

N° 279 - Février 2011

BASES

Connaître et bien utiliser
les banques de données

SOMMAIRE

Méthodologie

- Comment mieux innover grâce aux banques de données, pp.1-4

Actualités

- STN : la banque de données Beilstein change de nom, p.4

Web invisible

- L'Ompi lance une banque de données sophistiquée sur les marques internationales, pp.6-7
- Petale : l'écolosion de l'archive ouverte des universités de Lorraine, p.8
- Agorha rassemble les bases d'histoire de l'Institut national d'Histoire de l'art, p.9
- Corporama réunit les informations du Web sur une entreprise, p.10
- Nouvelle interface pour DOAJ.org, p.11

Agenda

- Web visible, Web invisible, Web 2.0, p.11
- Veille sur le Net, p.11

Comment mieux innover grâce aux banques de données

François Libmann

Rappelons brièvement une définition communément admise de l'**innovation**. Il s'agit de la « socialisation », c'est-à-dire la **diffusion à une échelle significative, d'une invention** – pure création de l'esprit qui peut être matérialisée, comme une molécule synthétique ou une formule comme $E=mc^2$, mais dont la caractéristique est qu'elle n'existait pas auparavant – **ou d'une découverte** – mise à jour d'une chose qui existait, mais que personne n'avait encore vue.

Il est évident que l'innovation a existé bien avant les banques de données, dont les premières ont été développées dans le courant des années 60 et commercialisées au début des années 70 (Dialog a été lancé commercialement en 1972).

Mais le développement des banques de données – et, plus tard, d'Internet – a contribué très certainement à l'accélération du processus d'innovation.

On rappellera d'ailleurs que d'une certaine façon, les banques de données ont été vues dès leur conception comme un moyen d'accélérer l'innovation.

En effet, c'est le lancement en 1958 du premier Spoutnik par l'Union Soviétique qui a conduit les Américains à s'interroger sur leur retard dans la course à l'espace. Une des raisons identifiées avait été le mauvais accès des chercheurs à l'information scientifique et technique. C'est dans le cadre de la résolution de ce problème que la Nasa et la société Lockheed ont été mêlées de très près à la création de Dialog.

Complexité et interactivité

Le premier apport des banques de données a consisté à améliorer considérablement la pertinence des DSI (Diffusions sélectives de l'information), en permettant d'utiliser des équations de recherche complexes au lieu de se contenter d'un nombre très limité de critères, comme c'est le cas pour des surveillances manuelles.

Quant à la recherche ponctuelle d'information, elle a bénéficié non seulement de la possibilité de poser des questions complexes, mais aussi de l'interactivité qui permet d'enchaîner plusieurs recherches modifiées, après analyse des résultats de la précédente.

BASES PUBLICATIONS

Commission paritaire 1110 I 87287 • ISSN 0765-1325

SARL de presse au capital de 10 000 €

RC Paris B335325007

Siège social : 27, rue de la Vistule - 75013 Paris

Tél.: 01 45 82 75 75 • Fax : 01 45 82 46 04

contact@bases-publications.com

Directeur de la Publication : François LIBMANN

Rédactrice en Chef : Béatrice FOENIX-RIOU



Innovation et banques de données *suite*

La possibilité d'effectuer des recherches complexes s'est retrouvée, à quelques nuances près, dans tous les langages d'interrogation développés par les concurrents (DataStar, racheté par Dialog, Questel ou STN) et existe toujours.

Preuve de son importance, la nouvelle plateforme Proquest Dialog, destinée à remplacer Dialog et DataStar par une seule interface, conservera la possibilité d'utiliser un langage d'interrogation sophistiqué, en le complétant par quelques possibilités devenues familières avec l'usage du Web.

Si l'on veut éviter de réinventer la roue, dans une démarche sur des sujets très technologiques, il est nécessaire que les recherches sur l'état de l'art puissent être faites à partir d'outils puissants en soi, mais qui permettent aussi d'utiliser la valeur ajoutée des différentes banques de données, telle que l'indexation qui leur est propre.

