

> Le Text Mining : une technologie Data Mining pour l'analyse approfondie des données textuelles

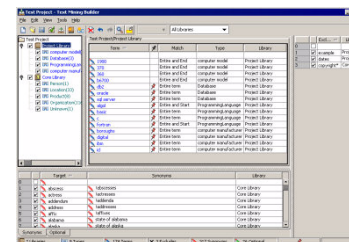


Text Extraction

Khalid EL HIMDI, Consultant Data Mining
Hamid MACHHOUR
SPSS Maghreb



Text Extraction



SPSS en quelques chiffres



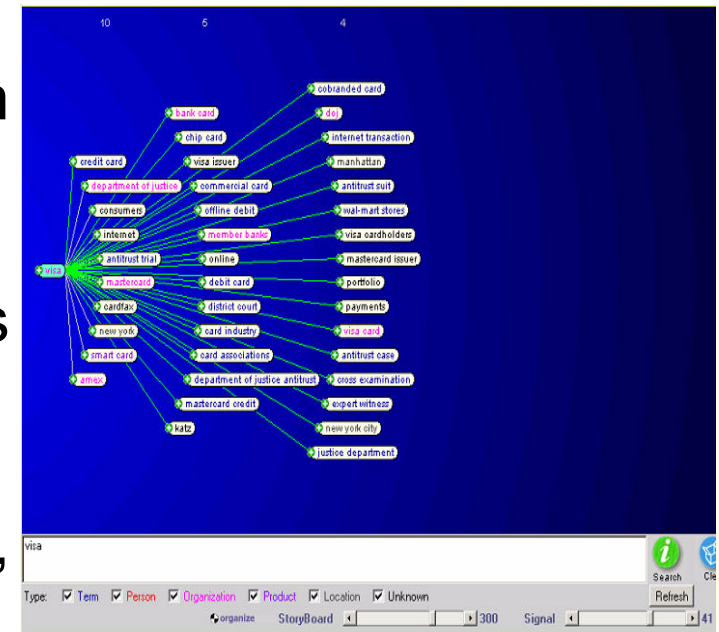
SPSS : Statistical Products for Solutions and Services

- Éditeur logiciel international fondé en 1968
- Cotée au NASDAQ depuis 1993
- Chiffre d'affaires FY2006 : 321 M\$
- Plus de 90% des Fortune 1000 sont clients SPSS
- Plus de 1300 personnes dans le monde
 - 13 collaborateurs au Maghreb (Maroc, Algérie, Tunisie et Mauritanie)
(ventes, marketing, services, R&D)
- Plus de 250 000 clients
 - 500 clients actifs au Maghreb
- Plus de 2 millions d'utilisateurs dans le monde

- Fondée in 1977 : Eril SA à Paris
- 20 premières années ont été consacrées à la R&D du Natural Language Processing (NLP) avec diverses agences gouvernementales européennes et Françaises.
- Racheté par SPSS en Février 2002
- Offres Produits SPSS LexiQuest : Trois produits Text Mining
 - LexiQuest Mine
 - LexiQuest Categorize
 - Text Mining pour Clementine
- Tous utilisent la même technologie d'extraction des concepts (NLP).

LexiQuest Mine

- Outil permettant d'identifier rapidement les concepts clés et les relations qui les lient parmi des milliers de documents.
- Affiche de manière conviviale au sein de cartes graphiques dynamiques.
- Analyse l'évolution des tendances dans le temps pour chaque concept.
- C'est l'outil de choix pour structurer, découvrir et anticiper l'information



LexiQuest Mine

SPSS

SPSS LexiQuest Mine (tm) - Wanadoo

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites

Address http://localhost/mine/applet.html

msn Search Web Games Form Fill

SPSS LexiQuest Mine ...



Keywords

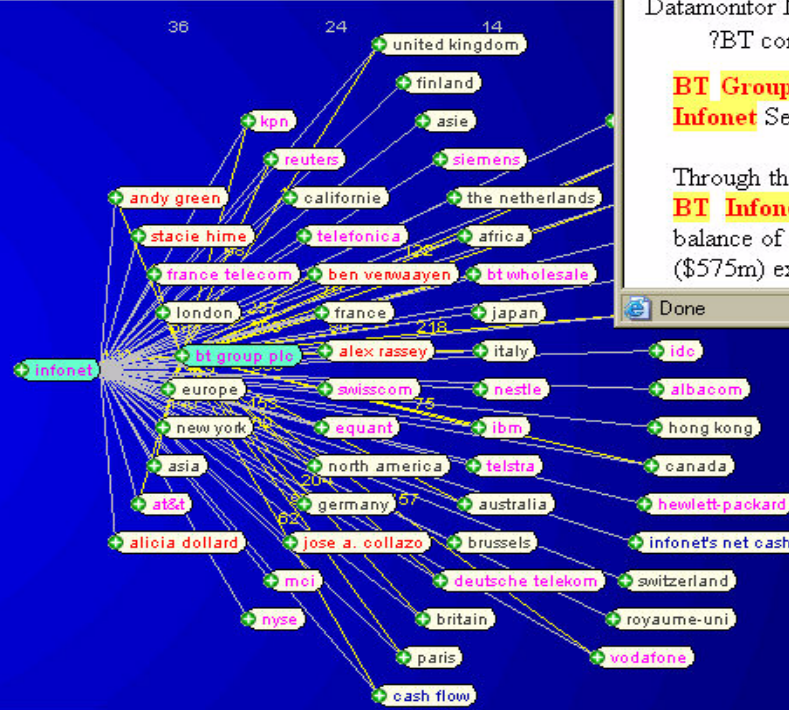
Refine

Base strategy2

Mine Search

Concepts by frequency

bt group plc (21069)
 europe (2362)
 infonet (2182)
 swisscom (2167)
 broadband (1602)
 provider (1142)
 london (1089)
 pac (1087)
 internet (1021)
 data (962)
 ofcom (927)
 pierre danon (865)
 service providers (847)
 communications (815)
 bt global services (811)
 microsoft (737)
 pierre audoin consultants (724)
 bt wholesale (724)



bt group plc
 infonet NEAR bt group plc

Types: ☒ Topic ☒ Person ☒ Organization ☒ Technology ☒ Location ☒ Non Ling ☒ My Category ☐ Unknown

organize

Storyboard

10000

Signal



Refresh

58

Datamonitor News and Comment

?BT completes **Infonet** acquisition ? (28/02/2005)?

BT Group has completed its acquisition of network services provider **Infonet** Services, and renamed the merged entity **BT Infonet**.

Through the deal, London, UK-based **BT** paid 2.06 per share, valuing **BT Infonet** at 510m pounds (\$965m) inclusive of **Infonet**'s cash balance of 205m pounds (\$390m) at March 31, 2004, or 305m pounds (\$575m) excluding the cash balance.

Done

Local intranet

LexiQuest Categorize



- **LexiQuest Categorize** catégorise automatiquement les documents en fonction de leur contenu et d'une taxinomie prédéfinie
- La compréhension de chaque document permet à Categorize de lui affecter des catégories pertinentes.
- Une fois la catégorisation effectuée, le document peut être envoyé à la bonne personne en fonction d'un profil pré-établi ou bien être exploité par un portail intra/extranet.

LexiQuest Categorize

SPSS

The screenshot displays the LexiQuest Taxonomy Manager interface. On the left, a tree view shows a taxonomy for 'CRM' with various sub-categories like 'Business', 'Careers', 'Complaints', etc. The main window shows a list of documents under the 'Documents' tab. A yellow callout points to the 'Developer Support' category in the tree, labeled 'Catégorie identifiée et Confiance'. Another yellow callout points to the 'activex control' descriptor in the table, labeled 'Concepts trouvés, Occurrences, et Poids'. At the bottom, a Notepad window shows the original message text.

Documents | Descriptors | Run 1 | Run 2

Statistics | Documents | By category | Uncategorized | All

Href

- \aster\corpus\corpuser\mustang\messages\business\diebold acquires nexus software.txt
- \aster\corpus\corpuser\mustang\messages\developer support\retrieving messages from mu...
- \aster\corpus\corpuser\mustang\messages\developer support\support inquiry mustang mes...

