

## Prévision automatisée des résultats de l'élection présidentielle de 2007

Voici une étude qui paraîtrait presque truqué si elle n'était pas aussi pragmatique.

L'idée d'essayer de prévoir les résultats des votes pour la présidence Française en 2007 m'est venu un peu par hasard et par gout de la Médiamétrie.

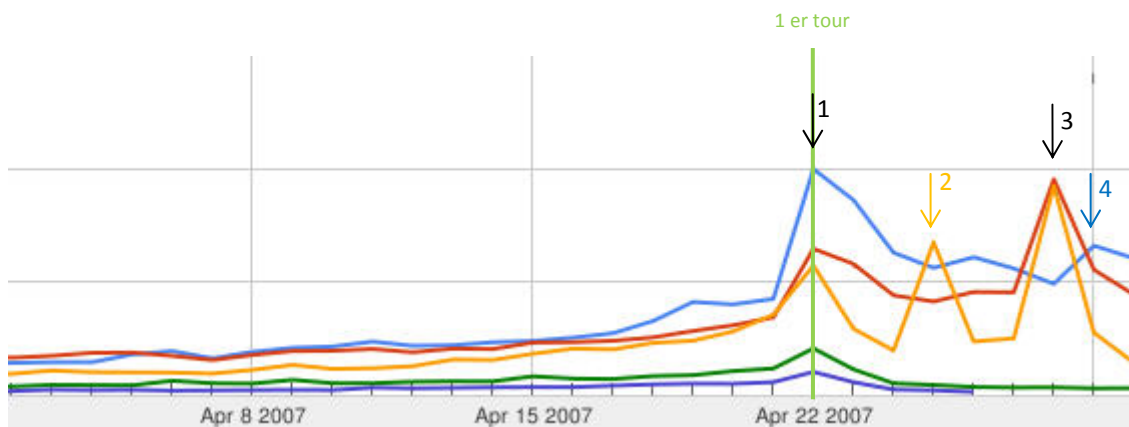
Le 23 avril 2007, lendemain du premier tour de l'élection, j'ai commencé à trifouiller dans les médias. Ce que je cherchais, c'était évidemment une analyse automatisée. Pas question d'y passer des semaines. Après avoir manipulé quelques outils, j'ai mis au point une méthode basée sur l'analyse automatisée des moteurs de recherche. Cette méthode consiste à analyser les mots clés recherchés par les utilisateurs "Français" sur les moteurs de recherche et de présenter les résultats sous forme graphique. Ce qui m'a interpellé dans cette méthode, c'est la formidable justesse des résultats obtenus.

Alors bien sûr, les résultats présents dans cette étude sont a posteriori me direz-vous. Ce qui est intéressant ce serait les mêmes résultats prévisible avant les élections, et vous auriez bien raison.

J'ai donc essayé de renouveler l'affaire sur le deuxième tour le 6 mai 2007. Le résultat n'est pas aussi juste dans le pourcentage que le premier, mais reste tout de même correct. Sarkozy devrait être élu président de la république.

Voici pourquoi j'ai décidé d'y consacrer un article. L'analyse a été faite sur les mois d'avril et de Mai 2007, car ce sont les plus cohérents.

Sur cette première courbe, on peut distinguer quatre piques majeurs.