La recherche d'état de l'art est un cas classique, qui remonte pratiquement à l'origine des banques de données, mais il existe de multiples autres façons d'utiliser les banques de données – et, plus généralement, les sources d'information électroniques – dans une démarche d'innovation.

On notera, en le regrettant, le manque d'appétence de la plupart des chercheurs pour l'utilisation d'outils sophistiqués, qui est confirmé par l'usage massif qu'ils font de Google Scholar.

Pourtant, l'apprentissage de l'utilisation de ces outils prend infiniment moins de temps qu'il ne leur en a fallu pour devenir expert dans leur domaine ou tout simplement apprendre l'anglais, mais c'est ainsi.

C'est la raison pour laquelle se sont développés, sur Internet notamment, des outils fonctionnant sur une autre logique et plus simple à interroger par l'utilisateur final.

Dans le même temps – cause ou effet, ou les deux à la fois –, on a observé, dans nombre d'entreprises ou d'organismes, une baisse significative – voire leur disparition – du nombre de professionnels de l'information capables d'utiliser les outils sophistiqués.

Emblématique de la nouvelle génération d'outils, Scopus ou Web of Science proposent de démarrer par une recherche simple, sinon très simple, et permettent de trouver rapidement l'auteur le plus connu de la publication la plus citée, dans un domaine donné.

Si cela peut être utile pour une première approche d'un domaine nouveau, on peut penser que dans les autres cas, cette démarche est à l'opposée d'une démarche d'innovation puisque, quelque part, c'est le sommet du conformisme.

D'autre part, ces outils mettent en valeur les citations et permettent de naviguer dans celles-ci dans un sens ou dans l'autre. Il est facile de se laisser emporter dans la navigation, de citations en documents et réciproquement. Mais on peut considérer alors que l'on perd la maîtrise de sa recherche, en se laissant mener par cet enchaînement, tout en priant le «dieu sérendipité» pour que des informations inattendues (par définition) mais pertinentes apparaissent.

On observe que cette démarche est très proche de celle qui permet de naviguer de lien en lien sur le Web, et qui est devenue une façon de faire extrêmement prégnante.

Bien qu'également destinés au chercheur, mais fonctionnant sur une autre logique, on a vu aussi apparaître – notamment dans le domaine de la pharmacie – des produits «verticaux» spécifiquement dédiés à un domaine.

Ces produits apportent des informations et outils qui peuvent être sophistiqués, à mesure de l'avancement de la démarche du chercheur, afin de l'optimiser.

Sciverse d'Elsevier (voir Bases n°274, septembre 2010) est ainsi dans cette tendance.

Dans tous les cas, la question de fond est de savoir si le chercheur reste maître du pilotage intellectuel de sa démarche de recherche/innovation, ou s'il se laisse guider par les outils, ce qui est un risque, même si des résultats peuvent être au rendez-vous.

L'intégration d'outils d'analyse

De plus en plus fréquemment, des outils d'analyse plus ou moins sophistiqués sont proposés en complément des résultats. Le plus classique est le classement des auteurs, des années de publication ou des affiliations.

Ces analyses peuvent être réalisées facilement avec des commandes comme RANK de Dialog ou ANALYZE de STN, mais les résultats sont de plus en plus souvent présentés automatiquement, comme sur Scopus ou Factiva.

Mais ce ne sont que des analyses élémentaires.



On trouve ainsi chez Orbit (Questel) avec les outils d'Intellixir ou STN avec AnaVist des outils plus sophistiqués, qui permettent par exemple d'identifier des réseaux de chercheurs ou un espace vacant pour des dépôts de brevets.

Ces outils favorisent-ils ou ont-ils favorisé l'innovation ?

On peut penser, comme on l'a dit au début, qu'ils en ont tout au moins accéléré les processus, et étendent le champ de vision des chercheurs.

Mais comme il est très difficile, voire impossible, de comparer les résultats des travaux de chercheurs ayant utilisé en partie ces outils avec ceux de chercheurs ne les ayant pas employés, il est très délicat de trancher précisément.