Category	Confidence	Relevance	Status	Descriptor	Occurrences	Weight
Developer Su...	24	6268	G	activex control	5	1369
				resizing	2	439
				architect	2	370
				pentium iii-500	1	298
				support inquiry	1	298
				system configuration	1	265
				reply asap	1	262
				dimension	1	238
				surprise	1	235
				delphi	1	216
				imc agent	3	202
				crippen	1	202
				steve crippen	1	202
				active	5	201
				resize	1	183
				patches	1	156
				control	5	136

Message Original

Support Inquiry Mustang Message Center.txt - Notepad

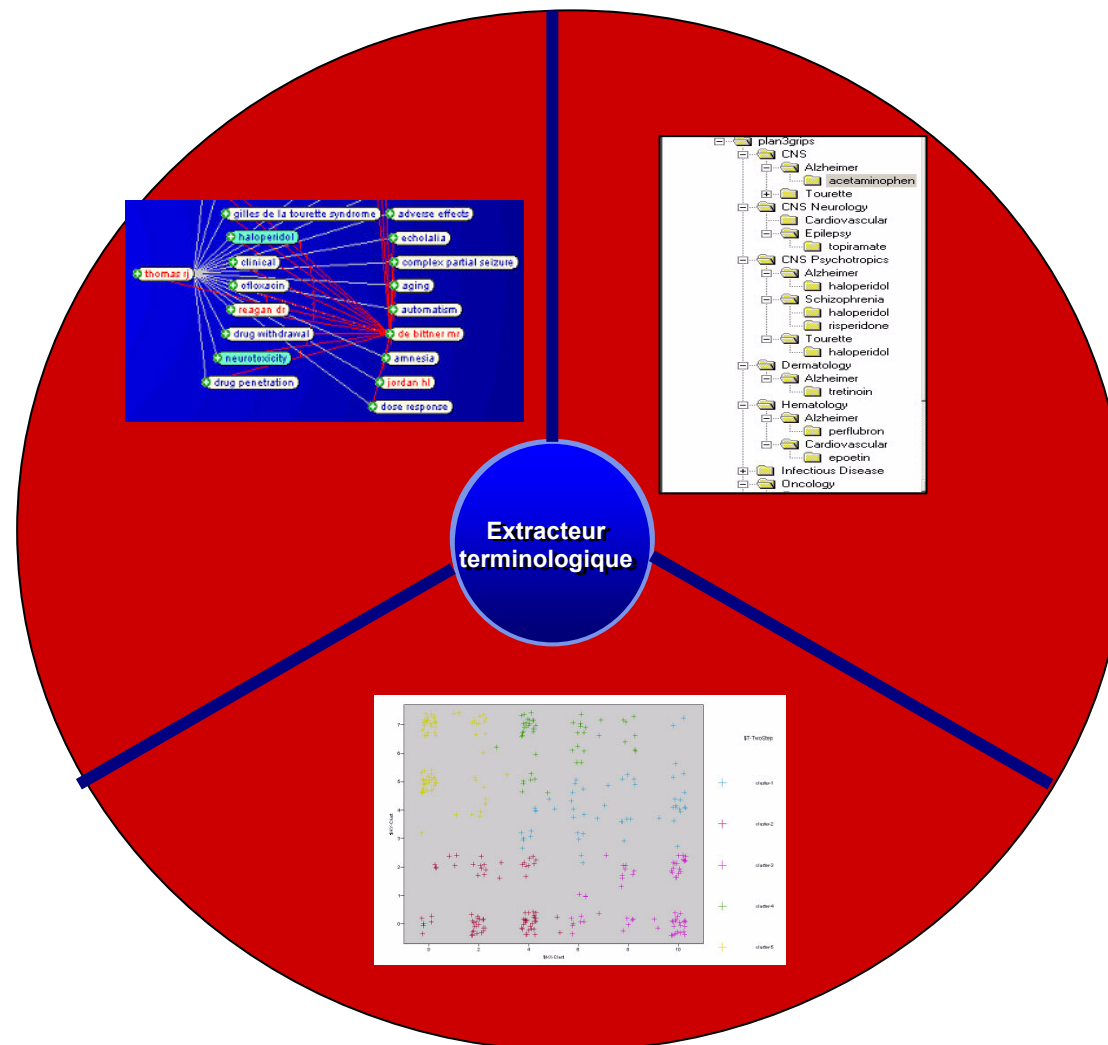
Currently, I am using your company Mustang Architect version 3.0 to develop the IMC Agent. To my surpris
I am using Delphi 5.0. When I placed the ActiveX Control on my form, it is OK. Compile & run .. still ok
I have worked with ActiveX Control & I have not faced this kinds of problem before.
Can you rectify the problem as soon as possible? Maybe in 1 or 2 days as it is urgent. Please include ar

Text Mining for Clementine



- Text-Mining pour Clementine est un composant de la plate forme Clementine permettant de tirer parti des informations textuelles non structurées.
- Elles peuvent ainsi être combinées aux informations provenant des bases de données pour améliorer la prédictibilité de modèles classiques en data mining

Des Concepts aux composants d'Analyse Prédictive



Comment le Text Mining fonctionne?



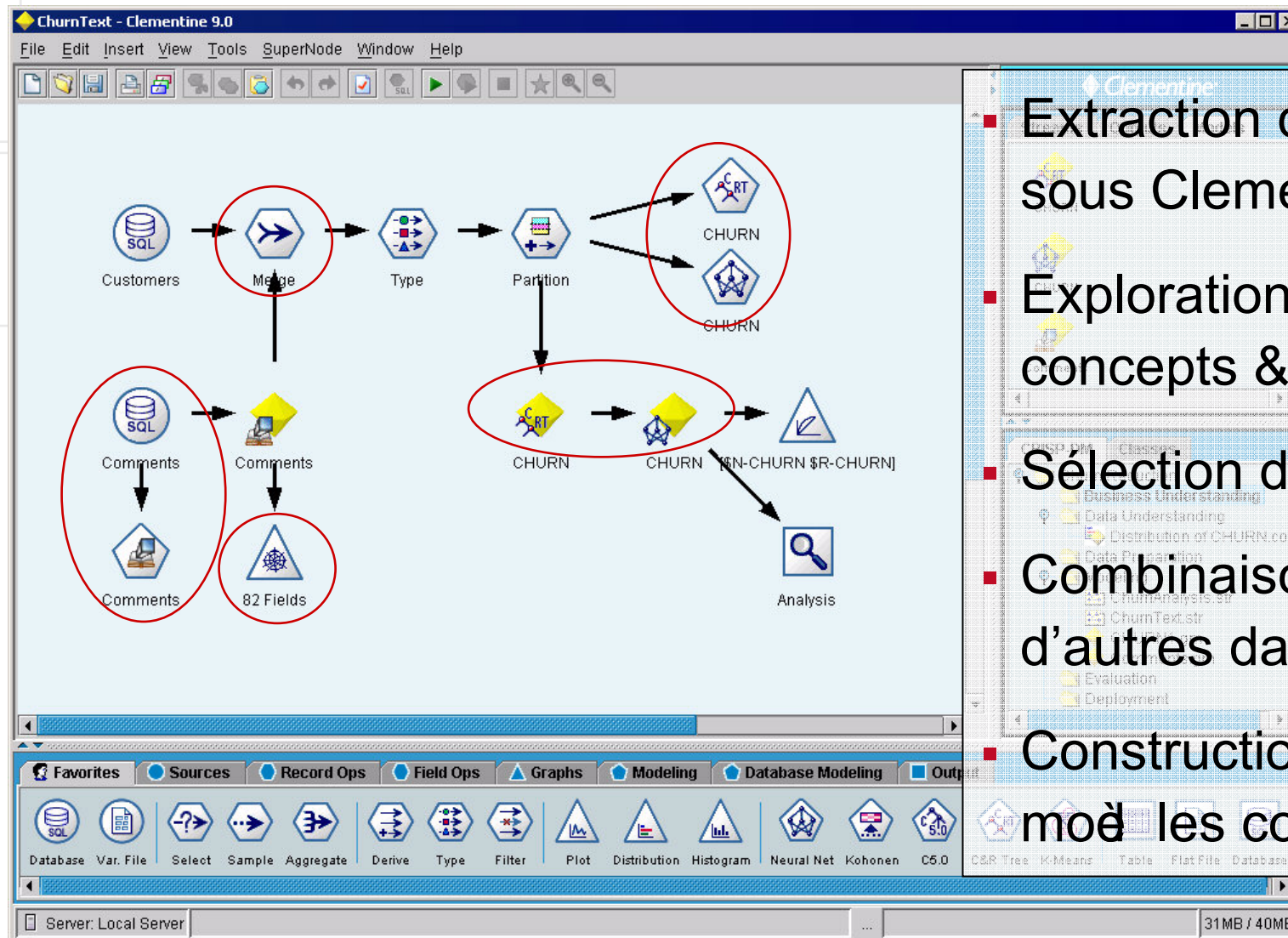
- Fondement: Extraction de Concepts
- Extraction de “concepts” – Mots ou combinaisons de mots tenant compte du contexte d’intérêt
 - Technologie du langage Naturel (**NLP** : Naturel Language Processing) basée sur la linguistique
- Par exemple:
 - Noms, organisations, lieux, choses: “Khalid El Himdi”, “SPSS Inc.”, “Aéroport Mohammed V”, “CODEM”
 - Entités Non-linguistiques : monnaie, temps, dates, numéros de téléphone, CIN, N° SS, etc.
 - Patterns: opinions positive et négative, relations entre d’autres concepts: “Khalid El Himdi a une belle Honda Civic.”

Data Mining et Text Mining Combinés



- Concepts extraits sont maintenant des données structurées
 - Les outils Data Mining peuvent être utilisés pour les analyser
- Les “data” Concepts peuvent être combinés avec d’autres données structurées pour l’analyse
 - Ex. data Concept issus de texte des Call center,
- Data Concept peuvent être générés efficacement de telle manière que le texte libre peut produire de résultats ...
 - Pouvant être utilisés pour la prise de décision
 - En temps réel

Intégration du Data Mining avec le Text Mining



- Extraction de Concept sous Clementine
- Exploration des concepts & relations
- Sélection de concepts
- Combinaison avec d'autres data
- Construction de modèles combinés

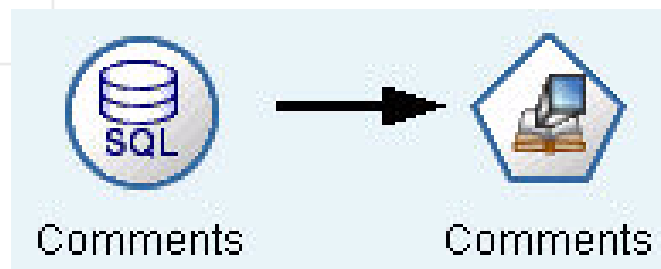
Creation d'un Modèle Text Mining



Text Extraction

SPSS

- Utiliser le noeud modélisation “Extraction de Texte” et indiquer le champ texte désiré.



