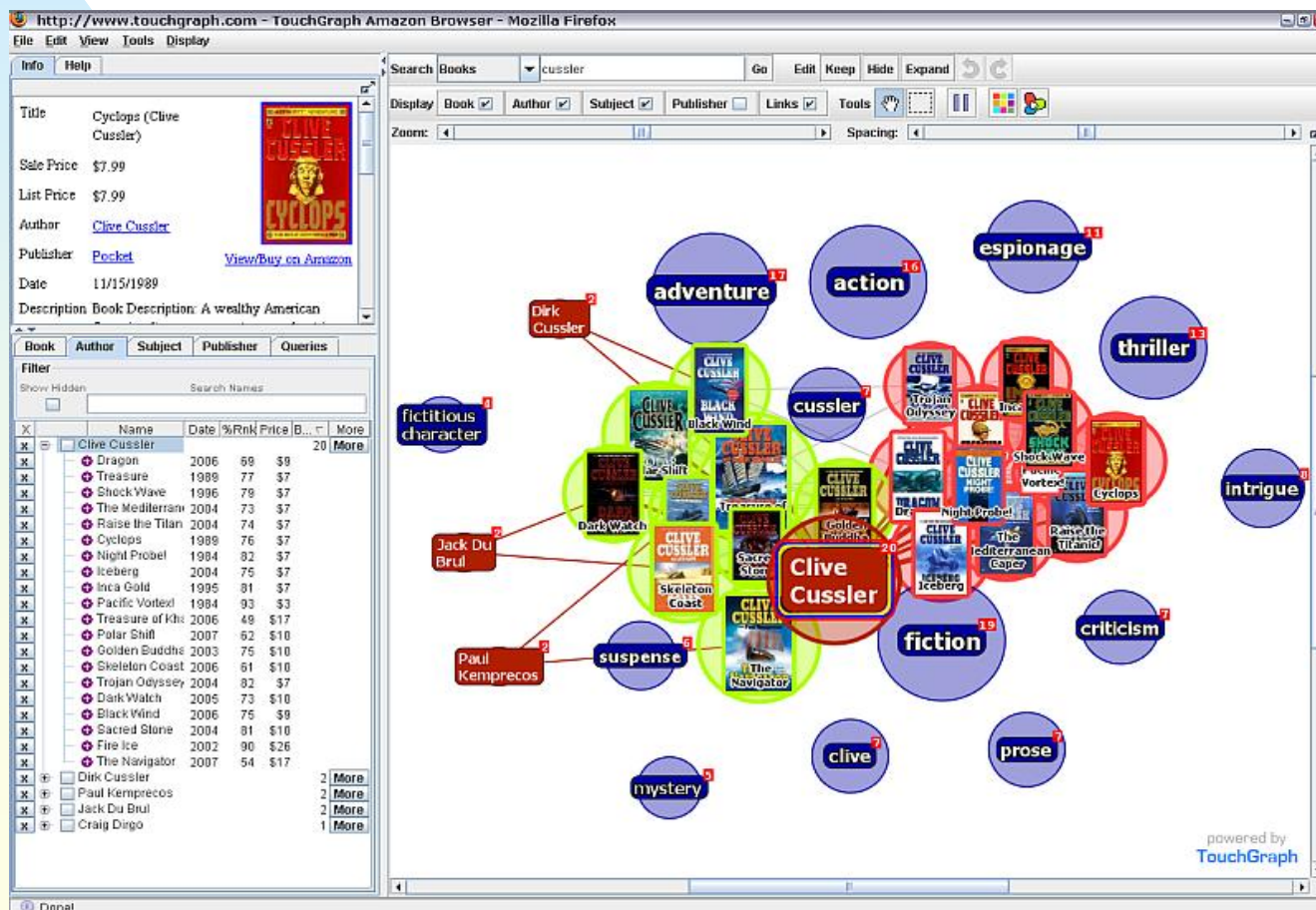
23 septembre 1894 à Newark, New Jersey, États-Unis - 9 mai 1968 (New York, États-Unis) est un scénariste, réalisateur et producteur américain. Il est connu pour avoir organisé un concours de

