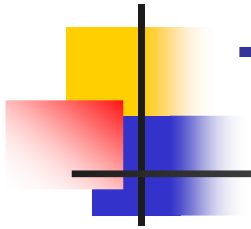


Text classifier (1/5)

- Feasability study of a web money manager
- Concept:
  - The user tags transactions by setting them into a category
  - An algorithm extract relevant words from text labels in classified transaction, in order to automatically classify future transactions

Text classifier (2/5)

- Result is a self-maintained accounting with tree-like categories (requires 3 month training)

Valeur		Mois					
Top Categ	**Categorie**	**Janvier**	**Février**	**Mars**	**Avril**	**Mai**	**Total**
Home	Common	-862,63	-1149,48	-779,05	-991,74	-996,25	-4779,15
	Remboursem	1237,41	-44,5	100			1292,91
	Impôts	-230	-230	-230	-230	-230	-1150
	Frais CA	-6,1	-6,1	-6,1	-23,6	-6,1	-48
	Multimedia					112,45	112,45
Somme Home		**138,68**	**-1430,08**	**-915,15**	**-1245,34**	**-1119,9**	**-4571,79**
Epargne	Furaxis	-15	-15	-15	-15	-15	-75
	Atout libre	-843,77					-843,77
	Codevi				-300	-550	-850
Somme Epargne		**-858,77**	**-15**	**-15**	**-315**	**-565**	**-1768,77**
Life	Téléphone	-87,78	-80,1	-73,01		-64,82	-305,71
	Retraits	-350	-360	-190	-370	-390	-1660
	Paiements	-1190,05	-1111,77	-442,8	-475,48		-3220,1
	Server		35,86				35,86
	Cheques	-69					-69
	Lecture				-52,6		-52,6
	Cadeau					-15	-15
Somme Life		**-1696,83**	**-1516,01**	**-705,81**	**-898,08**	**-469,82**	**-5286,55**
Assurance	Santé		184,28		-67,13	-18,98	98,17

Text classifier (3/5)

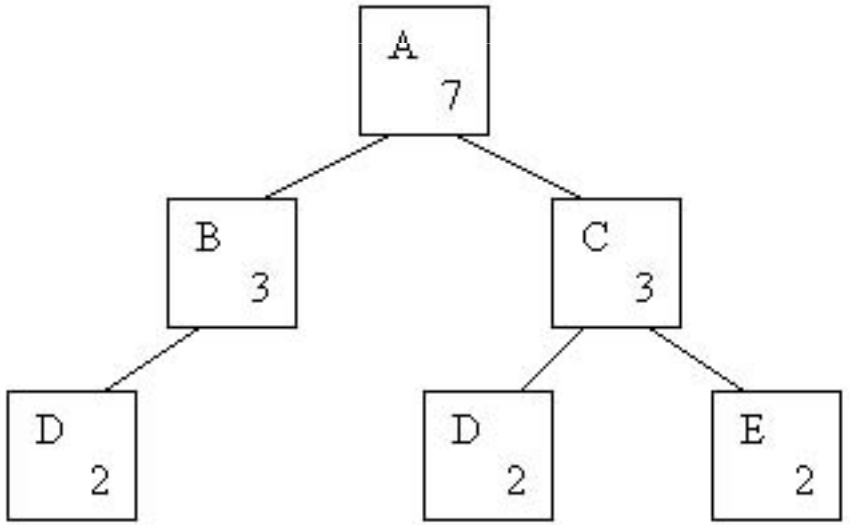
- Learning consists in retrieving relevant keywords (i.e. filters) for a category

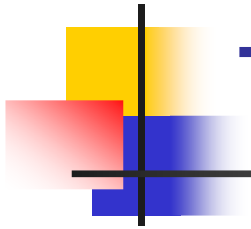
	Tokens > 2								
List	PAYEMENT	JANUARY	FEBRUARY	INTERNET	TELEPHON	SALARY	*Filters*		*Blacklist*
Group1									
PAYEMENT JANUARY INTERNET 0659885							PAYEMENT, INTERNET		JANUARY, FEBRUARY
PAYEMENT FEBRUARY INTERNET 8975462									
PAYEMENT JANUARY TELEPHON 145232							PAYEMENT, TELEPHON		
PAYEMENT FEBRUARY TELEPHON 46546578									
Group2									
SALARY JANUARY							SALARY		
SALARY FEBRUARY									

Text classifier (4/5)

- A Closed Item Sets algorithm detects most frequent and precise (i.e. complex) patterns

```
Category 1
A.B.D.M.*
A.B.D.N.*
A.B.E.O.*
A.C.D.P.*
A.C.D.Q.*
A.C.E.P.*
A.C.E.Q.*
```





Text classifier (5/5)

- Good « early » algorithm
- Nice prototype interface on Ruby/Ajax
- Too complex from a business point of view