

MÉTHODOLOGIE D'ANALYSE DE L'INFORMATION

L'ANALYSE DE BREVETS « FUEL CELL » ANNEE 2000 – 2002

Ce document est une étude courte, réalisée à titre d'exemple de notre méthodologie d'analyse des informations structurées. Elle a pour but de montrer, d'une part, comment en analysant les liens entre différentes informations, on se donne la possibilité de découvrir d'autres informations, jusque là cachées. On est alors en mesure d'aller plus loin dans l'appréhension des questions qui nous préoccupent. Et d'autre part, comment cette méthode d'analyse de l'information peut être utilisée pour traiter un corpus de brevets.

Ce document commence directement par l'étude menée sur les brevets. Nous y expliquons :

- le choix du corpus
- le modèle de données utilisé pour analyser ce corpus
- le traitement effectué sur ces données
- enfin, nous montrons comment nous analysons ce corpus et ce que cette analyse apporte.

La méthode ayant été démontrée par un cas précis, nous discuterons d'autres applications possibles de cette méthode.

Commentaires

Ce document n'est qu'un exemple développé pour ouvrir la discussion et la réflexion sur une méthodologie d'analyse de l'information structurée. Les résultats décrits dans ce document n'ont pas vocation à être utilisés dans un cadre opérationnel. Ce document est une étude courte, réalisée à titre d'exemple de notre méthodologie d'analyse des informations structurées. Elle a pour but de montrer, d'une part, comment en analysant les liens entre différentes informations, on se donne la possibilité de découvrir d'autres informations, jusque là cachées. On est alors en mesure d'aller plus loin dans l'appréhension des questions qui nous préoccupent. Et d'autre part, comment cette méthode d'analyse de l'information peut être utilisée pour traiter un corpus de brevets

ETUDE DE BREVETS : CHOIX DU CORPUS ET DES ELEMENTS ANALYSES

L'information brevet

Les brevets déposés dans les bases de données nationales ou internationales constituent une source d'information technique bien connue. Chaque fiche informatique décrivant un brevet est structurée et présente un certain nombre d'éléments identifiables d'information, comme :

- la référence du brevet
- sa date de dépôt
- le nom des inventeurs du brevet
- le nom des entreprises déposantes
- les codes internationaux ou nationaux décrivant les thèmes du brevet
- le nom des cabinets d'avocats qui ont traité le dépôt
- etc.

Particularité de l'information structurée

Ces bases de données fournissent de l'information structurée. C'est à dire que chaque type d'information contenu dans une fiche voit sa position et son format clairement identifiés et uniques au travers de la base de données. De fait, il est facile d'extraire ces informations par l'application de scripts informatiques.

Utilisation de l'information brevet

Les informations comme le nom des inventeurs, des entreprises, des cabinets d'avocats peuvent être lues comme des informations de collaboration : qui a inventé avec qui, quel cabinet travaille avec quelle entreprise, etc.

Nous proposons dans cette démonstration de ne conserver des informations brevets que les champs suivants :

1. la référence du brevet
2. son titre
3. le nom de ses inventeurs
4. le nom de ses déposants (s'il y a lieu)

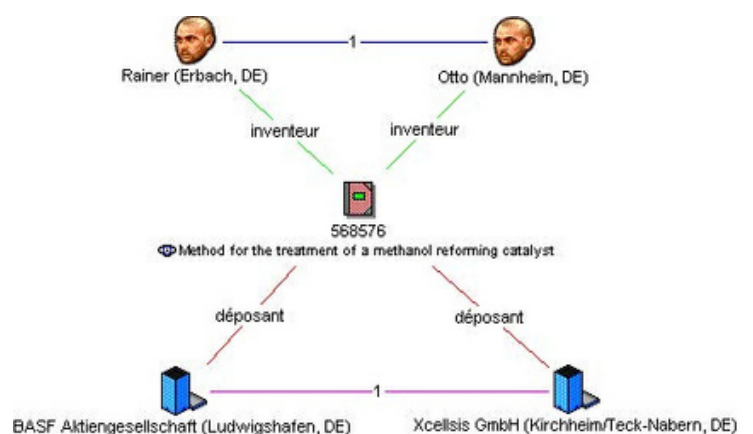
Nous allons utiliser ces informations pour faire ressortir les réseaux de collaboration qui lient chercheurs et entreprises.