Naissance	23 septembre 1894 Newark États-Unis
Nationalité	Américain
Mort	9 mai 1968 New York États-Unis
Profession(s)	Réalisateur

Cartographie de l'information- Famille d'outils orientés « bibliothèques »

2- Les outils de navigation et de recherche sur Internet

B-Les outils de visualisation des résultats : TouchGraph, application Java permettant de cartographier les résultats issus notamment de : Google, Amazon, PubMed, Alexa, FaceBook.



Cartographie de l'information- Famille d'outils orientés « bibliothèques »

2- Les outils de navigation et de recherche sur Internet (sélection de Pierre Nobis)

C- Les outils de recherche avec ajout de termes associés dans une colonne, regroupement thématique, nuage de « tags »

The screenshot displays the ESC Lille library website. At the top, there is a navigation bar with links for 'Mon compte', 'Présentations', 'Communautés', 'Coach info', and 'Contactez-nous'. Below this is a search bar with the text 'Suggestions' and a 'Rechercher' button. The main content area features a tag cloud on the left with terms like 'internet', 'web 2.0', 'site', 'navigation', and 'commerce'. The central part of the page shows a list of search results for 'web 2.0', including titles like 'Web 2.0 (Révolutions et nouveaux services d'Internet)', 'Initiation à ASP 3.0', and 'L'Internet collaboratif, communautaire et social, vers la prise de pouvoir de internautes?'. Each result includes details such as location, year, subject, and occurrence. On the right side, there is a sidebar with filters for 'Choisissez un fonds', 'Disponibilité', 'Support', 'Année', 'Sujet', and 'Auteur'. The bottom of the page has a legend for 'Association', 'Traduction', and 'Chemin suivi'.

D- Métamoteurs avec affichage de clusters sous forme de listes :

Carrot2 / Clusty-Vivissimo / Iboogie/ Polymeta / Turbo10 /Webclust

Cartographie de l'information- Famille d'outils orientés « bibliothèques »

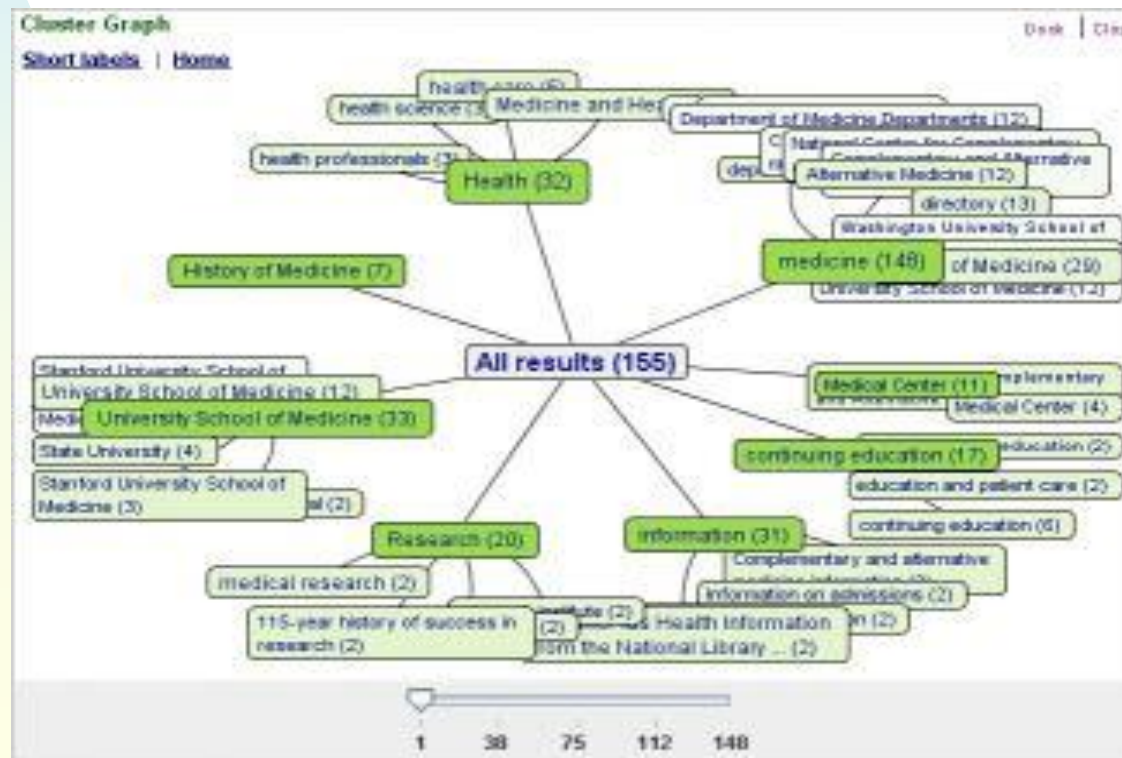
2- Les outils de navigation et de recherche sur Internet (sélection de Pierre Nobis)

