

Explorer les entités nommées avec iSeek

[Retour au sommaire de la lettre](#)

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

iSeek est un nouveau moteur de recherche qui base son originalité sur l'exploration des entités nommées (noms de personnes, de lieux, d'organismes) et l'analyse des relations entre elles. La société conceptrice de cet outil est déjà en possession d'une quarantaine de brevets en traitement automatique du langage et en intelligence artificielle. Présentation d'un outil très intéressant et qui devrait bientôt connaître une version française encore plus attrayante...

Pas une semaine sans que de nouveaux moteurs de recherche apparaissent. Si nombre d'entre eux n'apportent rien et disparaissent très vite de la circulation, d'autres en revanche sont réellement intéressants, tant dans les fonctionnalités qu'ils proposent que dans ce qu'ils nous permettent d'entreapercevoir de l'avenir de la recherche sur Internet. Tout comme Evri, présenté récemment, iSeek rentre pleinement dans cette seconde catégorie.

Mis en ligne depuis peu par Vantage Linguistics, une société américaine qui possède une quarantaine de brevets en traitement automatique du langage et intelligence artificielle, iSeek (www.iseek.com) est un moteur qui dispose de nombreuses fonctionnalités originales que nous vous proposons maintenant de découvrir. Sachez toutefois qu'il est pour l'instant optimisé pour la langue anglaise. Vantage Linguistics proposant des modules de reconnaissance automatique des langues, on peut espérer que cette limitation évoluera rapidement.

Compréhension du langage naturel

La recherche en langage naturel n'est certes pas une fonctionnalité nouvelle mais iSeek le fait très bien. Ainsi, si l'on pose la question : "*Give me a list of the french presidents*", on obtient comme premier résultat une page de la Wikipedia tout à fait pertinente. Reste toutefois à se poser la question de l'avenir de la recherche en langage naturel. Les internautes en quête d'informations sont dorénavant familiers des recherches par mots-clés et les utilisent de mieux en mieux. Une étude de la société Hitwise datée de février montrait par exemple que les requêtes composées d'un ou deux mots ont baissé respectivement de 3 et 5% en un an, alors que les requêtes comportant 4 mots et plus ont sensiblement augmenté (<http://searchengineland.com/search-queries-getting-longer-16676>). Heureusement, iSeek permet aussi la simple recherche par mots-clés.

Découverte de relations entre entités nommées

Là où iSeek apporte du nouveau c'est dans sa manière d'utiliser les entités nommées. Pour chaque requête effectuée il détecte en effet des entités de 8 types :

- Topics : concepts qui émergent des pages de résultats
- People : nom de personnes
- Places : nom de lieux
- Organizations : organisations publiques ou privées, associations
- Dates : dates, années trouvées dans les résultats
- Abbreviations
- Date published : dates de publication des pages de résultat
- Sources : magazines ou journaux en ligne

Ainsi pour la requête "*Dave Snowden*" (chercheur gallois en gestion des connaissances) nous obtenons la page de résultats suivante :

The screenshot shows the iSEEK search results for the query "dave snowden". The interface includes a search bar at the top with the query and the iSEEK logo. Below the search bar, there are four main categories of results: People, Places, Organizations, and Date and Time. Each category is represented by a folder icon and a list of items with a count next to each item.

Category	Item	Count
People	Dave Snowden	100
	David Snowden	3
	Larry Prusak	3
	Mark Jackson	3
	Patrick Lambe	3
	Daniele Chauvel	2
	Jon Husband	2
	Ian Glendinning	1
	More	
Places	London	3
	Australia	2
	Europe	2
	Singapore	2
	UK	2
	Houston	1
	Israel	1
	Kuala Lumpur	1
	More	
Organizations	IBM	4
	Microsoft	3
	Edge Pte Ltd	2
	IBM Institute	2
	Keynote	2
	Knowledge Management	2
	Business & Commerce	1
	Conference & Exposition	1
	More	
Date and Time	2007	
	2008	
	2003	
	2000	
	March 4, 2009	
	February 2009	

A partir de là, nous allons pouvoir affiner nos résultats en croisant entre elles les entités nommées. Cette opération ne se fait pas au hasard et l'on pourrait parler de filtrage semi-guidé puisqu'iSeek indique visuellement si une entité est en rapport avec une autre. Cela permet de ne pas perdre de temps à explorer des pistes qui ne mèneraient nulle part. Un exemple : pour savoir pourquoi le nom de Larry Prusak est associé trois fois à celui de Dave Snowden. Un clic sur son nom et on obtient les deux pages relatives dans lesquelles on apprend qu'ils ont participé à une interview portant sur l'avenir du *knowledge management* intitulé "Is KM dead ?".