Comments

Texte :

☒ Le champ de texte représente le texte réel
☐ Le champ de texte représente les noms de chemin vers les documents

Type de document :

Unité textuelle :

Nombre de caractères pour le mode paragraphe
Minimum : Maximum :

Codage du texte source :

☐ Ré-utiliser la traduction existante

Répertoire de traduction :

Mode de partitionnement : ☒ Utiliser les paramètres du noeud ... ☐ Utiliser les paramètres personnels...

Partition :

Champs Extraction Expert Annotations

OK Exécuter Annuler Appliquer Réinitialiser

Creation d'un Modèle Text Mining



Text Extraction

SPSS



Comments



Comments

Comments

Nom du modèle : ☒ Auto ☐ Personnalisé

☒ Utiliser les données partitionnées

Langue : Anglais

Limiter l'extraction aux concepts d'effectif supérieur ou égal à :

☐ Limiter l'extraction aux n premiers

☒ Sélectionner le plus fréquents

☒ Désélectionner les concepts apparaissant dans plus de

Options de traduction

Qualité de la traduction (1=priorité à la vitesse, 7=priorité à la précision)

☐ Texte déjà traduit

Champs Extraction Expert Annotations

OK Exécuter Annuler

Comments

☐ Utilisation de ressources linguistiques spécifiques

☐ Traitement des erreurs de ponctuation

☐ Traitement des fautes de frappe. Nombre de caractères minimum requis : 5

☒ Extraction des unitermes

☒ Extraction des entités non linguistiques

☒ Extraction des termes commençant par une majuscule

☐ Utilisation des patrons dynamiques

Nombre de mots pleins soumis à permutation pour le regroupement : 2

Champs Extraction Expert Annotations

OK Exécuter Annuler Appliquer Réinitialiser

Le Modèle Text Mining

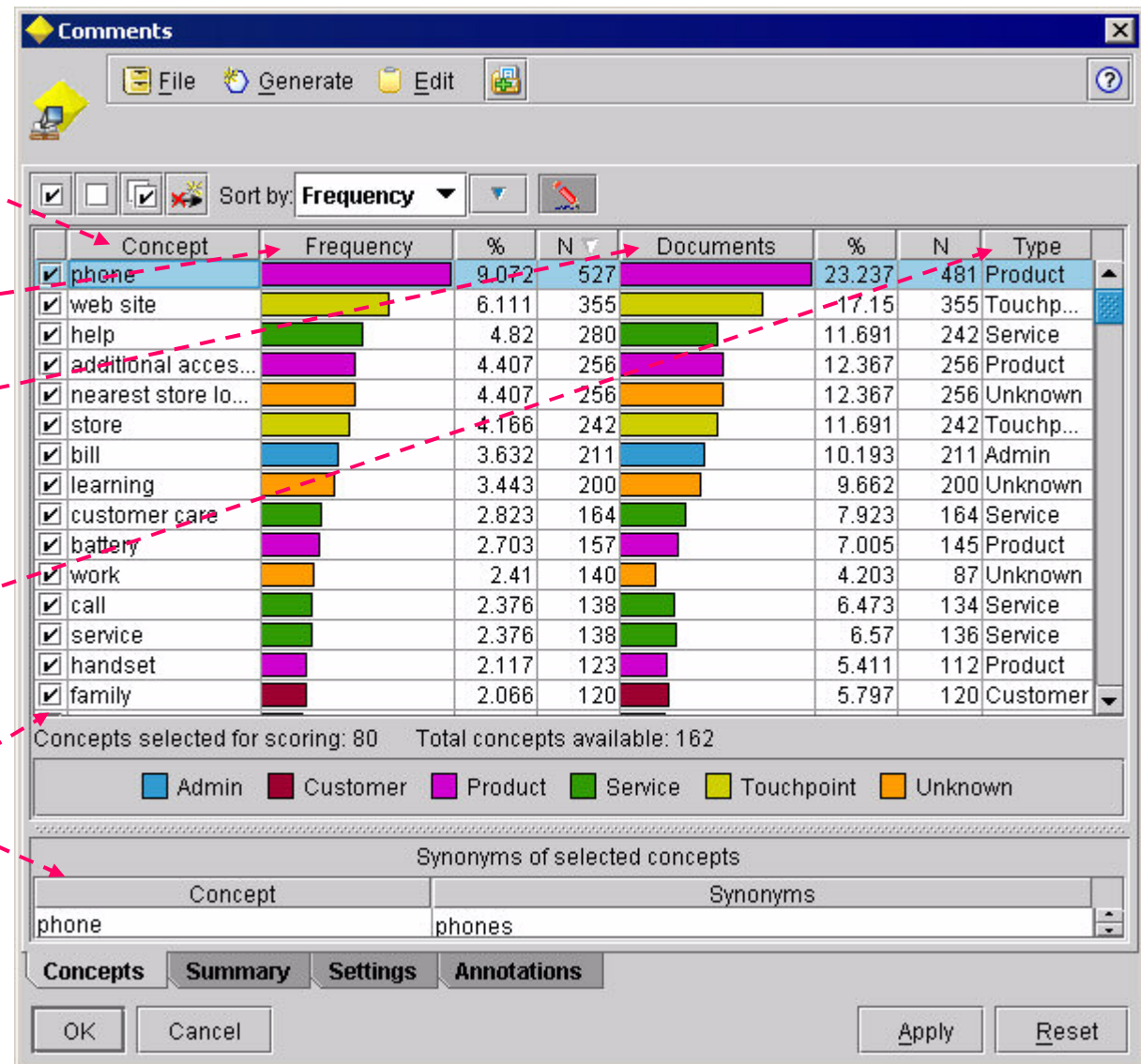


SPSS

Liste de concepts
découvert

Avec:

- Fréquence
et nombre de
documents avec
concept
- type du Concept
- Synonymes
- Sélection de
concepts
 - différentes
méthodes



Sélection de Concept

Select All
Unselect All
Selection options
Import selection from
filter node – e.g. can
generate selection
from a model



Comments

File Generate Edit

Sort by: Frequency

Concept	Frequency	%	N	Documents	%	N	Type
<input checked="" type="checkbox"/> phone	9.072	527	23.237	481	Product		
<input checked="" type="checkbox"/> web site	6.111	355	17.15	355	Touchp...		
<input checked="" type="checkbox"/> help	4.82	280	11.691	242	Service		
<input checked="" type="checkbox"/> additional acces...	4.407	256	12.367	256	Product		
<input checked="" type="checkbox"/> nearest store lo...	4.407	256	12.367	256	Unknown		
<input checked="" type="checkbox"/> store	4.166	242	11.691	242	Touchp...		
<input checked="" type="checkbox"/> bill	3.632	211	10.193	211	Admin		
<input checked="" type="checkbox"/> learning	3.443	200	9.662	200	Unknown		
<input checked="" type="checkbox"/> customer care	2.823	164	7.923	164	Service		
<input checked="" type="checkbox"/> battery	2.703	157	7.005	145	Product		
<input checked="" type="checkbox"/> work	2.41	140	4.203	87	Unknown		
<input checked="" type="checkbox"/> call	2.376	138	6.473	134	Service		
<input checked="" type="checkbox"/> service	2.376	138	6.57	136	Service		
	2.117	123	5.411	112	Product		
	2.066	120	5.797	120	Customer		

