



# ***Automatic profiling system for ranking candidates answers in Human Resources***



Rémy Kessler, Juan Manuel Torres-Moreno  
Mathieu Roche, Nicolas Bechet et Marc El-Bèze

[remy.kessler@univ-avignon.fr](mailto:remy.kessler@univ-avignon.fr)



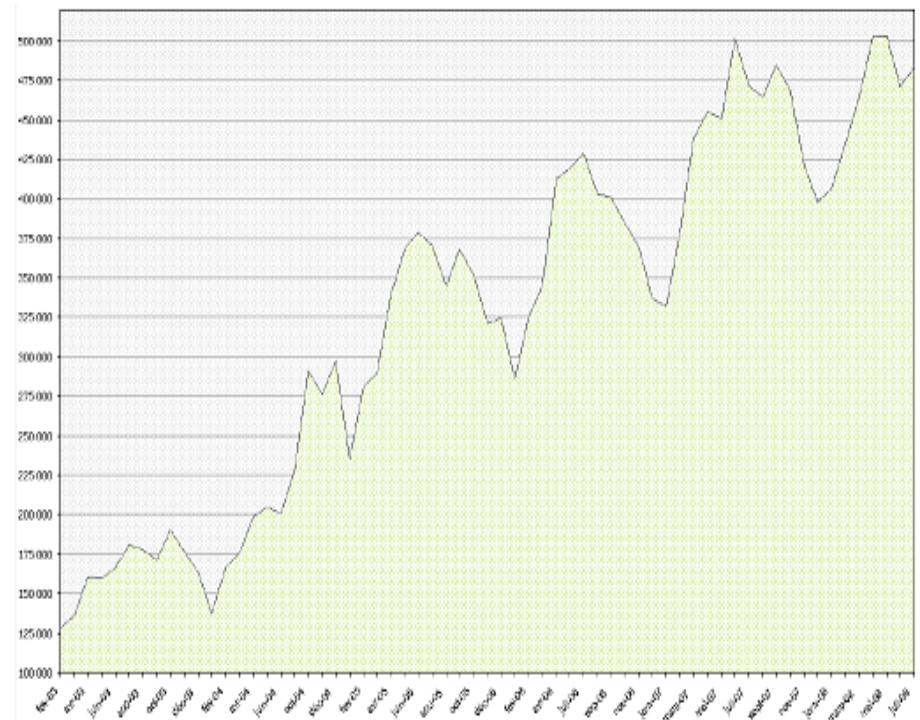
# Introduction

□ exponential growth of Internet  
=>online job-search sites market

2003 : 177 000 jobs offers

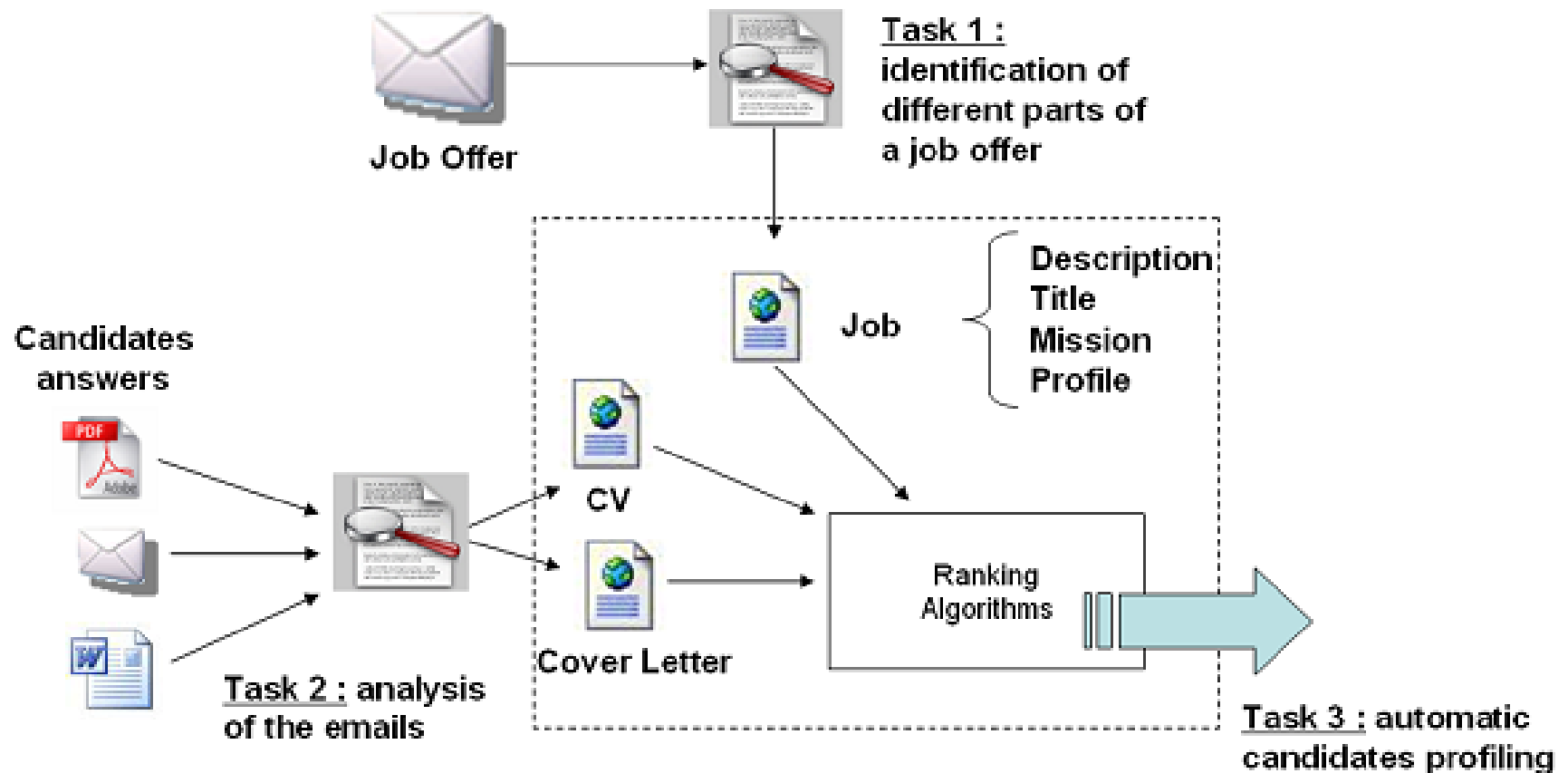
2008 : 500 000 jobs offers

□ can not be managed efficiently by  
companies



Source Keljob.com

# System Overview



## *Corpus informations and statistics*

Total number of jobs offers	25
jobs offers with less 10 candidates	2
jobs offers with more 10 candidates	8
jobs offers with more 50 candidates	6
jobs offers with more 100 candidates	9
Total Number of candidates	2916
candidates tagging <b>positive</b>	220
candidates tagging <b>negative</b>	2696

### ■ Several topics :

- jobs in accountancy
- business enterprise
- computer sciences
- cook
- etc...

### ■ Tagged positive or negative by a recruiting consultant

**Positive** : potential candidate for a given job

**Negative** : a irrelevant candidate for the job (decision of the recruiting consultant)

### ■ Each job offer is associated to at least 4 candidates

## Job Offer analysis (DTMP)

Description  
of the company (D)

This french firm, specialised in chemical analysis, is looking for:

PERSON IN CHARGE OF LABORATORY  
TRANSFER  
South East

← Title (T)

Mission (M)

→ You will be in charge of regrouping the transfer activities of different analysis laboratories.

You will analyse, conduct and implement the necessary phases of the project, respecting budgets and previously defined, dead lines.

Your solution will need to consider different parameters of the project (social, logistic, materials, data processing...) and integrate a roadmap (production, methods, accreditations, development, commercial... ).