E- Moteurs avec affichage de clusters sous forme graphique

E1- représentation centrée : Allplus / WebBrain

Allplus est un métamoteur à clustérisation graphique qui interroge les bases de *Ask*, *Google*, *Yahoo* et *Live* par le biais de l'outil *Polymeta*. L'interface "cluster graph" permet de visualiser les résultats sous une forme graphique.

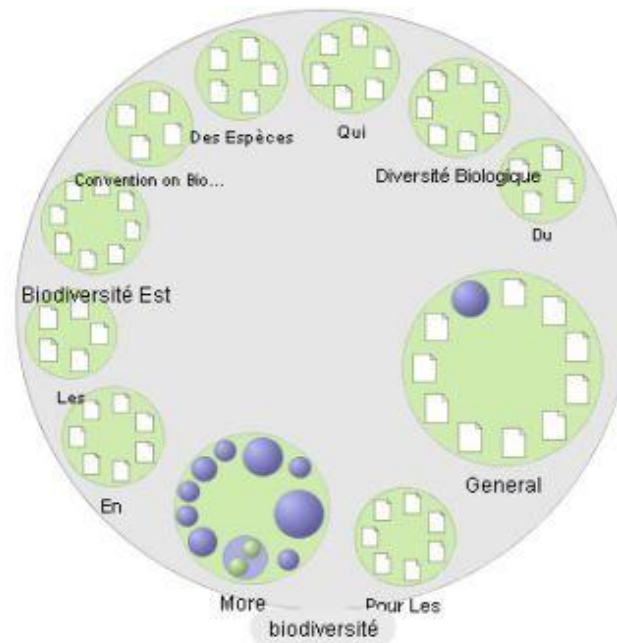
Webbrain permet de visualiser et d'affiner les résultats d'une requête issus de l'*Open Directory Project*.



2- Les outils de navigation et de recherche sur Internet (sélection de Pierre Nobis)

E- Moteurs avec affichage de clusters sous forme graphique

E2- Représentation circulaire : Grokker



E3- Représentation topographique : TooLeNet s'appuie sur les technologies de *KartOO* et sur l'index de *Yahoo* pour effectuer des recherches selon 3 modes : moteur, annuaire et carte auxquels s'ajoutent des options de personnalisation et de classification assez avancées.

4- Les outils de création (carto)graphique

A- Pour mémoire : les logiciels de création de cartes heuristiques et conceptuelles

- Cmaptools est un logiciel gratuit (usage éducatif) de création de cartes conceptuelles
- Freemind est un logiciel libre et gratuit qui permet de réaliser des cartes heuristiques.
- Thinkgraph est un logiciel gratuit de dessin 2D orienté vers la réalisation de cartes conceptuelles.