Légende : —▶ : pique impliquant plusieurs acteurs

—▶ : pique Bayrou

—▶ : pique Sarkozy

—▶ : pique Royal

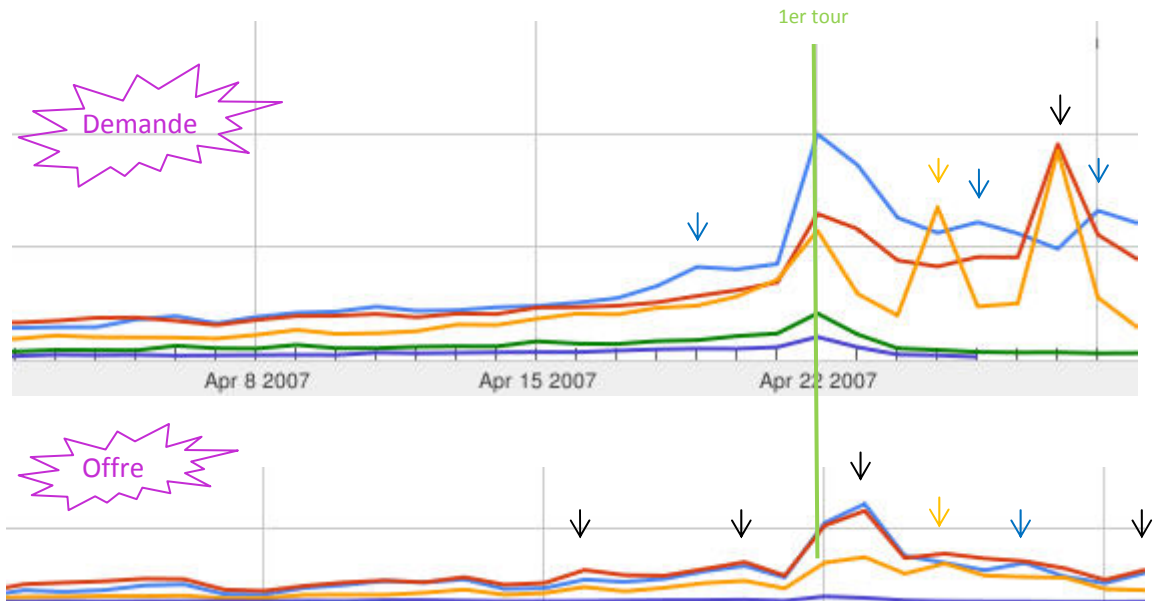
**Pique 1** : le vote du premier tour de l'élection présidentielle explose l'engouement et la curiosité des français.

**Pique 2** : Bien que Bayrou ne soit pas au second tour, il a fait un excellent score post-électoral. Les 5 % de différence entre Royal et Sarkozy renvoie à s'intéresser aux électeurs de Bayrou et sa consigne de vote. Car celle-ci va être décisive dans la balance pour la présidentielle. C'est pourquoi il y a eu une forte augmentation du terme "Bayrou" recherché entre le 24 et le 26 avril, date du discours de François Bayrou sur ses positions.

**Pique 3** : Ce pique renvoie au débat très controversé entre Royal et Bayrou qui s'est déroulé le Samedi 28 avril 2007 à 11 heures.

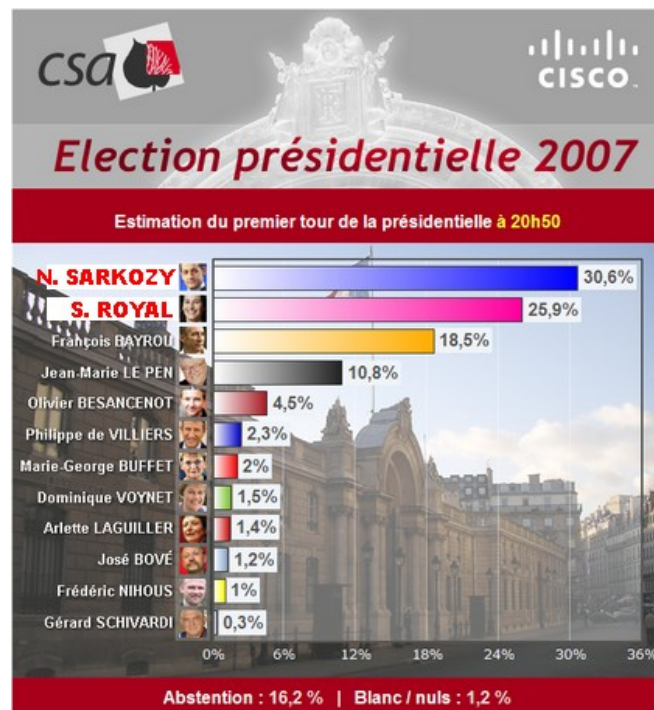
**Pique 4** : Le Dimanche 29 avril, Nicolas Sarkozy remonte légèrement dans la courbe du graphique. Ceci s'explique dans les médias par la date de son meeting à 16 heures, salle omnisport de Paris Bercy. "Ensemble tout devient possible" : c'est au court de ce rassemblement que Nicolas Sarkozy fit un discours post premier tour.

