UNIVERSITÉ DE MONTRÉAL

ÉTUDE EXPLORATOIRE DE LA COMMERCIALISATION DES NANOTECHNOLOGIES AU CANADA : L'UTILISATION DE LA FOUILLE DE DONNÉES DE CONTENU WEB

CONSTANT RIETSCH

DÉPARTEMENT DE MATHÉMATIQUES ET DE GÉNIE INDUSTRIEL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION

DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES

(GÉNIE INDUSTRIEL)

SEPTEMBRE 2015

© Constant Rietsch, 2015.

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé :

ÉTUDE EXPLORATOIRE DE LA COMMERCIALISATION DES NANOTECHNOLOGIES AU CANADA : L'UTILISATION DE LA FOUILLE DE DONNÉES DE CONTENU WEB

présenté par : RIETSCH Constant

en vue de l'obtention du diplôme de : Maîtrise ès sciences appliquées

a été dûment accepté par le jury d'examen constitué de :

M. AGARD Bruno, Doctorat, président

Mme BEAUDRY Catherine, Ph. D., membre et directrice de recherche

M. ELIA Elie, Ph. D., membre

REMERCIEMENTS

J’aimerais remercier toutes les personnes qui ont contribué, de près ou de loin, à l’aboutissement de ce mémoire.

En premier lieu, il convient de souligner le rôle de ma directrice de recherche, Catherine Beaudry, pour m’avoir guidé tout au long de ce projet et m’avoir transmis des connaissances précieuses ainsi que pour le soutien financier qu’elle m’a octroyé durant ma maîtrise.

Je remercie également Bruno Agard pour son aide sur la mise au point de la méthodologie et plus particulièrement pour la comparaison des documents, Carl Saint-Pierre pour son aide sur l’analyse statistique, Ellias Zainea pour ses scripts et Benoît Forêt pour l’installation du serveur MySQL.

Je remercie également Bruno Agard et Elia Elie d’être membres du jury pour l’évaluation de ce mémoire.

Finalement, j’aimerais remercier ma famille, mes amis ainsi que l’ensemble des membres de la Chaire de recherche du Canada sur la création, le développement et la commercialisation de l’innovation pour leur soutien tout au long de ce projet.

RÉSUMÉ

Les entreprises comptent tous les jours plus sur Internet, partageant de nombreuses informations avec leurs clients, mais aussi à leurs investisseurs et à leurs associés. La majorité des entreprises partagent des informations telles que leurs produits, services, partenariats, activités de recherche et de développement sur leur site Internet. Toutes ces informations sont autant d'opportunités d'analyse pour autant que nous puissions les récupérer et les interpréter.

Dans ce contexte, est-ce que le contenu d'un site Internet peut être suffisant pour mener une étude sur l'innovation et la commercialisation?

Les sites Internet sont des sources d'information disponibles gratuitement et accessibles par tout le monde. La majorité des entreprises ayant un site Internet, la couverture d'une étude basée sur Internet est un avantage.

Pour notre travail de recherche sur la commercialisation de la nanotechnologie au Canada, nous récupérons le plus grand nombre de données sur des sites Internet d'entreprises canadiennes répertoriées par diverses sources comme ayant des activités en nanotechnologie. À partir de ces données, nous classons les entreprises selon leur domaine nanotechnologique grâce à une liste de mots clés provenant d'articles scientifiques pour estimer le niveau d'avancement des entreprises canadiennes dans la recherche nanotechnologique.

Enfin grâce à plusieurs indicateurs d'innovation et de commercialisation nous décrivons l'état actuel de la nanotechnologie dans l'industrie canadienne.

ABSTRACT

Companies rely more every day on the Internet, sharing a lot of information to their customers but also their investors and their associates. Most companies share information such as products, services, partnerships, and research and development activities on their website. All this information is much analytical opportunities provided that we can retrieve and interpret.

In this context, is the content of a website can be enough to conduct a study on innovation and commercialization?

Websites are sources of information available and accessible to everyone. The majority of companies with a website, the coverage of an Internet-based study is an advantage.

For our research on the commercialization of nanotechnology in Canada, we recover as many data on websites of Canadian companies listed by various sources as having activities in nanotechnology. From these data we cluster companies according to their nanotechnology field using a list of keywords from scientific papers to estimate the level of advancement of Canadian companies in nanotechnology research.

Finally thanks to several indicators of innovation and marketing we describe the current state of nanotechnology in the Canadian industry.