total concepts available: 162

Product Service Touchpoint Unknown

synonyms of selected concepts

Synonyms

phones

Annotations

Apply Reset

Check Options

☒ Check top 75 concepts based on frequency

☐ Check top 75 concepts based on document count

☐ Check all concepts

☐ Un-check concepts that occur in more than 95 % of records

☐ Un-check all concepts

OK Apply Cancel Help

Scoring avec le modèle Text Mining

- Le modèle Text mining “score” les concepts sélectionnés



- Un champ par concept
 - Default
- Peut produire un enregistrement par concept

Table (32 fields, 2,070 records) #1

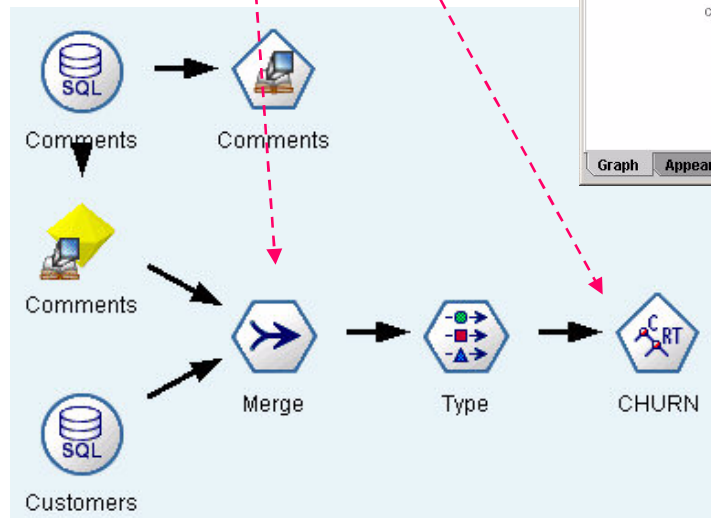
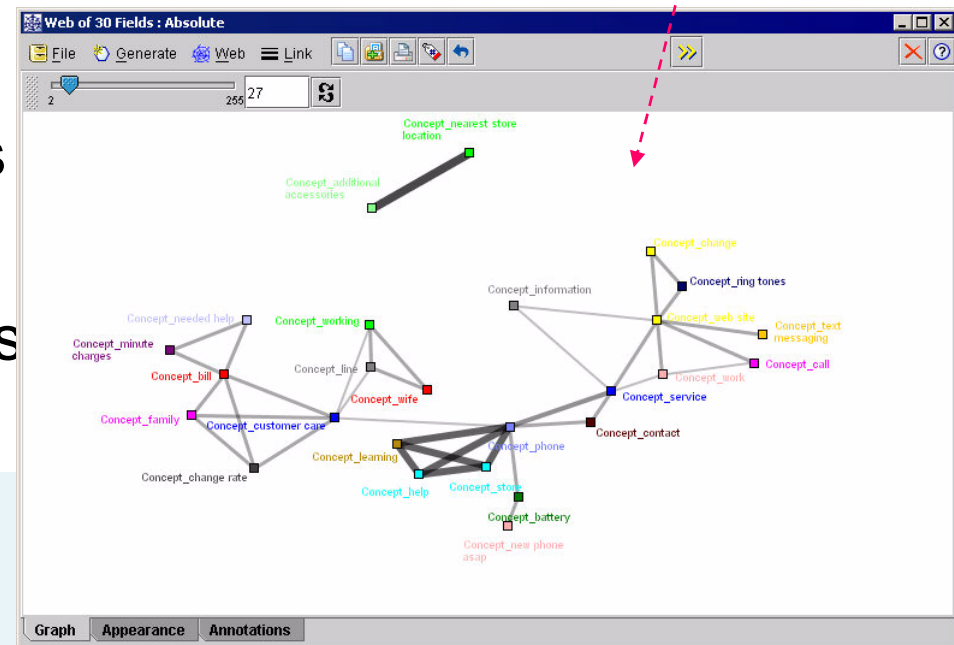
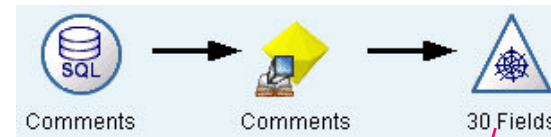
	Concept_battery	Concept_bill	Concept_call	Concept_change	Concept_change address	Concept_change rate
1	F	F	F	F	F	F
2	F	F	F	F	F	F
3	F	F	F	F	F	F
4	F	F	T	F	F	F
5	F	F	F	F	F	F
6	F	F	F	F	F	F
7	F	F	F	F	F	F
8	T	F	F	F	F	F
9	F	F	F	F	F	F
10	F	F	F	F	F	F
11	T	F	F	F	F	F
12	T	F	F	F	F	F
13	F	F	F	T	F	F
14	F	F	F	F	F	F
15	F	F	F	F	F	F
16	F	F	T	F	F	F

Table (5 fields, 4,523 records)

	ID	Comments	Concept	Type	Count
1	1309	Does not like the way the phone works. It is to difficult ...	phone	Unknown	2
2	3556	Wanted to know the nearest store location. Wants to b...	additional accessories	Unknown	1
3	3556	Wanted to know the nearest store location. Wants to b...	nearest store location	Unknown	1
4	2230	Wants to know how to do text messaging. Referred hi...	text messaging	Unknown	1
5	2230	Wants to know how to do text messaging. Referred hi...	web site	Unknown	1
6	2312	Asked how to disable call waiting. referred him to web...	call	Unknown	1
7	2312	Asked how to disable call waiting. referred him to web...	web site	Unknown	1
8	3327	Needs help learning how to use the phone. I suggest...	help	Unknown	1
9	3327	Needs help learning how to use the phone. I suggest...	learning	Unknown	1
10	3327	Needs help learning how to use the phone. I suggest...	phone	Unknown	1
11	3327	Needs help learning how to use the phone. I suggest...	store	Unknown	1
12	3789	Wanted to know the nearest store location. Wants to b...	additional accessories	Unknown	1
13	3789	Wanted to know the nearest store location. Wants to b...	nearest store location	Unknown	1
14	1060	Said his battery never has worked well. Wants a new ...	battery	Unknown	1

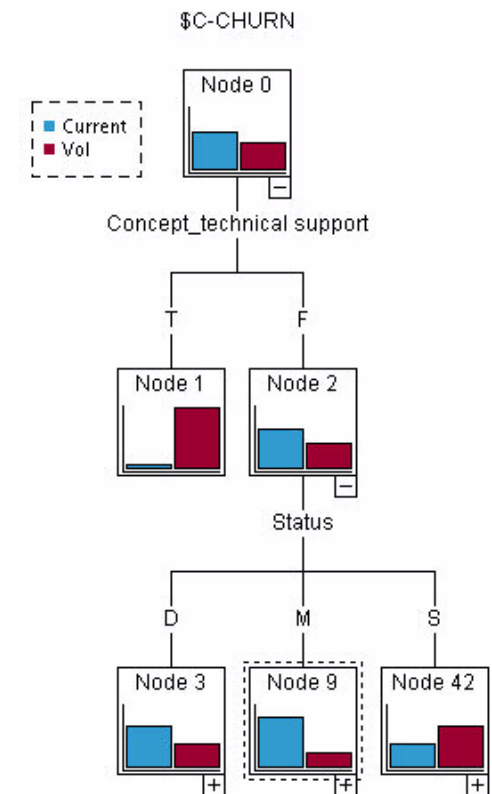
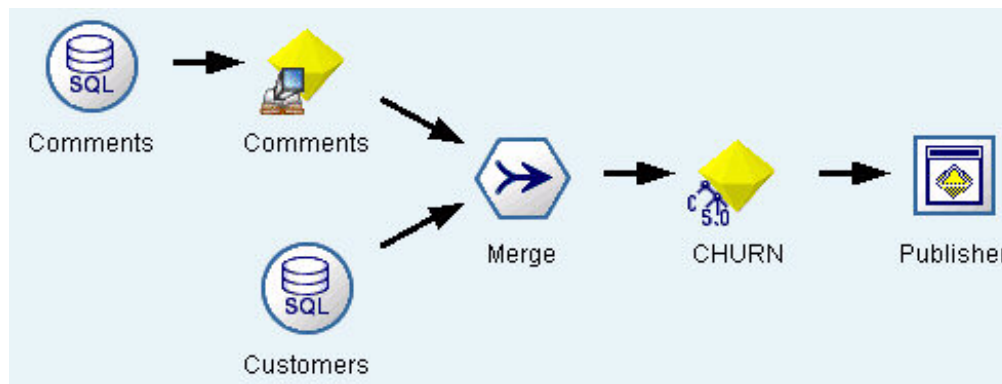
Et puis...

- Extrêmement facile de:
 - Explorer relations entre concepts
 - Fusionner les concepts avec d'autres data
 - Construire des modèles combinés



Et mieux encore...

- Facile de déployer des modèles text mining

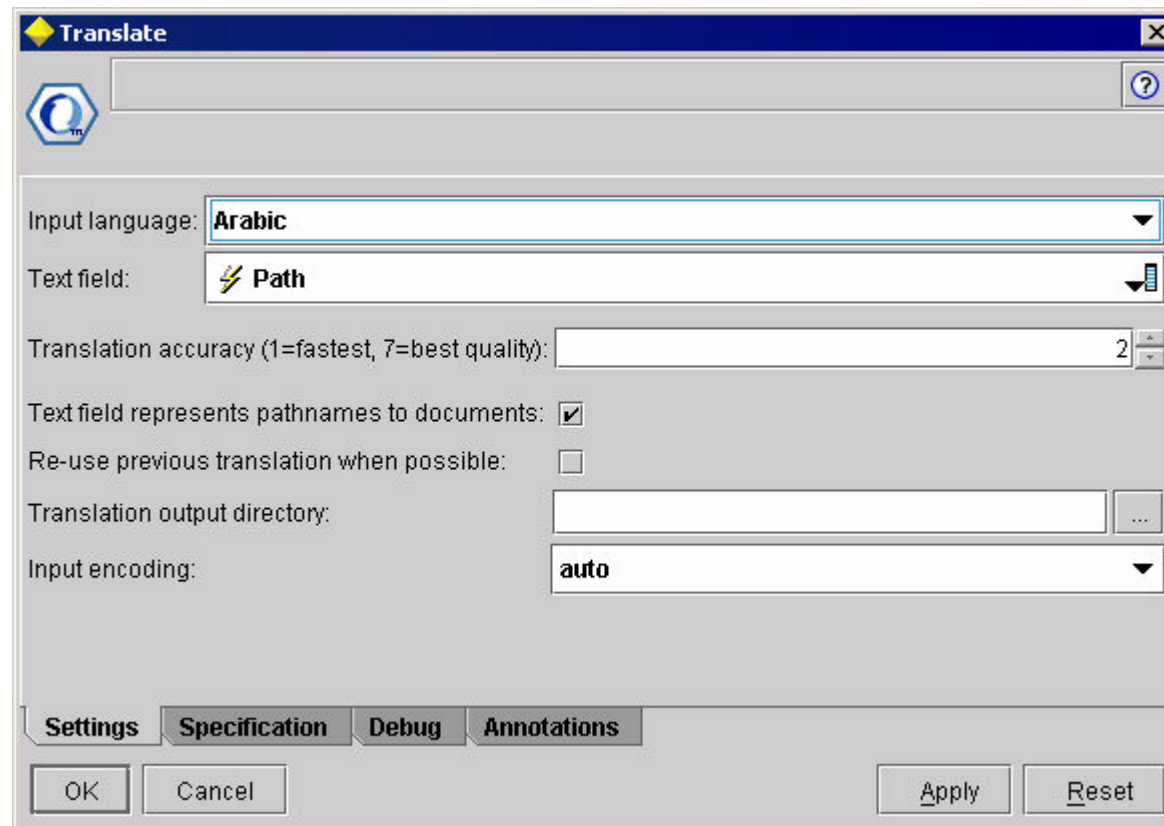


Noeud de Traduction

- Language Weaver translation technology
 - Arabe & Chinois standard
 - Traduire à l'anglais puis appliquer du TM4C