Being a focus on environmental analytical chemistry, you have already led an activity transfer project.  
Fluent English required. Please send your CV and cover letter indicating reference number VA 11/06 to [beatrice.lardon@atalan.fr](mailto:beatrice.lardon@atalan.fr)

← Profile (P)

# Example of Curriculum Vitae (CV)

## LAETITIA BRUNET

7 rue Barrée

69 230 GRIGNY

Port : 06.06.06.06.06

Date of birth : 2 Juin 1985

E-Mail : [laetitia.brugnet@gmail.com](mailto:laetitia.brugnet@gmail.com)

Driver licence

### Educational background

- 2004/2006 : BTEC Higher National Diploma (HND) NRC in Auxerre
- 2002/2004 : baccalaureat: high school diploma in business in Auxerre

### Professional experience

⇒ Training course in a regional newspaper

⇒ Training course in Fiat Automobile

• 10 weeks

⇒ since september 2004 : supermarket cashier in Géant Casino in Auxerre (week ends and holidays).

⇒ 2004 : Seasonal worker in the LCL bank agency at the reception

⇒ Training course in Citroën Automobile

• 18 Weeks

### **Others**

English (strong knowledge)

Tennis (good level – plays in competition)

- not sentence
- summerize of idea
- visual segmentation
- relevant collocation

## *Example of cover letter (CL)*

Nom : LADET  
prénom : Marc

Name : LADET  
Firstname : Marc

Monsieur

Votre annonce en référence a retenue toute mon attention, vous trouverez donc ci-joint mon curriculum vitae. Vous constaterez à la lecture de mon CV une bonne expérience de structures touristiques dont j'assume les directions depuis 15 ans. Je me suis toujours impliqué dans les installations que je dirigeais, aussi bien au niveau de la gestion des hommes, que financière, et je suis particulièrement attaché à la préservation du patrimoine et au respect des conditions de vente.

Disponible pour vous rencontrer à la date qui vous conviendra, veuillez agréer, monsieur, mes salutations distinguées.

Dear Mr,

I would like to express my interest for your job offer. You can find enclosed my Curriculum Vitae. You will see from reading my CV good experience in tourist structures which I manage since 15 years. I am always involved in installations that I was heading, both in terms of managing men, and financial, and I am particularly attached to heritage preservation and compliance with the conditions of sale.

Available to meet you at the time that suits you, please accept, sir, my highest consideration.

- End of cover letter : salutation
- Summarizes important points in connection with the job's offer
- Facultative



# Preprocessing



## *Filtering / lemmatization*

- Deletion of verbs and functional words (stoplist)

to be, to have, to be able to, to need,...

- Deletion of common expressions (stoplist/collocations)

for example, that is, each of,...

- Lemmatization processing

sing, sang, sung, singer will be transformed into sing.

=► These process allows to decrease the curse of dimensionality

## Vectorial representation

$\Gamma_i^\mu$  = Frequency of the term  $i$  in the  $\mu$  segment

	1	2	3	...	$i$	...	$N-1$	$N$	Term
1	1	3	1	0	...		0	0	
2	0	1	0	0	...		0	2	
3	0	0	0	4	...		1	0	
...	⋮	⋮	⋮	⋮	⋱		⋮	⋮	
$\mu$									
	1	2	0	1	...		0	0	
segment $P$	0	0	0	0	...		1	0	

**Frequency Matrix for each term / segment**



# Ranking algorithms

## ***Measures of similarity (1/2)***

- A number of similarity measures have been tested

- Cosine (Manning&Schutze(2000))

$$sim_{cosine}(j, d) = \frac{j_i \cdot d_i}{\sqrt{\sum_{i=1}^n |j_i|^2 \cdot \sum_{i=1}^n |d_i|^2}}$$

- Enertex (Fernandez(2007))

$$E_{\mu, \nu} = -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N s_{\mu}^i J^{i,j} s_{\nu}^j$$

- Overlap (Manning&Schutze(2000))

$$sim_{Overlap}(j, d) = \frac{j_i \cdot d_i}{Min \left( \sum_{i=1}^n |j_i|^2, \sum_{i=1}^n |d_i|^2 \right)}$$

## *Measures of similarity (2/2)*

- number of similarity measures to determine which is most effective

- Okabis (El-beze & Bellot, 2001)

$$Okabis(j, d) = \sum_{w \in d \cap j} \frac{TF_{w,d}}{TF_{w,d} + \frac{\sqrt{|d|}}{M_S}}$$

- Minkowski(Fernandez(2007))

$$sim_{Minkowski}(j, d) = \frac{1}{1 + (\sum_{i=1}^n |j_i - d_i|^p)^{\frac{1}{p}}}$$