Notre modèle de données

Les données que nous étudions présentent trois 'objets' définis de la façon suivante :

- le brevet proprement dit identifié de manière unique par son numéro, et caractérisé par son titre
- l'inventeur, identifié de manière unique par son nom, et le plus souvent par l'ensemble (nom + adresse)
- la société déposante, identifiée de manière unique par son nom, et parfois par l'indication de son adresse

Nous représentons les liens entre ces objets de la façon suivante :





Représente un brevet unique. Il est identifié par son numéro. Sous ce numéro apparaît le titre du brevet



Représente un inventeur unique. Il est identifié par son nom de famille, et bien souvent aussi par son prénom et une indication d'adresse (pays ou état pour les USA, parfois une ville)



Représente une entreprise unique (parfois une université ou une association). Elle est identifiée par son nom et souvent par une indication d'adresse (pays ou état pour les USA, parfois une ville)

Entre ces entités existent des liens :

1. Entre deux inventeurs, le lien indique le nombre de brevets que ces deux personnes ont inventé conjointement.
2. Entre deux entreprises, le lien indique le nombre de brevets que ces deux entreprises ont déposé conjointement
3. Entre un brevet et un inventeur, le lien précise simplement 'inventeur'
4. Entre un brevet et une entreprise, le lien précise simplement 'déposant'

Choix du corpus

Pour effectuer cette démonstration de notre méthodologie d'analyse de l'information, nous avons choisi :

- d'extraire les informations de la base de données des brevets américains, base de données gratuitement exploitable sur le site <http://www.uspto.gov/>
- d'étudier les brevets contenant le terme « fuel cell » et déposés dans les années 2000 et 2001
- de ne garder que les brevets, largement majoritaires, qui font apparaître une ou plusieurs entreprises déposantes.

Autrement dit nous interrogeons la base de données <http://www.uspto.gov/> avec l'équation : "fuel cell" AND APD/1/1/2000->12/31/2001

Nous obtenons 95 brevets. Sur ces 95 brevets, 89 sont déposés par des entreprises : ce sont ces 89 brevets que nous étudions ici.

ETUDE DE BREVETS : ANALYSE DES INFORMATIONS

Quelles entreprises déposent ?

Avant d'étudier les réseaux de collaboration entre entreprises et entre chercheurs, voyons quelles sont les entreprises qui ont déposé des brevets sur notre sujet :

- 66 entreprises ont déposé ces 98 brevets
- seules 8 de ces 66 entreprises en ont déposés plus de 1

7 Ford Global Technologies, Inc. (Dearborn, MI) □

7 General Motors Corporation (Detroit, MI) □

7 Honda Giken Kogyo Kabushiki Kaisha □ (Tokyo, JP) □

5 Fantom Technologies Inc. (Welland, CA) □

3 XCELLSIS GmbH (Kirchheim/Teck-Nabern, DE) □

2 DaimlerChrysler AG (Stuttgart, DE) □

2 Plug Power Inc. (Latham, NY) □

2 The United States of America as represented by the
Administrator of the (Washington, DC) 1

Graphe complet des collaborations

Nous importons ensuite dans notre outil d'analyse les éléments sélectionnés. Nous obtenons le graphe des collaborations entre entreprises et chercheurs, sur les brevets sélectionnés.

Ce graphe est le suivant :

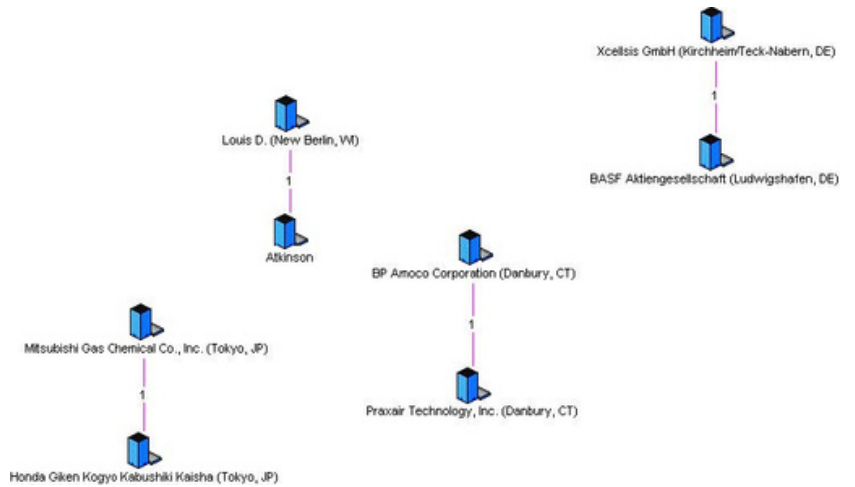


Quoique ne traitant qu'un nombre restreint d'informations (89 brevets) ce graphe est déjà complexe. C'est là où commence le travail d'analyse.