Translate





> Étude de cas: Application de text mining dans la veille stratégique

Présenté par:

M. Khalid ELHIMDI
M. Hamid MACHHOUR

- 1. But et objectifs**
- 2. Définition des besoins**
- 3. Les tâches matiques de la veille**
- 4. Choix des sources d'information**
- 5. Collecte et organisation de l'information**
- 6. Analyse et traitement de l'information**
 1. Élaboration d'un dictionnaire de concepts
 2. Élaboration d'un modèle de classification de documents
 3. Classement et catégorisation de documents
- 7. Conclusion et perspectives**

1. But et objectifs

- Le secteur de télécommunication au Maroc en mouvement: Maroc Telecom, Meditel et Wana (Maroc Connect)
- Identification de la dynamique du secteur de télécommunication au Maroc à travers l'analyse
 - des mouvements des opérateurs concurrents,
 - des clients,
 - des produits et des technologies, etc.

2. Définition des besoins

- L'identification des organisations actives dans le domaine telecom au Maroc:
 - Terminologie technique et industrielle → dictionnaire de concepts
- A base du dictionnaire on doit être capable de dégager des informations à propos de:
 - Emplacement des opérateurs
 - leurs dirigeants
 - leurs partenaires stratégiques
 - ce qui est acheté ou vendu récemment
 - leurs principales compétences
 - leurs situation financière,
 - leurs principaux clients,
 - la part de marché par opérateur,
 - les nouveaux produits ou services en développement,
 - leurs stratégies et activités de commercialisation,
 - leurs fournisseurs, etc.

3. Les thématiques de la veille

- **Offres de produits ou de services** : détecter les nouvelles offres de services et/ou de produits, ainsi que tout signe de dégradation ou de détérioration de ces offres, etc.
- **Ressources financières et de gestion** : détection de noms de dirigeants ou l'embouche de nouvelles compétences, détection de noms de partenaires (banques, universités, bourses, etc.), etc.
- **Concurrents** : détection des opérateurs concurrents sur le marché telecom.
- **Distributions** : détection d'ajout de nouveaux emplacements, points de ventes, centres, etc.
- **Stratégies et démarches de commercialisation** : détection de nouvelles opérations d'investissement, de partenariat et coopération, de vente et/ou d'achat et établissement de prix, etc.

4. Choix des sources d'information

- www.iam.ma: site de Maroc telecom,
- www.meditel.ma: site web de Meditel,
- www.wana.ma: site de Wana(Maroc connect),
- www.mobileiam.ma: site du pôle mobile de Maroc telecom,
- www.menara.ma: site du fournisseur Internet de Maroc telecom,
- www.itmaroc.com

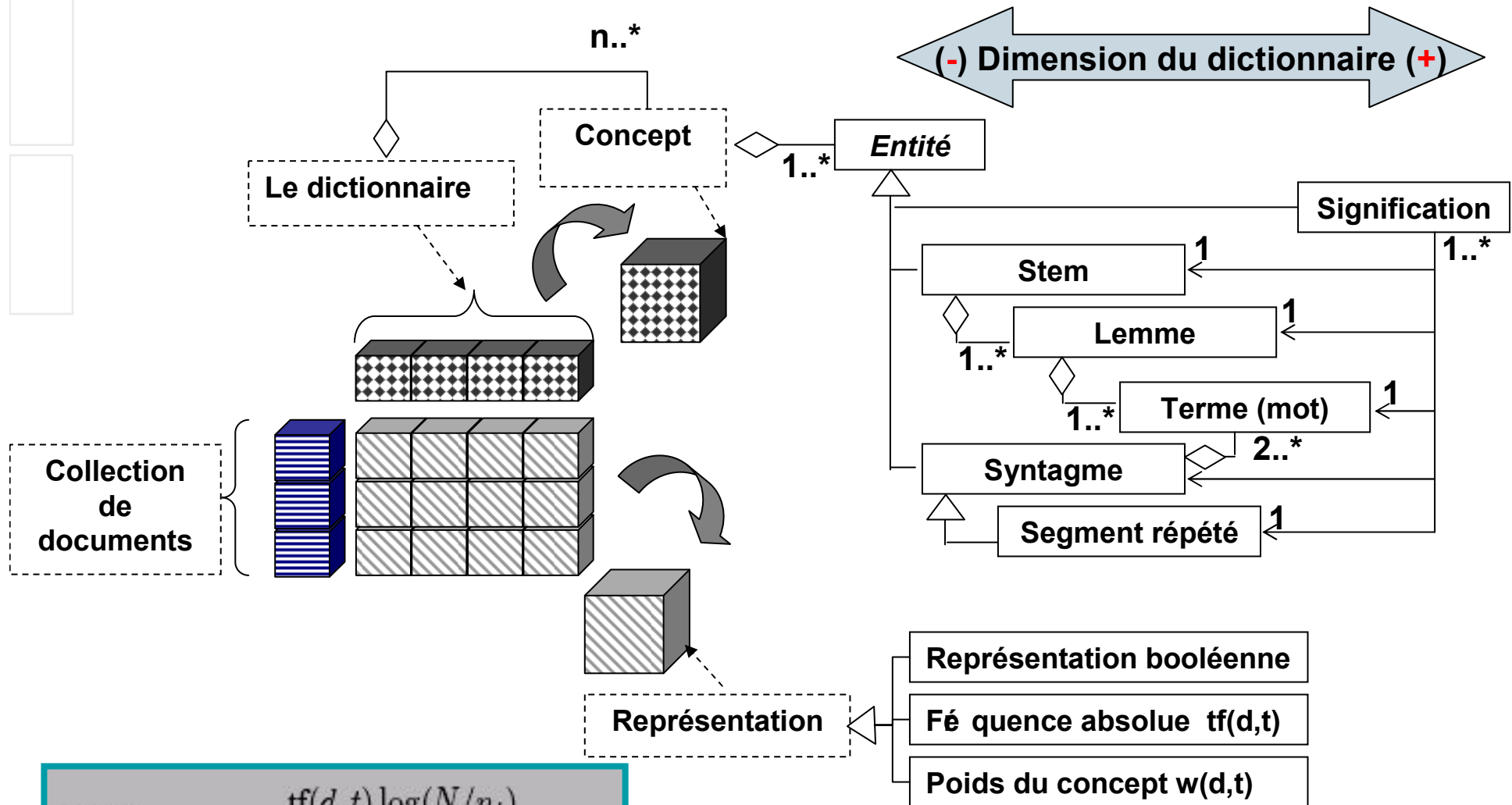
5. Collecte et organisation de l'information

Site web	Nombre de pages téléchargées
www.iam.ma	88
www.meditel.ma	120
www.marocconnect.ma	59
www.mobileiam.ma	132
www.menara.ma	184
www.itmaroc.com	183
Total	766