The screenshot shows the iSEEK search results for the query "Larry Prusak". The interface includes a search bar at the top with the query and the iSEEK logo. Below the search bar, there are two main categories of results: Topics and People. Each category is represented by a folder icon and a list of items with a count next to each item.

Category	Item	Count
Topics	Larry Prusak Dave Snowden Patrick Lambe	2
	1145 Pm Permalink Comments 7	0
	3rd Contactivity Event In Israel	0
	Articles By Dave Snowden	0
	Basketball Jackson	0
	Cynefin Centre	0
	Dave Snowden's Blog	0
	Jon Husband Interview	0
	More	
	People	Dave Snowden
Larry Prusak		2
Patrick Lambe		2
Daniele Chauvel		1
David Snowden		1
Ian Glendinning		1

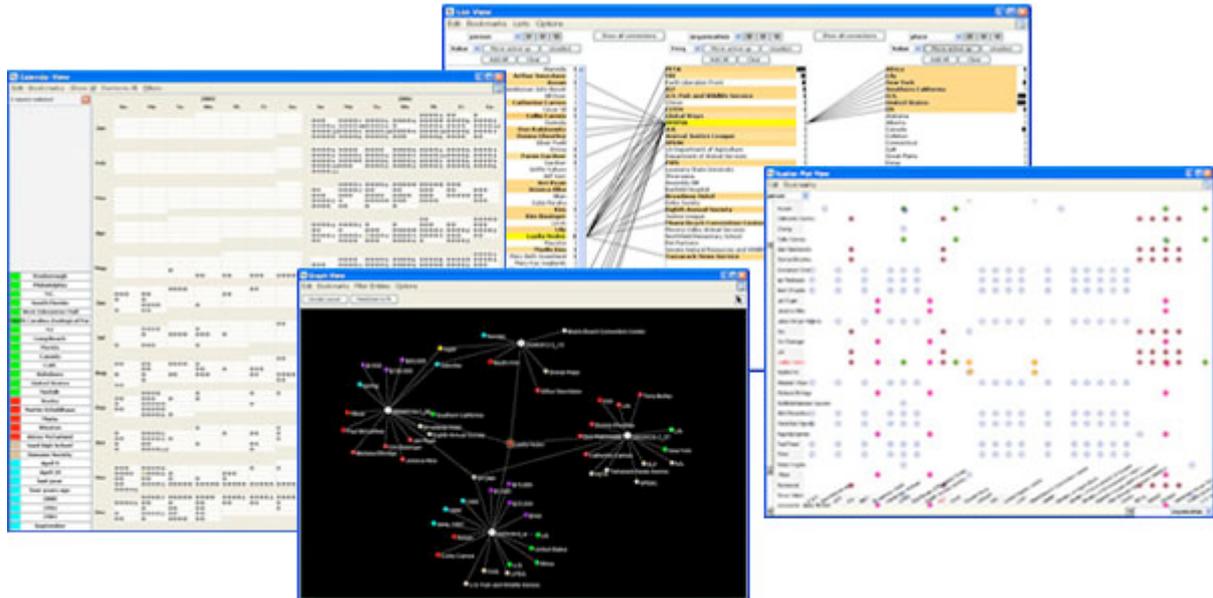
Below the search results, there is a section titled "Viewing 1-2 of 2 total results". It contains two search results:

- The Business of Knowing: Is KM Dead? Great video of Dave Snowden ...**
Jul 9, 2008 ... Great video of **Dave Snowden** and Larry Prusak. This is a fantastic video
thebusinessofknowing.blogspot.com/2008/07/is-km-dead-great-video-of-dave-...
- Is KM Dead? Larry Prusak, Dave Snowden, Patrick Lambe**
Jul 31, 2008 ... It is actually a video interview that Patrick Lambe, author of the super fir
it.toolbox.com/blogs/elsua/is-km-dead-larry-prusak-dave-snowden-patrick-l...

Mais ce n'est pas tout. Si on regarde les entités nommées actives (non grisées), on s'aperçoit qu'apparaît également le lieu de l'interview, Kuala Lumpur, ainsi que l'année où elle a eu lieu, 2008.

Tout ce qui est lié sur les pages de résultats apparaît ici d'une manière ou d'une autre et je vais ainsi pouvoir découvrir et explorer des pistes autour du sujet de ma requête afin de découvrir des relations que je n'imaginai pas. Ainsi si on clique ici sur IBM, on apprend que Dave Snowden a précédemment dirigé le laboratoire *Cynefin Center for Organisational Complexity*, créé par cette compagnie pour l'aider à mieux comprendre les attentes des organisations en terme de gestion des connaissances.