Collaborations directes entre entreprises

Une première question : peut-on identifier des entreprises qui ont déposé conjointement des brevets ?

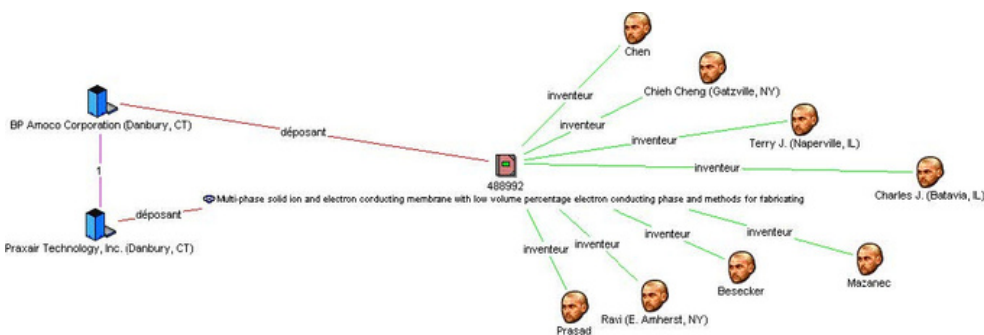
Nous identifions les réseaux de collaboration suivant :



Dans le cadre de cette démonstration nous choisissons l'une de ces collaborations et développons plus avant son analyse.

Collaboration BP - Praxair

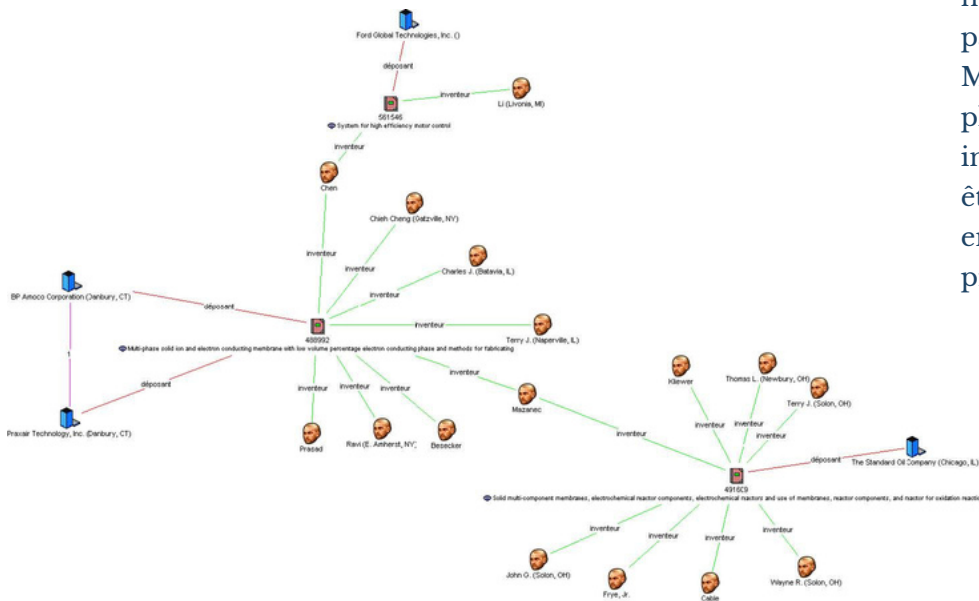
La collaboration identifiée entre BP Amoco Corporation et Praxair Technology Inc. se développe de la façon suivante :



Le sujet du brevet objet de cette collaboration est :

Multi-phase solid ion and electron conducting membrane with low volume percentage electron conducting phase and methods for fabricating.

Les chercheurs de cette collaboration (notons que les brevets ne permettent pas a priori de savoir qui travaillait directement pour quelle entreprise) ont-ils eux mêmes participé à d'autres dépôts de brevets ?