- Needleman-Wunsch algorithm(Needleman-Wunsch(1970))

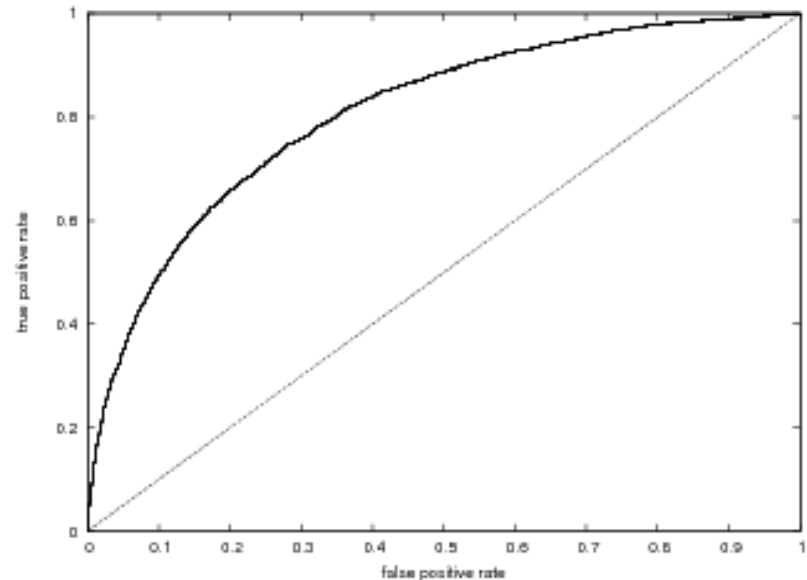
=> Measures combined by an Algorithm of Decision (AD)



# Results

# Roc Curve

- AUC (Area Under the Curve) can be interpreted as the effectiveness of a measurement of interest
  - Equivalent to the statistical test of Wilcoxon-Mann-Whitney
  - Advantage: independence of quantity of +/- examples
- In the case of candidate answers ranking, a perfect ROC curve corresponds to obtain all relevant candidate answers at the beginning of the list and all irrelevant at the end (AUC=1)



# Results

	AUC	Enertex	Cosine	Minkowski	Manhattan	NW	Overlap	Okapi	Decision
DTMP	CL	0,524	0,567	0,561	0,591	0,481	0,573	0,521	<b>0,596</b>
	CV	0,524	<b>0,604</b>	0,510	0,503	0,532	0,543	0,541	0,562
	CL+CV	0,523	<b>0,621</b>	0,539	0,532	0,509	0,522	0,523	0,571
TMP	CL	0,524	0,560	0,559	0,580	0,473	0,562	0,513	<b>0,591</b>
	CV	0,523	<b>0,622</b>	0,508	0,501	0,544	0,538	0,542	0,561
	CL+CV	0,523	<b>0,642</b>	0,538	0,528	0,526	0,531	0,532	0,592