Les banques de données comme élément d'aide à la créativité

Pour la simplification de l'exposé, nous prendrons un exemple dans le domaine des produits de grande consommation ; mais cela est parfaitement transposable à des produits plus techniques.

Voici un exemple qui permet d'identifier rapidement des concepts que l'on peut ensuite injecter dans le domaine dans lequel on souhaite innover.

On peut, par exemple, choisir la banque de données PROMT sur Dialog (fichier n° 16), qui est un classique de l'information

business et qui présente la particularité d'avoir une indexation très développée.

On peut, par exemple, prendre le code événement 336 (*Product introduction*) et le croiser avec le secteur alimentaire.

On écrira tout simplement (dans Dialog) *SEC=336 AND PC = 20*.

Cela donne ... 78 666 réponses depuis 1990. On n'est, évidemment, pas obligé de les visualiser toutes. On commence par visualiser les 20 plus récentes dans le format (gratuit) titre + indexation, qui donne une idée de l'information.

On sélectionne celles qui semblent présenter le plus d'intérêt et on les visualise alors en format complet payant, évidemment.

C'est ainsi qu'on peut retenir le lancement au Royaume-Uni d'une saucisse allégée pour le petit-déjeuner, de pâtes protéinées ou de «Gourmet Burger», c'est-à-dire des hamburgers haut de gamme.

Dans cet exemple, les concepts d'aliment allégé ou protéiné, de même que la montée en gamme ne sont pas nouveaux, mais c'est un moyen commode d'identifier rapidement des concepts et de voir ce que donne leur mise en relation/application dans le domaine ou sur les produits qui font l'objet d'une recherche d'innovation.

Cette mise en relation est très libre.

On peut en effet chercher à alléger mais aussi à alourdir, rajouter des protéines mais aussi d'autres choses à un produit banal, etc.

Explorer les technologies liées à un concept

Parmi les grandes tendances du moment, le développement de l'interactivité avec le consommateur tient une place importante, que l'on illustrera par la multiplication des «tag codes» sur les publicités, permettant à un consommateur équipé de smartphone d'obtenir immédiatement de l'information complémentaire.

Si l'on veut identifier les technologies et les acteurs intervenant dans ce secteur, on pourra utiliser plusieurs types de sources :

- la presse française dans EDD et internationale dans Factiva et/ou LexisNexis ;
- le Web ouvert ;
- les banques de données brevet.

Dans la mesure où il s'agit d'une recherche ouverte, exploratoire, la stratégie n'a pas besoin d'être sophistiquée, ce qui engendrera un taux de bruit relativement important mais permettra quand même d'identifier des technologies et entreprises intéressantes.

On enrichira bien sûr les termes de recherche avec ceux trouvés dans les documents pertinents.

On ira aussi visiter les sites web des entreprises identifiées et on relancera des recherches à partir de leur nom.

On s'intéressera particulièrement à des articles faisant le tour de la question ou, au moins, d'une technologie, telle que « Augmented reality ads on the rise around Europe », pour prendre l'exemple de la réalité augmentée. C'est dans ce type d'articles que l'on trouvera des noms d'entreprises actives dans le secteur, ainsi que des exemples d'applications.

Au cours de ces recherches, on découvrira que si les technologies de réalité augmentée sont particulièrement utilisées pour les interactions avec le consommateur, il existe aussi d'autres technologies comme les écrans tactiles.

Pour une problématique telle que celle-ci, les sources électroniques sont un outil particulièrement puissant pour identifier rapidement un grand nombre d'acteurs et de technologies.

Le plus long sera d'analyser les informations trouvées, puis ensuite de prendre contact avec une sélection d'entreprises, si l'on souhaite utiliser ces technologies.

On peut aussi, bien sûr, utiliser ces informations pour déterminer un champ de recherche dans ce domaine qui ne soit pas déjà occupé, si l'on souhaite en devenir un acteur.