Quelles questions fait apparaître le graphe BP – Praxair ?

Nous identifions une collaboration sur un brevet entre BP Amoco Corporation et Praxair Technology Inc. Deux des chercheurs participant à cette collaboration ont par ailleurs participé chacun séparément à une autre collaboration.

- il s'agit en premier lieu d'un certain 'Chen' ayant travaillé sur le dépôt d'un brevet pour Ford Global Technologies Inc. (System for high efficiency motor control)
- il s'agit en second lieu d'un certain 'Mazanec' ayant travaillé sur le dépôt d'un brevet pour The Standard Oil Company (Solid multi-component membranes, electrochemical reactor components, electrochemical reactors and use of membranes, reactor components, and reactor for oxidation reactions)

Doit-on effectivement y voir des collaborations indirectes entre BP Amoco et Praxair d'une part, et Ford Global Technologies ou Standard Oil d'autre part ? Les liens entre BP et Standard n'étant pas nouveaux, que signifierait cette présence de Standard ? Le lien avec Ford est-il réel ou simplement dû au fait que le nom 'Chen' est très commun aux USA ?

Résultat de notre méthode

Nous voyons qu'une analyse poussée de l'information structurée au travers de notre méthode fait apparaître des questions et non pas des réponses. Mais ces questions étaient, pour la plupart, inconnues du client avant nos études. Et être en mesure de poser une question est le premier pas pour pouvoir y répondre !

Nous apportons ainsi un regard autre sur le sujet nous préoccupant.

Les méthodes classiques de recherche de l'information ouverte doivent permettre un premier défrichage. Ce défrichage doit permettre au Client d'identifier si oui ou non la question identifiée est importante, si les éléments identifiés méritent une surveillance particulière.

Enfin, il faut noter qu'il est souvent intéressant pour le Client de travailler une journée avec l'analyste. Le Client étant expert de son domaine technique, il est alors à même de guider l'analyste dans son exploration des informations obtenues.

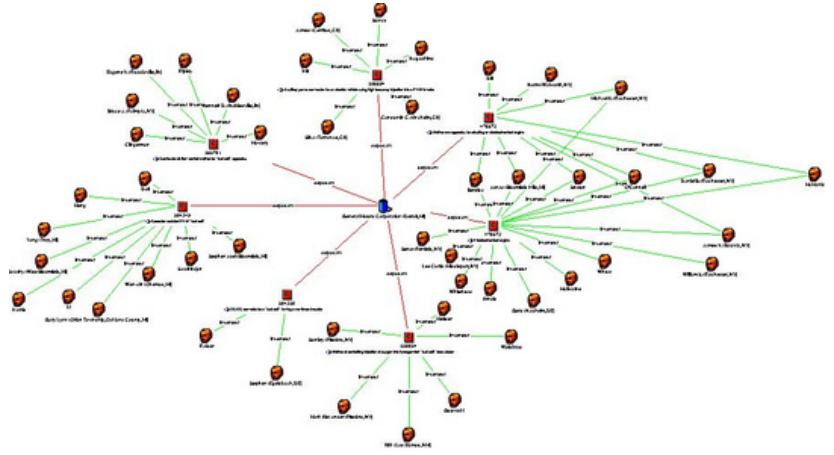
Autres axes d'analyses avec le même corpus

L'étude des collaborations entre entreprises n'est que le début de l'exploitation de notre corpus. D'autres recherches sont effectuées :

- Quels sont les associations de chercheurs les plus prolifiques ?
- Comment s'articulent les équipes de recherche au sein d'une entreprise ?

□ La richesse des questions que l'on peut poser augmente avec les paramètres que l'on prend en compte : date de dépôt, cabinets d'avocats, citations d'autres brevets, etc.

Par exemple au sein de General Motor, nous identifions les équipes et donc les personnes pivots :



METHODOLOGIE D'ANALYSE DE L'INFORMATION QUELQUES PISTES SUR L'UTILISATION DE LA METHODE

Principe général

L'exemple décrit précédemment permet de comprendre la philosophie générale de notre méthode :

Exploiter en masse la structure de certaines informations pour y identifier des relations significatives.

Il est évident que les brevets ne sont qu'un type d'information disponible, et que l'exploitation présentée précédemment n'en est lui même qu'un exemple.