Voici la courbe qui montre, par une simple analyse de mots clés, la présence des noms dans les médias. J'ai trouvé intéressant de la montrer car elle permet d'avoir une vision de l'offre, en comparaison avec la courbe précédente correspondant à la demande au niveau des médias.

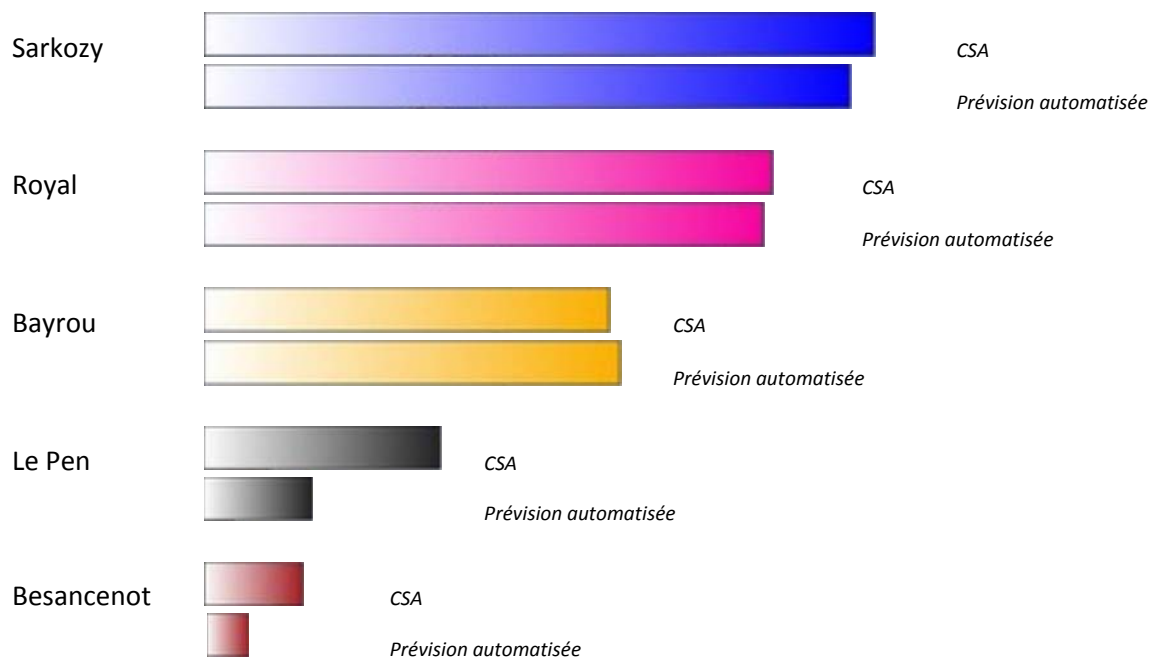


Premier constat, peu de piques correspondent réellement. Ceci est en partie dû au fait que les médias ont besoin d'un temps de rédaction et de publication. Mais qu'en partie seulement. Car on peut voir que certaines augmentations de publication dans les médias ne voient pas d'augmentation dans les recherches des internautes et inversement. C'est notamment le cas pour les deux premiers piques des médias, et les deux grands piques de Royal et Bayrou post-élection.

La fin du premier tour écoulé, voici les résultats du CSA à 20h50.