4- Les outils de création (carto)graphique

B- Les outils d'édition de graphes

Guess est un logiciel libre et gratuit développé par Eytan Adar qui permet de visualiser des graphes statiques. Options de modification de certains attributs (taille, couleur, forme, algorithmes de positionnement...).

Pajek est un logiciel libre et gratuit (usage non commercial) développé par V. Batagelj et A. Mrvar, qui permet de traiter des données de réseaux de grande taille. *Pajek* signifie "araignée" en slovène.

C- Les outils en ligne

ManyEye : service collaboratif développé par IBM grâce auquel on peut générer selon un des modèles proposés sa propre visualisation de l'information (après avoir téléversé ses propres données).

IssueCrawler : outil de web social mining, utilisable à distance (chargement des URL), récupération graphique.

2- Contextes applicatifs

- Éléments d'évaluation
- Les perspectives de la cartographie sémantique

2- Contextes applicatifs / Éléments d'évaluation

2.1- Evaluation scientifique

Question de méthode : *Évaluation des Interfaces Utilisateur d'Information* : Bonnel et Chevalier, 2006

Objectif : proposer une méthodologie d'évaluation permettant de comparer, du point de vue de la visualisation des résultats à une requête, les outils de recherche : Google, Vivisimo, Ujiko, Grokker, Kartoo, Vios, SmartWeb (prototype).

Evaluation des représentations graphiques : généralités (Tricot et Roche 2006) :

- Arborescences simples :
 - ✓ habitude des utilisateurs (explorateur de fichiers informatiques) : pas de problèmes avec volume limité,
 - ✓ au-delà d'une certaine quantité d'information, les arborescences deviennent inefficaces.
- Arbres de cônes en 3D :
 - ✓ posent des problèmes à l'utilisateur qui n'a pas l'habitude des espaces en 3D,
 - ✓ effort cognitif plus élevé et prise en main plus longue.
- Arbres hyperboliques :
 - ✓ problèmes des étiquettes des éléments (occlusion),
 - ✓ effet de zoom imprévisible.

L'une des conclusions : une seule visualisation ne permet pas de répondre à toutes les questions ... multiplier les possibilités de visualisation.

2- Contextes applicatifs / Eléments d'évaluation

2.1- Evaluation scientifique : doctorat en cours à l'EBSI (Aline Crédeville)

- Contexte : visualisation de l'information dans le contexte de recherche d'information
- Point de vue de la sémiologie graphique :
 - principes les plus couramment utilisés sont la proximité (Grokker), l'englobage (Kartoo) et la connectivité (Auquabrowser)
 - métaphores visuelles pas toujours adéquates au web
- Résultats :
 - l'information doit être structurée
 - apports possible pour l'exploration mais pas pour la recherche

2- Contextes applicatifs / Eléments d'évaluation

2.2- Evaluation en contexte d'usage de recherche d'informations (Ciaccia, Ihadjadene, Martins 2006)

Objectif : analyse de l'influence du mode de présentation des informations sur les processus cognitifs mis en jeu par l'utilisateur dans l'activité de recherche d'informations sur Internet

Outils évalués : Kartoo (pour 18 usagers) et Mapstan (pour 16 usagers).

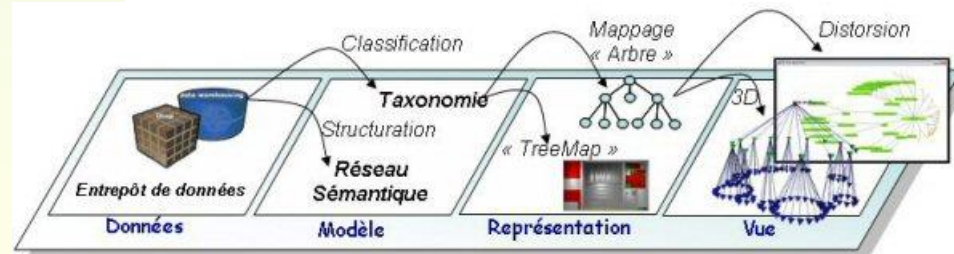
Principaux résultats :