6. Analyse et traitement de l'information

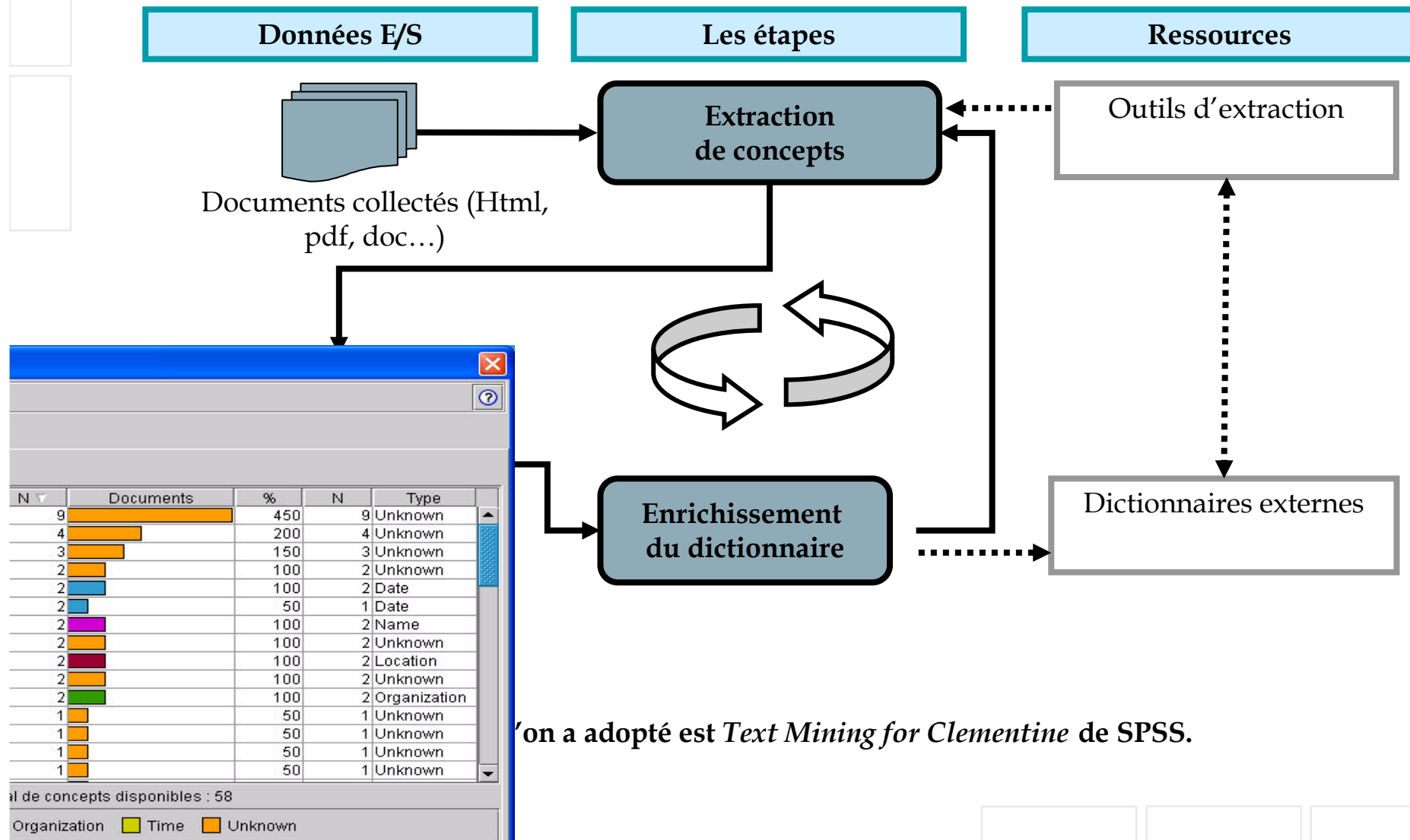
1. Élaborer un dictionnaire de concepts : phase de préparation des données.
2. Élaborer un modèle de classification de documents : phase de modélisation.
3. Classer de nouveaux documents : phase de test et d'évaluation.

6.1. Élaboration d'un dictionnaire de concepts

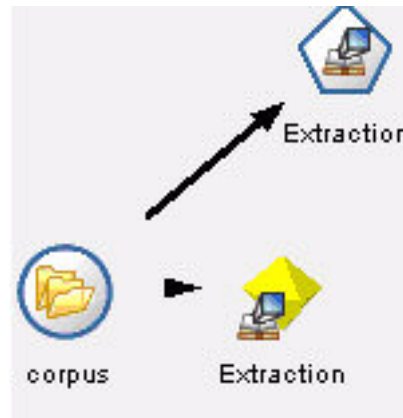


$$w(d,t) = \frac{tf(d,t) \log(N/n_t)}{\sqrt{\sum_{j=1}^m tf(d,t_j)^2 (\log(N/n_{t_j}))^2}}$$

6.1. Élaboration d'un dictionnaire de concepts

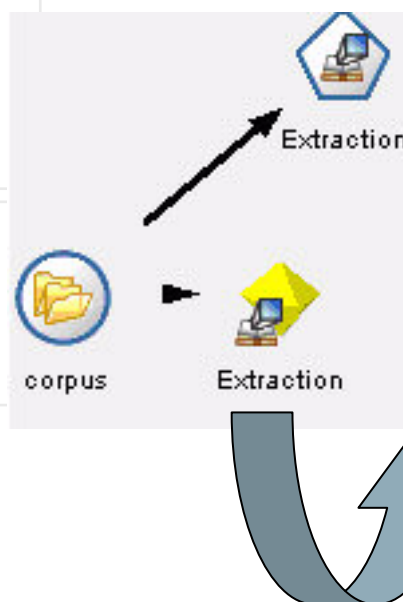


6.1. Élaboration d'un dictionnaire de concepts



- L'ajout de concepts dans ces dictionnaires doit respecter une syntaxe prédéfinie par l'outil TM4C → Text Mining Builder de SPSS
- Le raffinement de concepts nécessite une bonne connaissance de la terminologie du domaine spécifique étudié.

6.1. Élaboration d'un dictionnaire de concepts:



Extraction

Fichier

Générer

Edition

Trier par:

Fréquence

	Concept	Fréquence	%	N	Documents	%	N	Type
<input checked="" type="checkbox"/>	produit		28,487	1399		66,667	178	Product
<input checked="" type="checkbox"/>	service		25,677	1261		86,142	230	Type de concept
<input checked="" type="checkbox"/>	distribution		11,342	557		70,412	188	Location
<input checked="" type="checkbox"/>	stratégie-commerciale		9,55	469		68,165	182	Term
<input checked="" type="checkbox"/>	ressource-financière		6,312	310		34,457	92	Term
<input checked="" type="checkbox"/>	meditel		4,663	229		43,446	116	Organization
<input checked="" type="checkbox"/>	maroc-telecom		4,541	223		31,835	85	Organization
<input checked="" type="checkbox"/>	ressource-gestion		3,054	150		33,708	90	Term
<input checked="" type="checkbox"/>	maroc connect		2,79	137		21,348	57	Organization
<input checked="" type="checkbox"/>	amélioration		2,26	111		21,723	58	Term
<input checked="" type="checkbox"/>	innovation		0,774	38		11,236	30	Term
<input checked="" type="checkbox"/>	détérioration		0,55	27		8,24	22	Term

Concepts sélectionnés pour le calcul des scores : 12 Nombre total de concepts disponibles : 12