Si, à notre connaissance, iSeek est le seul à proposer gratuitement ce type de fonctionnalités pour rechercher dans le contenu non-structuré du web, il faut noter la présentation récente par des étudiants de l'Université de Georgia Tech de Jigsaw (<http://www.cc.gatech.edu/gvu/ii/jigsaw/>). Il s'agit d'une application d'aide à l'analyse de l'information, destinée aux services de renseignement et de polices qui pousse les principes d'iSeek encore plus loin.



Autres fonctionnalités

iSeek propose d'autres fonctionnalités moins déterminantes mais qui contribuent à faire de lui un moteur de recherche à la fois personnalisable et complet. Outre le fait qu'il est possible de modifier d'un clic l'emplacement de la fenêtre de présentation des entités nommées (iViews), on peut aussi s'enregistrer et créer un compte MySeek. A l'instar de ce que propose Netvibes, il sera alors possible de se créer un portail d'information personnalisé composé d'actualités, de flux RSS et de widgets variés. Le fait d'avoir un compte permet également d'effectuer un certain nombre d'actions pour chaque résultats :

The Business of Knowing: Is KM Dead? Great video of Dave Snowden ...
 Jul 9, 2008 ... Great video of **Dave Snowden** and Larry Prusak. This is a fantastic video, which actually hits the nail on the head
thebusinessofknowing.blogspot.com/2008/07/is-km-dead-great-video-of-dave-... Average Rating (1 votes)

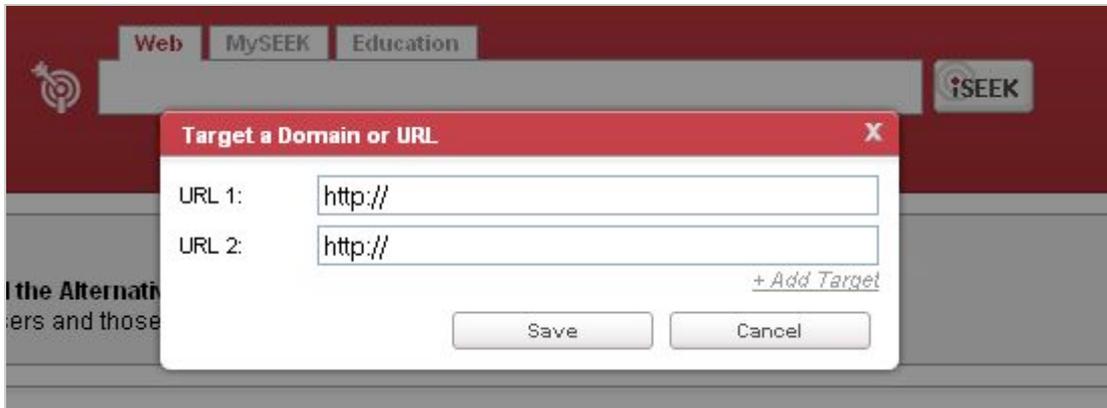
- Etoile : enregistrement d'un résultat dans ses favoris MySeek.
- Enveloppe : envoi du résultat par email.
- 5 étoiles : possibilité de noter un résultat de 1 à 5.

La première de ces fonctionnalités est très intéressante puisqu'elle permet d'entrer des résultats dans des dossiers personnalisés dans lesquels il sera ensuite possible de lancer des recherches ciblées.

Target Your Results By:		
MySEEK My Favorites 2	Topics 1145 Pm Permalink Comments 7 1 Articles By Dave Snowden 1 Dave Snowden's Blog 1 Knowledge Work And Knowledge Management 1 3rd Contactivity Event In Israel 0 Basketball Jackson 0	People Dave Snowden 2 David Snowden 0 Ian Glendinning 0 Larry Prusak 0 Patrick Lambe 0

On pourra ainsi se créer autant de dossiers thématiques que nécessaire et y rechercher par mots-clés avec l'avantage d'une présentation des résultats enrichie grâce à iSeek.

Notons enfin qu'en cliquant sur la cible rouge à gauche de la fenêtre de recherche vous pourrez limiter votre recherche à un ou plusieurs sites web. C'est en fait l'équivalent de l'opérateur "Site : " de Google, avec ici la possibilité d'en ajouter plusieurs, ou encore d'un Google CSE qui ne serait pas permanent.



Utilisée de cette manière, la détection d'entités nommées pourrait s'avérer être un concurrent de taille à la catégorisation automatique telle qu'elle est pratiquée par Clusty ou Xsearch.

Christophe Deschamps

Consultant et formateur en gestion de l'information.

Responsable du blog Outils Froids (<http://www.outilsfroids.net/>)

Réagissez à cet article sur le blog des abonnés d'Abondance :
<http://abonnes.abondance.com/blogpro/2009/03/explorer-les-entitees-nommees-avec.html>