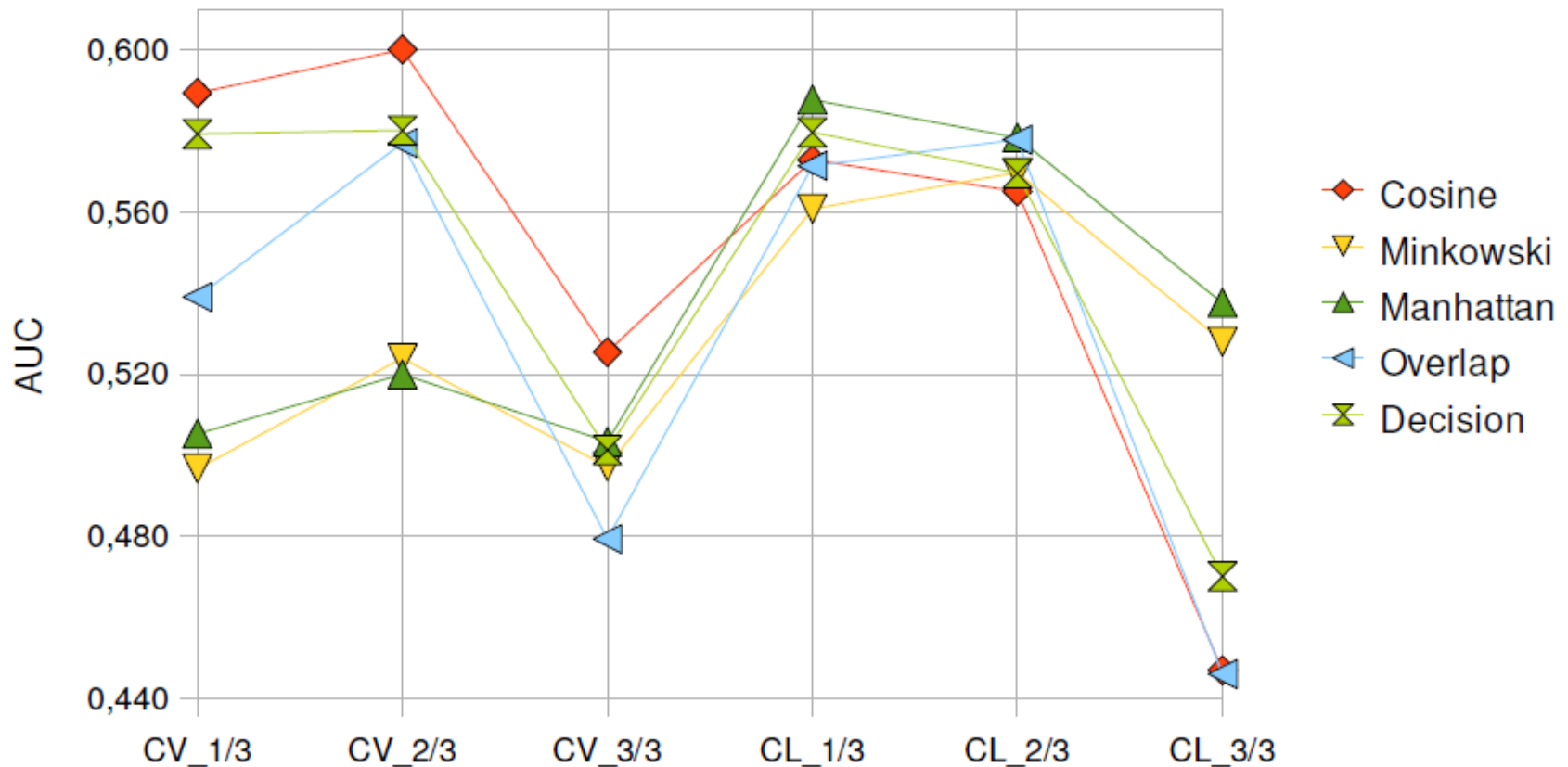
AUC obtained with different segmentations

AUC	Enertex	Cosine	Minkowski	Manhattan	NW	Overlap	Okapi	Decision
CV_1/3	0,525	<b>0,589</b>	0,497	0,505	0,533	0,539	0,569	0,579
CV_2/3	0,524	<b>0,600</b>	0,524	0,520	0,515	0,577	0,560	0,580
CV_3/3	0,526	<b>0,526</b>	0,497	0,503	0,510	0,479	0,506	0,501
CL_1/3	0,527	0,573	0,561	<b>0,588</b>	0,480	0,571	0,528	0,580
CL_2/3	0,533	0,565	0,570	<b>0,578</b>	0,481	0,578	0,543	0,570
CL_3/3	0,516	0,447	0,528	<b>0,538</b>	0,416	0,446	0,439	0,470

AUC obtained with different parts of **CV** and **CL**



## *AUC values vs parts of CV/CL (several similarity measures are compared)*



# ***Conclusion***

- Processing job information is a difficult task
- First results obtained are interesting but needs to be improved
  - cosine measure gives best results for almost all approaches
  - Decision Algorithm are noisy by the poor performance of certain measures
  - Last part of CV or CL contains less information
- Currently testing a part-of-speech tagger combining with term weighting to improve performance
- E-Gen is multilingual, database independent and portable
- First (Job offer analysis) and second module (sort of candidate emails ) are currently in test on Aktor's server



Thank you for your attention

Gracias por su atención