Date

Location

Name

Organization

Percent

Product

Synonymes des concepts sélectionnés

Concept	Synonymes
---------	-----------

Concepts

Récapitulatif

Paramètres

Annotations

OK

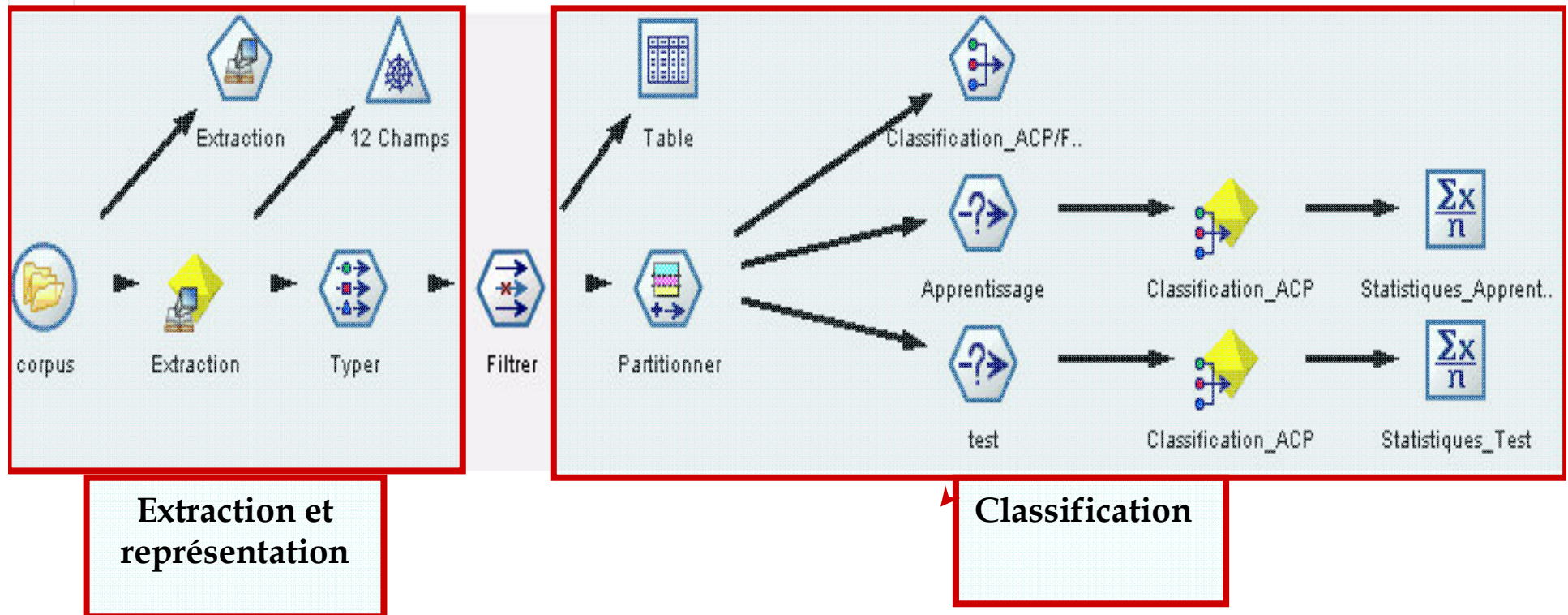
Annuler

Appliquer

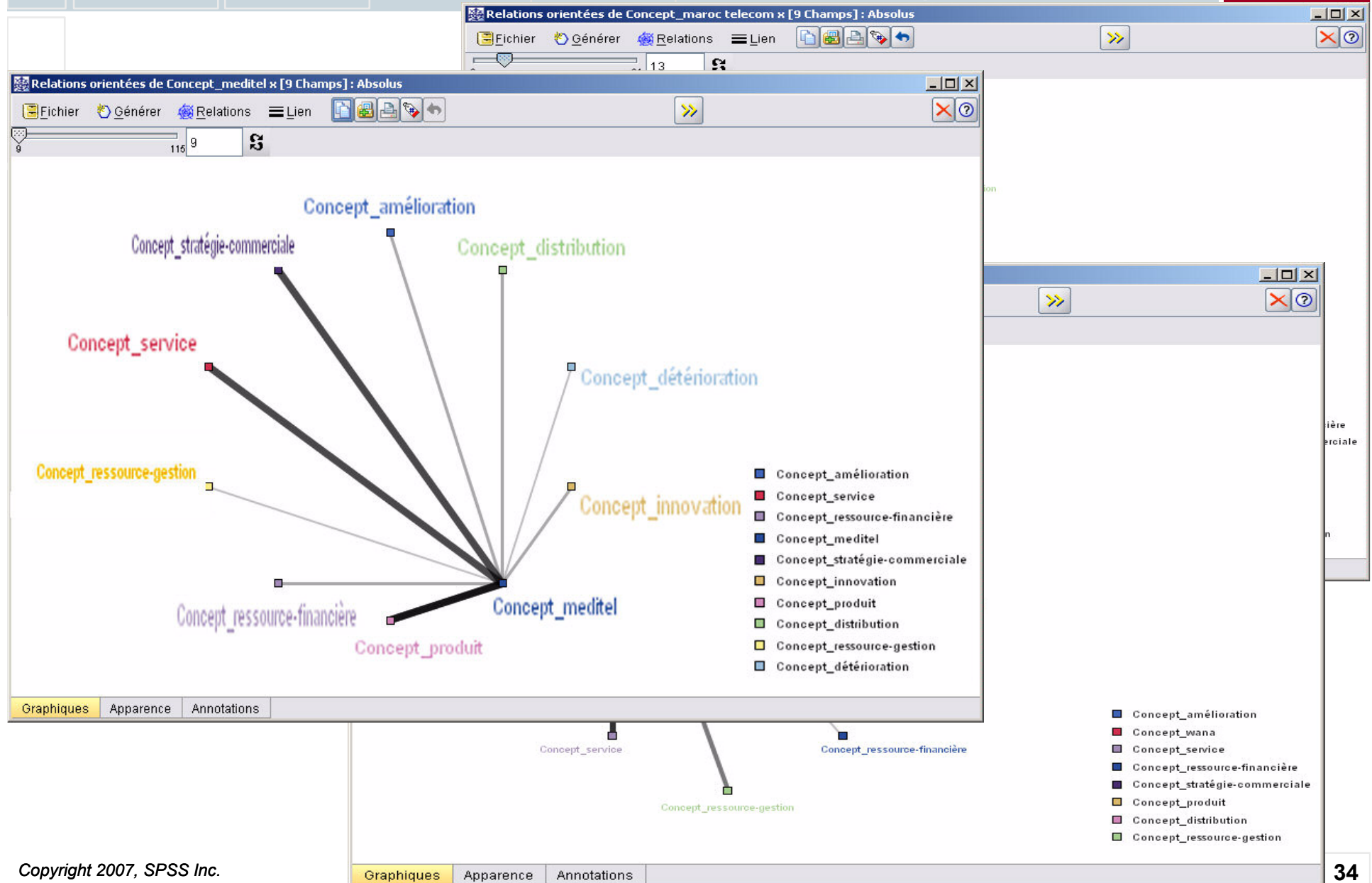
Réinitialiser

Liste de concepts clés extraits

6.2. Élaboration d'un modèle de classification



6.2. Élaboration d'un modèle de classification: Les relations entres les concepts



6.2. Élaboration d'un modèle de classification: Les modèles de classification ACP générés

	Composante			
	1	2	3	4
Concept_stratégie-commerciale	.790			
Concept_service	.754			
Concept_distribution	.741			
Concept_maroc connect	-.551		.501	
Concept_produit	.548	-.540		
Concept_maroc-telecom		.931		
Concept_ressource-financière		.807		
Concept_meditel	.542	-.645		
Concept_amélioration			.748	
Concept_ressource-gestion			.538	
Concept_innovation			.520	
Concept_détérioration				.835

Valeur booléenne

	Composante				
	1	2	3	4	5
Concept_service	.811				
Concept_meditel	.793				
Concept_stratégie-commerciale	.741				
Concept_distribution	.678				
Concept_innovation		.832			
Concept_amélioration		.726			
Concept_ressource-gestion		.518			.475
Concept_ressource-financière			.880		
Concept_détérioration				.915	
Concept_maroc-telecom			.585	.625	
Concept_maroc connect					.769
Concept_produit					.484

Fé quence absolue

	Composante				
	1	2	3	4	5
p_meditel	.842				
p_distribution	.809				
p_service	.685				
p_str_com	.600				.484
p_deter		.795			
p_produit		.755			
p_maroc_con			-.857		
p_maroc_tel	-.520		.673		
p_res_fin			.506		
p_innovation				.789	
p_amélioration				.661	
p_res_ges					.796

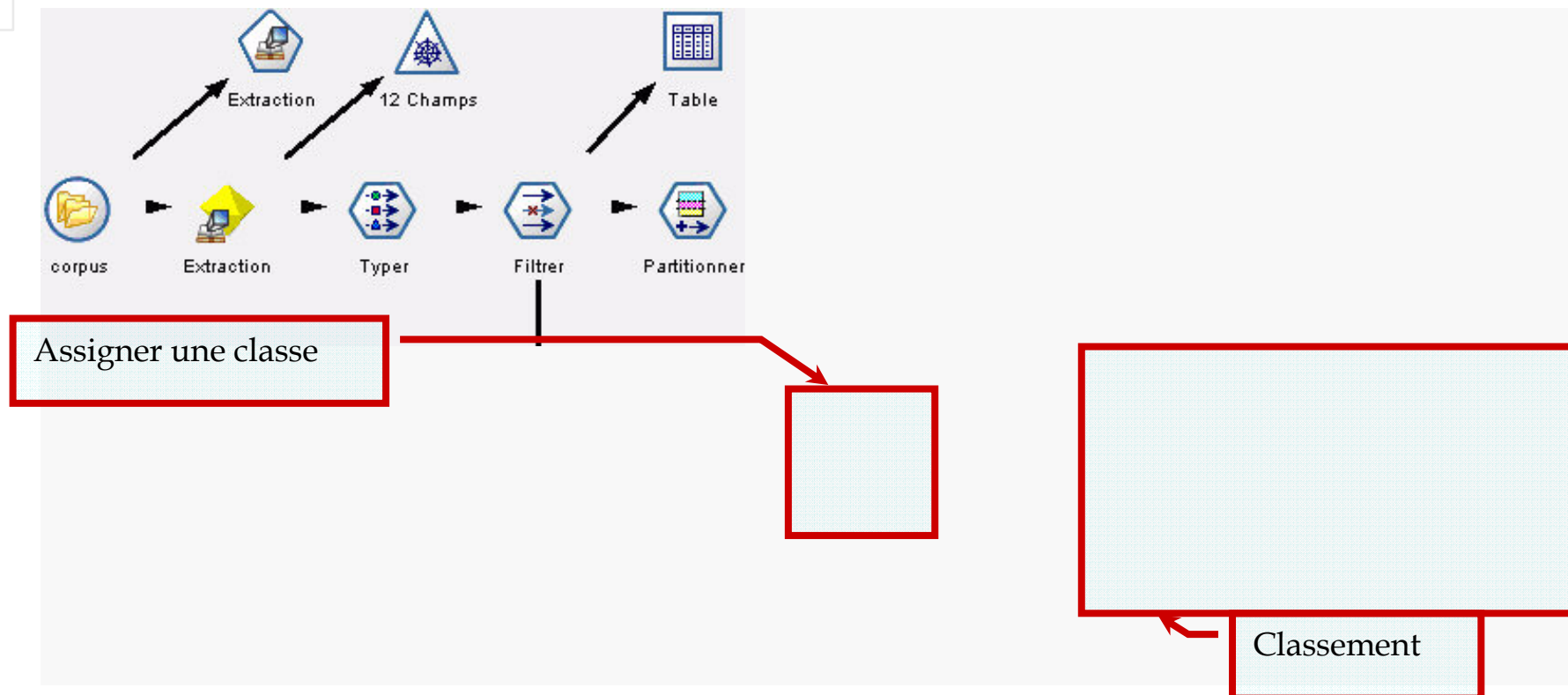
Poids

6.2. Élaboration d'un modèle de classification: Interprétations

- La relation entre les concepts « Meditel », « Stratégie commerciale », « Service » et « Distribution » persiste dans les 3 cas, cela signifie que l'opérateur Meditel adopte une stratégie commerciale qui se base sur l'offre de services aux différents points de vente.
- Alors que la persistance de la relation entre les concepts « Maroc Telecom » et « Ressource financière » indique que le mouvement actuel de l'opérateur Maroc Telecom s'oriente vers la recherche de ressources financières à travers des coopérations et des partenariats.
- l'apparition simultanée des concepts « Wana », « ressource de gestion » indique que l'opérateur Wana s'oriente vers l'innovation et l'amélioration de ses ressources de gestion.