- Coût cognitif, sans pour autant conduire à de meilleures performances (comparaison avec une présentation en liste).
Hypothèse : nouveauté de la présentation *versus* habitude des présentations verbales.
- Stratégie optimisée : préciser la requête d'une recherche à but flou et se concentrer sur les informations reliées à l'objectif de la question *versus* informations non reliées.
- Le sens de la visualisation doit être clarifié ; les systèmes doivent faciliter la compréhension des éléments visualisés.

Cartographie de l'information- Famille d'outils et contextes applicatifs

2- Contextes applicatifs / Les perspectives de la cartographie sémantique (C. Tricot : de la sémantique à la carte)

- Pour résoudre la problématique de la représentation de l'information, la visualisation de l'information seule ne suffit pas.
- La visualisation de l'information permet uniquement de présenter des données en leur associant une représentation graphique.
- La cartographie sémantique se veut :
 - ✓ Une méthode de cartographie, reposant sur la sémantique d'un domaine
 - ✓ Un processus de production de connaissances
données brutes -> données structurées -> données représentées -> données visualisées
 - ✓ Un formalisme de description, type Réseaux sémantiques
 - ✓ Un langage de description cartographique
 - ✓ Une architecture multi-agents



Cartographie de l'information- Famille d'outils et contextes applicatifs

2- Contextes applicatifs / Les perspectives de la cartographie sémantique (C. Tricot : de la sémantique à la carte)