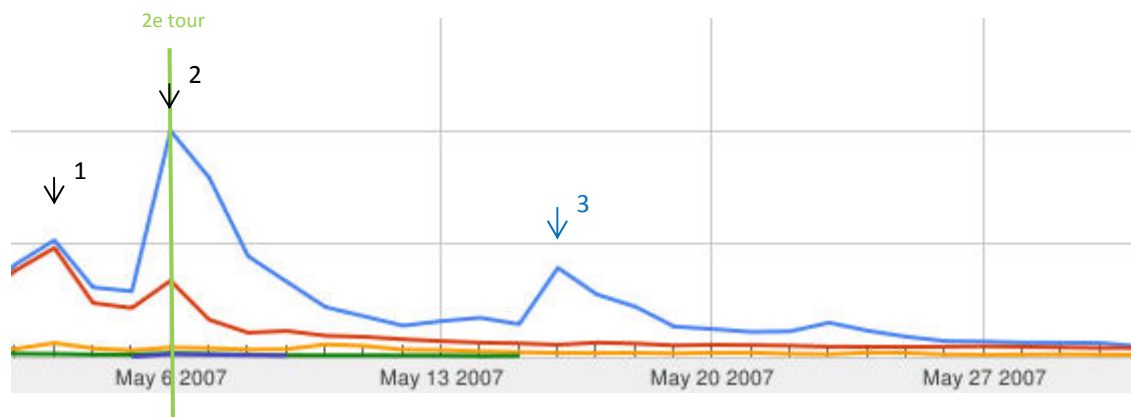


Comparons maintenant les résultats du CSA avec ceux obtenus par la recherche automatisée sur les requêtes des internautes dans les moteurs de recherches.



Le résultat est surprenant! Les approximations de “Le Pen” et de “Besancenot” sont assez large. En revanche, les résultats des trois leaders sont extrêmement proche des résultats réels (source CSA). C’est effectivement une telle précision dans la prédiction qui ont attiré mon attention sur cette méthode d’analyse automatisée. Devant un tel résultat, je n’ai pu que réitérer cette analyse au second tour pour m’assurer qu’elle n’était pas le fruit du hasard. Voici donc la meme analyse mais cette fois-ci pour le second tour de la présidentielle.

Sur cette courbe, nous avons trois piques majeurs.



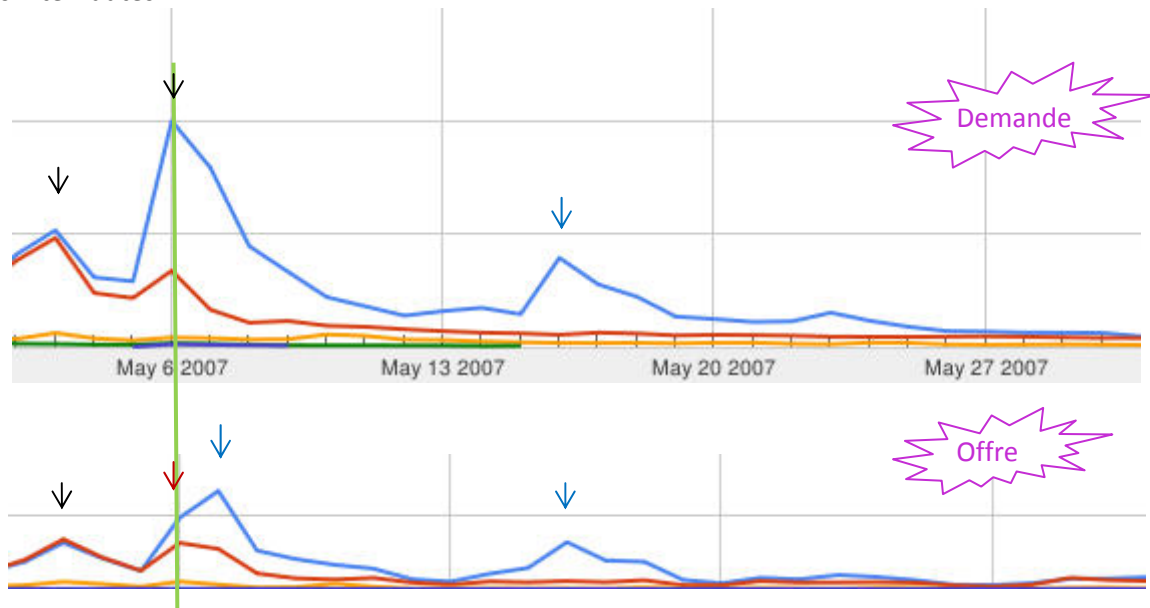
**Pique 1** : correspond au face à face musclé entre Royal et Sarkozy. Ce débat a vu une formidable audience. Très suivi, même à l’internationale.