Nous allons présenter quelques possibilités, parmi toutes celles qui peuvent être imaginées.

Exemples proposés

Dans les paragraphes qui vont suivre, nous présentons les exemples d'applications suivant :

- Etude des liens entre laboratoires
- Etude de la restructuration des équipes de recherche après fusion de deux entreprises
- Couplage de notre méthodologie avec des logiciels d'analyse des textes
- Analyse des stratégies des fonds de placement sur la bourse de New York
- Analyse des liens hypertextes entre sites Web
- Analyse des communications sur les forums, newsgroups, IRC, listes de diffusion.

Liens entre laboratoires

Pour rester dans les brevets, nous notons que la majorité des noms d'inventeurs viennent avec une adresse (un nom de ville et un pays, ou un état si nous sommes aux USA). Nous pouvons tracer le graphe des collaborations entre les inventeurs au sein d'une même entreprise, puis ne garder que ces adresses. Ensuite, une recherche d'information ou les données fournies par le Client, vont nous permettre d'identifier les laboratoires, centres de recherches ... de l'entreprise étudiée, et donc de faire le lien entre une adresse et un centre de recherche.

Le graphe représente alors les collaborations entre ces laboratoires au sein d'une même entreprise.

Fusion de deux entreprises

Lors de la fusion de deux entreprises en une seule entité, nous pouvons utiliser les informations de date pour étudier les restructurations entre les équipes de recherche et les laboratoires.

Pour cela nous créons :

- Un graphe des collaborations pour chacune des deux entreprises pour (l'année de la fusion - 1)
- Un graphe des collaborations pour chacune des deux entreprises pour (l'année de la fusion)
- Un graphe des collaborations pour la nouvelle entreprise pour (l'année de la fusion + 1)

En identifiant les inventeurs les plus prolifiques et en les suivant dans le temps, ou en se limitant comme expliqué ci-dessus aux centres de recherches, nous allons observer la restructuration des efforts de recherche de la nouvelle entité.

Couplage avec un outil d'analyse des textes

De nombreuses cellules de veille utilisent des outils d'analyse de textes. Hors des particularités techniques de ces outils, la plupart sert à regrouper les textes dans un plan de classement. Ce plan de classement peut être prédéfini (outils du type Arisem par exemple) ou non (outil de clustering du type d'Extended Technology Watch par exemple).

Ce regroupement effectué devient alors lui-même un élément de la structure de chaque texte : on peut effectivement adjoindre à chaque texte sa position dans le plan de classement. Toutes les méthodes présentées dans ce document peuvent alors être combinées à une pré-étude à l'aide de ces outils d'analyse de texte.

Nous obtenons alors les graphes de collaboration entre telles et telles entités, en fonction des sujets traités par les documents initiaux.

Analyse de la bourse de New York

Divers services gratuits sur Internet permettent d'obtenir pour toutes les entreprises cotées en bourse, à Wall Street ou au NASDAQ, la liste des principaux actionnaires de ces entreprises.

L'exploitation de ces données permet d'obtenir une cartographie de la bourse. A partir de celle-ci, il est possible par exemple d'identifier quel fond de placement possède la même stratégie d'investissement qu'un autre fond.

Analyse des liens Web

Lors d'une crise, il de plus en plus courant de voir apparaître des sites Web devant servir d'outils de pression pour tel ou tel acteur de cette crise. Ces sites font le lien vers d'autres sites, dans le but d'argumenter leur point de vue.

L'analyse des liens entre ces sites, associée éventuellement à des données comme celles concernant l'enregistrement des adresses Web, les e-mails, etc. peut permettre de remonter aux sites centraux dans l'argumentation, aux hiérarchies entre ces sites, à leur chronologie d'apparition.

Analyse des participations sur forums, listes de diffusion, IRC ...

Les forums, newsgroups, listes de diffusion, IRC ... sont autant de lieux de discussion, qui peuvent devenir des caisses de résonances idéales lors d'événements comme ceux décrits dans le paragraphe précédent.

L'adresse mail, l'adresse IP, le message auquel telle personne répond, la date ... font autant de structures démontrant des liens, et susceptibles d'être étudiées.

La surveillance de ces canaux et la récupération des informations pertinentes permettent alors d'identifier les réseaux de discussions, les leaders d'opinion dans le médium choisi, de remonter les chronologies ...