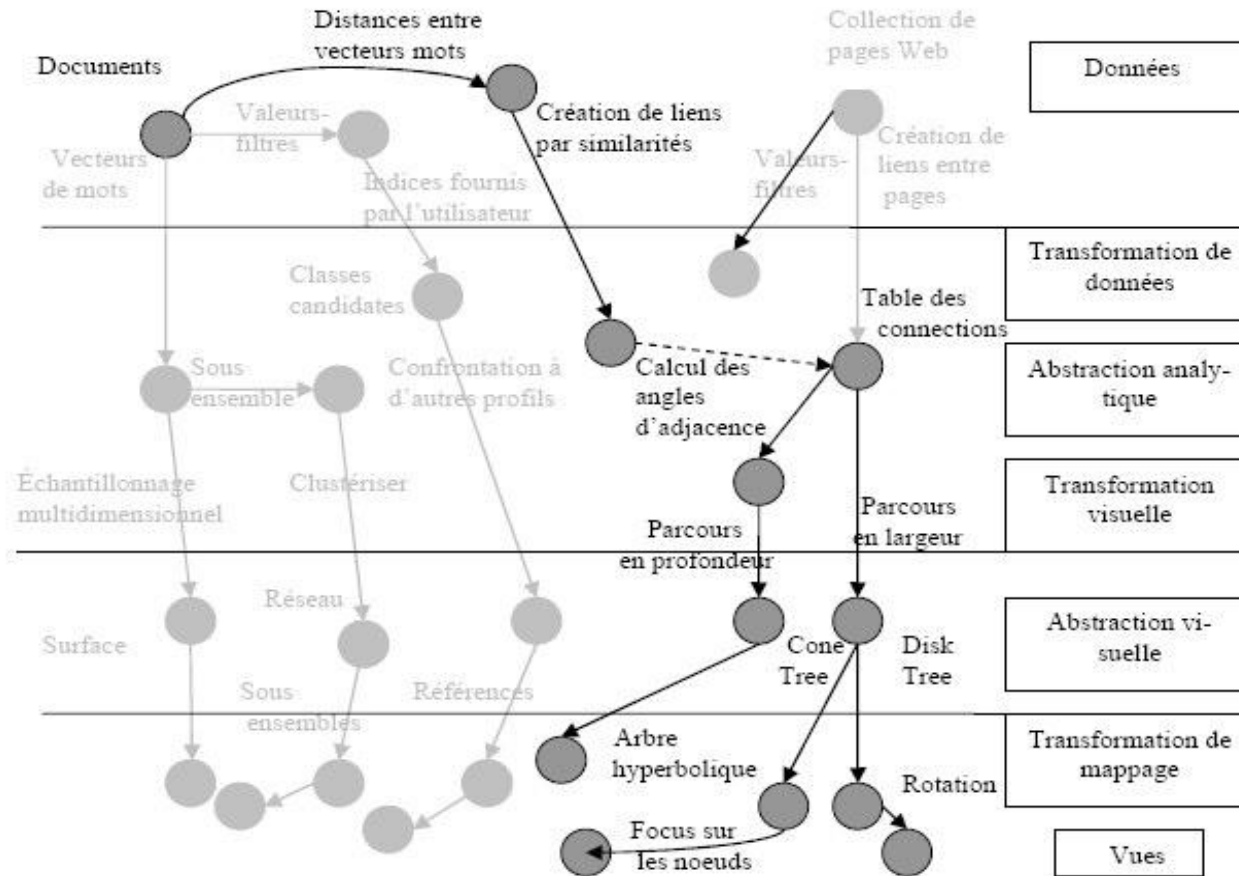


FIG.3- framework d'interactivité basé sur le modèle de Chi&Riedl model³

2- Contextes applicatifs / Les perspectives de la cartographie sémantique

“Une bonne structure représentative graphique pour un arbre de données est celle qui déforme le moins les similitudes originales entre les unités de données (similitudes qui sont donnés par l'outil de classification) ou en d'autres termes : deux individus (ou classes) qui sont proches l'une de l'autre le restent aussi possible que peut après projection dans l'espace de visualisation ; et ceci doit être évident (et vérifiable par une mesure de calcul de divergence) dans l'environnement de visualisation “, Tricot 2006.

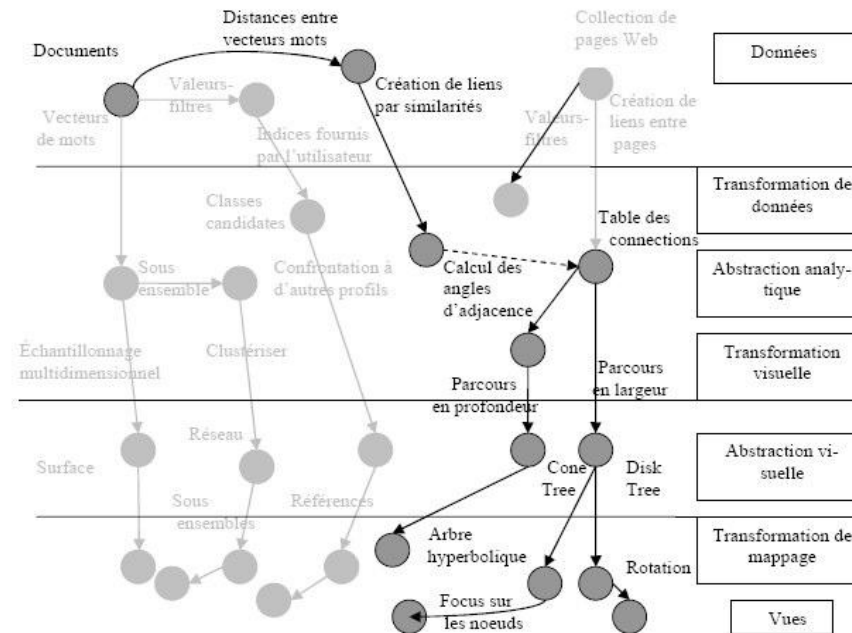


FIG.3- framework d'interactivité basé sur le modele de Chi&Riedl model³