**Pique 2** : élection du deuxième et dernier tour dans la course à la présidentielle.

**Pique 3** : départ de Jacques Chirac qui laisse sa place à l’Elysée au nouveau président de la république Française, Mr Nicolas Sarkozy.

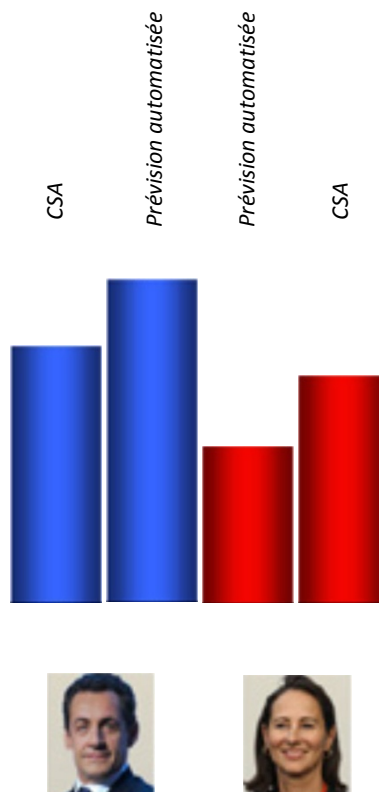
Nous pouvons également annoter une brève remontée le 23 Mai pour la requête “Sarkozy” probablement dû à la sortie dans les médias de la photo officielle de Nicolas Sarkozy. Il a choisit comme arrière plan non plus le parc de l’Elysée comme l’avait fait Chirac, mais la bibliothèque de l’Elysée.

Voici la même analyse que précédemment, sur la différence entre l'offre des médias et la demande des internautes.



Cette fois-ci, les courbes sont très proches. La légère suprématie de Sarkozy sur la courbe de la demande au niveau du pique pré-électoral n'est pas tout à fait représentée à sa juste valeur sur la courbe de l'offre. Les médias proposent un peu plus de Royal que de Sarkozy. On peut dire que les médias ont plutôt bien répondu à la demande, avec parfois le petit décalage de rédaction (premier pique de Sarkozy post-élection). Même la légère remontée du 23 Mai est représentée.

Voici ce que cela donne pour les prévisions nationales sur le résultat présidentiel du second tour.



Certe, la précision n'est pas aussi grande que pour le premier tour, mais la prévision automatisée a réussi à prédire qui serait élu nouveau président de la république Française le 6 Mai 2007 et ce, plusieurs jours à l'avance.

En psychologie Sociale, on parle bien souvent du décalage entre ce que les gens disent, ce qu'ils pensent et ce qu'ils font. Cette étude à le mérite de montré non pas comme les sondages, ce que les gens disent ou ce que les gens pensent, mais bien ce que les gens font. De plus, ils ne subissent pas de pression directe de l'interlocuteur qui lui pose généralement la question du sondage. Ici, il est libre de toutes pressions sociales directes face à la manière dont nous avons recueilli ces résultats statistiques.

Cette méthode pourrait permettre aux médias de connaître les attentes au jour le jour de leurs lecteurs. Et ce, pouvant être opéré géographiquement (Par pays, régions, départements et même par villes). Pouvant ainsi construire des offres plus en adéquation avec la demande. Elle pourrait surtout permettre de prédire le résultat des votes dans une campagne comme celle-ci (applicable également aux législatives, cantonales etc...).

Voici une analyse automatisée qui n'a suffi que d'une connexion internet, de quelques outils simples et d'un soupçon de matière grise.

Christophe Jorge, 8 Mai 2007