6.3. Classement et catégorisation

- Le classement vise à assigner des classes prédéfinies aux documents. Quel que soit la méthode spécifique employée, le classement commence avec un ensemble de documents dit « d'apprentissage », ces documents sont déjà étiquetées avec une classe C.



6.3. Classement et catégorisation:

Catégorisation d'un nouveau document

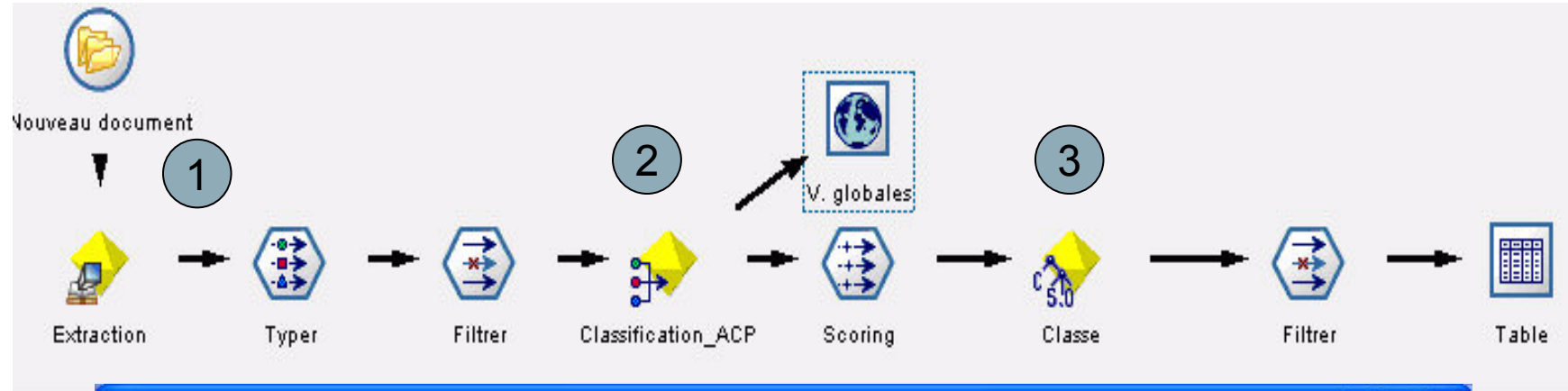








Table (3 champs, 183 enregistrements)

	DocID	Document	\$C-Classe
1	1	C:\Documents and Settings\sims\Bureau\hamid\nouveau doc\itmaroc\www.itmaroc.com\01.htm	C2
2	2	C:\Documents and Settings\sims\Bureau\hamid\nouveau doc\itmaroc\www.itmaroc.com\012.htm	C8
3	3	C:\Documents and Settings\sims\Bureau\hamid\nouveau doc\itmaroc\www.itmaroc.com\013.htm	C5
4	4	C:\Documents and Settings\sims\Bureau\hamid\nouveau doc\itmaroc\www.itmaroc.com\014.htm	C2
5	5	C:\Documents and Settings\sims\Bureau\hamid\nouveau doc\itmaroc\www.itmaroc.com\015.htm	C1
6	6	C:\Documents and Settings\sims\Bureau\hamid\nouveau doc\itmaroc\www.itmaroc.com\016.htm	C5
7	7	C:\Documents and Settings\sims\Bureau\hamid\nouveau doc\itmaroc\www.itmaroc.com\017.htm	C2
8	8	C:\Documents and Settings\sims\Bureau\hamid\nouveau doc\itmaroc\www.itmaroc.com\018.htm	C2
9	9	C:\Documents and Settings\sims\Bureau\hamid\nouveau doc\itmaroc\www.itmaroc.com\019.htm	C2
10	10	C:\Documents and Settings\sims\Bureau\hamid\nouveau doc\itmaroc\www.itmaroc.com\110\100.htm	C2
11	11	C:\Documents and Settings\sims\Bureau\hamid\nouveau doc\itmaroc\www.itmaroc.com\110\101.htm	C2
12	12	C:\Documents and Settings\sims\Bureau\hamid\nouveau doc\itmaroc\www.itmaroc.com\110\102.htm	C2
13	13	C:\Documents and Settings\sims\Bureau\hamid\nouveau doc\itmaroc\www.itmaroc.com\110\103.htm	C2
14	14	C:\Documents and Settings\sims\Bureau\hamid\nouveau doc\itmaroc\www.itmaroc.com\110\104.htm	C2
15	15	C:\Documents and Settings\sims\Bureau\hamid\nouveau doc\itmaroc\www.itmaroc.com\110\105.htm	C1
16	16	C:\Documents and Settings\sims\Bureau\hamid\nouveau doc\itmaroc\www.itmaroc.com\110\106.htm	C8
17	17	C:\Documents and Settings\sims\Bureau\hamid\nouveau doc\itmaroc\www.itmaroc.com\110\107.htm	C8
18	18	C:\Documents and Settings\sims\Bureau\hamid\nouveau doc\itmaroc\www.itmaroc.com\110\108.htm	C8
19	19	C:\Documents and Settings\sims\Bureau\hamid\nouveau doc\itmaroc\www.itmaroc.com\110\109.htm	C2
20	20	C:\Documents and Settings\sims\Bureau\hamid\nouveau doc\itmaroc\www.itmaroc.com\111\110.htm	C5

6.3. Classement et catégorisation:









Catégorisation d'un nouveau document

Valeur booléenne

Valeur ▲	Proportion	%	Comptage
C1		6,56	12
C2		65,03	119
C3		0,55	1
C4		2,73	5
C5		8,2	15
C8		16,94	31

« Maroc Telecom »
« ressources financières »

Féquence absolue








Valeur ▲	Proportion	%	Comptage
C1		25,14	46
C2		0,55	1
C3		30,6	56
C4		1,09	2
C5		17,49	32
C7		1,09	2
C8		1,64	3
C9		22,4	41

« Meditel », « Stratégie commerciale »,
« Service » « Distribution »,

« Maroc Telecom »
« ressources financières »

« Wana »
« ressources de gestion »

Poids

Valeur ▲	Proportion	%	Comptage
C1		28,42	52
C2		36,61	67
C3		10,93	20
C4		3,83	7
C5		12,02	22
C6		0,55	1
C9		7,65	14

« Meditel », « Stratégie commerciale »,
« Service » « Distribution »,

« détérioration » de « produit »

« Maroc Telecom »
« ressources financières »

« stratégie commerciale »
« ressources de gestion »

7. Conclusion et perspectives

1. La représentation des concepts par rapport aux documents est une valeur numérique : **poids** du concepts, **la fréquence** absolue ou une **valeur booléenne**.
2. Pour mener un projet de text mining correctement, en particulier un projet de veille, on doit révéler les besoins latents en information correctement.
3. Dans un environnement en constant changement, les besoins en information des organisations évoluent très rapidement
 - Actualisation et évolution des axes de veille
 - Enrichissement itératif des dictionnaires utilisés lors de l'extraction de concepts
 - Une bonne connaissance de la terminologie du domaine spécifique étudié
5. L'information extraite doit être **diffusée** au bon moment et à la bonne personne afin qu'elle soit utilisée.
6. **L'utilisation** de l'information justifie la mise en œuvre du processus de veille. Elle permet de mesurer l'utilité de l'information diffusée et de guider le veilleur dans ses actions futures.
7. **Le retour d'expérience** donne de la valeur à l'information et fournit au veilleur des **précisions** telles que:
 - l'utilité et la pertinence de l'information diffusée,
 - la véracité de cette information,
 - la nécessité de poursuivre les recherches d'information
 - la possibilité de passer à un autre axe de veille.

SPSS en quelques chiffres



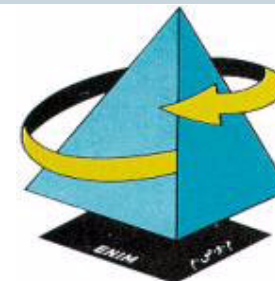
SPSS : Statistical Products for Solutions and Services

- Éditeur logiciel international fondé en 1968
- Cotée au NASDAQ depuis 1993
- Chiffre d'affaires FY2006 : 321 M\$
- Plus de 90% des Fortune 1000 sont clients SPSS
- Plus de 1300 personnes dans le monde
 - 13 collaborateurs au Maghreb (Maroc, Algérie, Tunisie et Mauritanie) (ventes, marketing, services, R&D)
- Plus de 250 000 clients
 - 500 clients actifs au Maghreb
- Plus de 2 millions d'utilisateurs dans le monde

- Création depuis : 1999
- Représentant exclusif de SPSS Inc. au Maroc, en Algérie, en Tunisie et en Mauritanie
- Plus de 80 clients dans la région - Plus de 900 licences vendues
- Portefeuille diversifié de clients opérant dans différents secteurs d'activités

Quelques clients SPSS Maghreb

SPSS



Royaume du Maroc

Haut Commissariat au Plan
Direction de la statistique

MINISTRE DES FINANCES ET
DE LA PRIVATISATION



Quelques clients SPSS Maghreb

SPSS



جامعة محمد الخامس السويسي
Université Mohammed V - Souissi