Résoudre un point de blocage

Dans une démarche d'innovation, il arrive que l'on définisse un concept prometteur, mais qu'il manque un composant, essentiel ou non, pour le réaliser correctement.

L'interrogation des banques de données de diverses natures ou du Web pourra alors s'avérer intéressante, pour tenter d'identifier des entités susceptibles de détenir des technologies constituant la partie manquante du puzzle ou des experts / expertises susceptibles de mener vers la clé du problème.

On pourra s'intéresser, en particulier, aux offres de technologies proposées par Oséo ou diverses universités.

On pensera aussi aux licences disponibles dans la partie Legal Status d'Inpadoc/Fampat (voir Bases, décembre 2010, n° 277 : «Comment identifier les licences en utilisant les banques de données»).

Quand une innovation devient enfin réalisable

Dans le paragraphe précédent, nous avons implicitement supposé que la solution, la pièce – en général technologique – manquante, était à la portée d'une recherche ponctuelle ou disponible rapidement.

Nous voudrions citer, un exemple dans lequel le décalage est beaucoup plus grand dans le temps, entre la conception d'une innovation et le moment où elle devient réalisable.

Nous citons dans notre livre «Méthodes pour innover et se diversifier», publié en 1980 aux Editions d'Organisation – et, donc, écrit avant –, l'exemple souvent utilisé dans les formations à La Créatique (une méthode de créativité d'origine française, constituant une logique de la découverte). Le thème choisi était : « Quelles innovations peut-on apporter à la carte routière ? ».

Un des outils de la méthode permettait d'imaginer deux types de solution :

■ des réponses ne mettant pas en cause le «principe de fonctionnement» de la carte routière. C'est le cas, par exemple, lorsque l'on ajoute sur la carte des informations d'un type spécifique : des restaurants ou certains types d'entre eux, des hôtels d'une certaine chaîne ou des stations services d'une certaine marque. Tout cela était facile à faire et a été réalisé depuis longtemps ;

■ des réponses remettant en cause l'actuel «principe de fonctionnement» de la carte routière, mais qui répondent à la même attente que l'on peut, par exemple, exprimer de façon abstraite par l'expression «maîtriser un espace inhabituel».

On passera ainsi de l'objet / principe actuel (de la carte routière papier) «support d'information sous forme de représentation imprimée préétablie embarquée», à «réservoir d'informations sous forme de système audiovisuel en temps réel par rapport à des besoins connus ou non».

On constatera que cette formulation décrit à peu près exactement les systèmes de GPS disponibles couramment aujourd'hui. Or, les briques technologiques nécessaires à leur réalisation à grande échelle n'existaient pas en 1980.

C'est donc en ayant conscience de ce type de solution qu'une veille suffisamment large et bien orientée permet, si l'on a présent à l'esprit l'innovation «en attente», de détecter l'apparition des briques nécessaires puis d'être, si on a les moyens de le réaliser, parmi les premiers ou le premier à mettre en œuvre ce type de solution, et de bénéficier alors d'une «prime à l'innovation».

Un concept commun : le mouvement

L'innovation est mouvement. Pour être plus précis, l'innovation est un cheminement, cheminement mental des personnes participant au projet et cheminement en terme d'avancement du projet lui-même.

La recherche régulière d'information (veille) a aussi une dimension temporelle et les allers-retours entre la démarche d'innovation et celle de la recherche d'information sont et doivent être très fréquents.

En effet, à mesure que l'on avance dans la démarche d'innovation, les sujets d'intérêt évoluent naturellement.

Réciproquement, les informations recueillies ont un impact sur la démarche d'innovation.

STN : La banque de données BEILSTEIN change de nom pour s'appeler ReaxysFile.

Cette banque de données contient, pour des substances organiques, des documents évalués issus du «*Handbook of organic chemistry*» publié par Friedrich Beilstein, ainsi que des données issues des 176 plus importants périodiques en chimie organique, couvrant la période de 1771 à nos jours.

Pour chaque substance, on trouve un nombre important de données, qui sont à la fois recherchables et visualisables.