TABLE DES MATIÈRES

REMERCIEMENTS III

RÉSUMÉ IV

ABSTRACT V

LISTE DES TABLEAUX IX

LISTE DES FIGURES X

LISTE DES SIGLES ET ABRÉVIATIONS XII

LISTE DES ANNEXES XIII

INTRODUCTION 1

CHAPITRE 1 RECENSION DE LA LITTÉRATURE 3

1.1 Innovation 3

1.1.1 Production de connaissance 4

1.1.2 Processus de transfert de technologie 8

1.1.3 Mesure de l'innovation 10

1.2 Commercialisation 14

1.3 Synthèse 17

CHAPITRE 2 LES NANOTECHNOLOGIES 18

2.1 Contexte 18

2.2 Définition 20

2.3 Classification 21

2.4 État de l'art 25

2.5 Commercialisation des nanotechnologies 27

2.5.1 Phases de commercialisation 27

2.5.2 Facteurs contribuant à la commercialisation ou à son frein 28

2.6 Synthèse 33

CHAPITRE 3 SUJET ET MÉTHODOLOGIE DE RECHERCHE 34

3.1 Problématique 34

3.2 Hypothèse de recherche 35

3.3 Objectifs de recherche 37

3.4 Données 38

3.4.1 Entreprises 38

3.4.2 Source de financement 40

3.5 Méthodologie de recherche 40

3.5.1 L'exploration de données sur Internet 41

3.5.2 Mise en place d'un protocole d'analyse 44

3.5.3 Récupération des données 45

3.5.4 Traitement des données 49

3.6 Forage de données 51

3.6.1 Classification automatique par apprentissage 52

3.7 Regroupement géographique 60

3.8 Indicateur d'innovation et de commercialisation dans les sites Internet 60

3.8.1 Création des indicateurs 60

3.8.2 Méthodologie de récupération des mots clés 64

3.8.3 Construction des indicateurs 66

3.8.4 Statistique descriptive des variables par indicateurs 66

3.8.5 Statistique descriptive des indicateurs 70

3.8.6 Stockage des indicateurs 72

3.8.7 Construction des regroupements suivant les indicateurs 72

3.9 Synthèse 74

CHAPITRE 4 RÉSULTATS ET ANALYSE 76

4.1 Commercialisation des produits en nanotechnologie par domaine 76

4.2 Regroupement géographique 80

4.3 Innovation et commercialisation 81

4.3.1 Corrélation entre les indicateurs 82

4.3.2 Regroupement des entreprises par indicateur 83

CHAPITRE 5 DISCUSSION ET CONCLUSION 90

BIBLIOGRAPHIE 94

ANNEXES 96

LISTE DES TABLEAUX

Tableau 2.1 : Définitions générales de la nanotechnologie (OECD 2010) 20

Tableau 3.1 : Sources des entreprises en nanotechnologie 39

Tableau 3.2 : Données récoltées.....51

Tableau 3.3 : Mots clés selon le domaine de la nanotechnologie 57

Tableau 3.4 : Matrice de confusion.....59

Tableau 3.5 : Description statistique des variables pour l'indicateur de financement.....67

Tableau 3.6 : Description statistique des variables pour l'indicateur de R-D 67

Tableau 3.7 : Description statistique des variables pour l'indicateur de protection intellectuelle .68

Tableau 3.8 : Description statistique des variables pour l'indicateur de collaboration 69

Tableau 3.9 : Description statistique des variables pour l'indicateur de produit.....69

Tableau 3.10 : Description statistique des variables pour l'indicateur de page.....70

Tableau 3.11 : Description statistique des indicateurs 72

Tableau 3.12 : Critère BIC en fonction du nombre de grappe 73

Tableau 3.13 : Répartition des individus dans les grappes suivant la méthode employée.....74

Tableau 4.1 : Produits en nanotechnologie 79

Tableau 4.2 : Description statistique des regroupements en nanotechnologie.....80

Tableau 4.3 : Matrice de corrélation des indicateurs 82

Tableau 4.4 : Regroupement en deux grappes 83

Tableau 4.5 : Regroupement en trois grappes 85

LISTE DES FIGURES

Figure 1.1 : Modèle linéaire 5

Figure 1.2 : Le modèle « Triple hélice » du milieu universitaire - industrie – gouvernement adapté de Etzkowitz and Leydesdorff (2000) 6

Figure 1.3 : Modèle de transfert universitaire des technologies (Bradley et al. 2013) 9

Figure 1.4 : Modèle intrant/extrant (Godin 2004) 10

Figure 1.5 : Flux de financement et acteurs clés de la commercialisation des nanotechnologies (Crawley 2007) 15

Figure 2.1 : Application des nanoparticules par secteur d'activité (Tsuzuki 2009) 19

Figure 2.2 : Diagramme Venn des sous-domaines de la nanotechnologie. *Source* : « Field scope of nanotechnology » développé par Georgia Tech Technology and Assessment Center (GT CNS-ASU Group) 22

Figure 2.3 : Publication scientifique par domaine de la nanotechnologie (Islam and Miyazaki 2010). 24

Figure 2.4 : Échéancier de début de prototypage industriel et commercialisation de la nanotechnologie: quatre générations imbriquées de produits et de procédés (Renn et Roco 2006) 28

Figure 3.1 : Méthodologie générale de recherche 41

Figure 3.2 : Processus d'acquisition et de traitement des données 45

Figure 3.3 : Processus de récolte de donnée du logiciel Nutch 47

Figure 3.4 : Processus de classification automatique 55

Figure 3.5 : Regroupement des sites Internet par domaine 58

Figure 4.1 : Classification des entreprises 77

Figure 4.2 : Régression linéaire des regroupements géographiques 81

Figure 4.3 : Regroupement des entreprises en deux grappes 84

Figure 4.4 : Regroupement des entreprises en trois grappes 86

Figure 4.5 : Regroupement des entreprises avec distinction de la recherche en deux grappes.....87

Figure 4.6 : Regroupement des entreprises avec distinction de la recherche en trois grappes 88

LISTE DES SIGLES ET ABRÉVIATIONS

AFM	Atomic Force Microscope
CAD	Dollar canadien
CIRST	Center for Interuniversity Research on Science and Technology
CNRC	Conseil National de Recherche Canada
DBI	Davies-Bouldin Index
EPO	European Patent Office
K-NN	K-Nearest Neighbors
NEMS	Nanoelectromechanical Systems
NINT	National Institute for Nanotechnology
NM	Nanomètre
NNI	National Nanotechnology Institute
R-D	Recherche et Développement
SME	Small and Medium Enterprises
STM	Scanning Force Microscope
SVM	Support Vector Machine
TF-IDF	Term Frequency-Inverse Document Frequency
URL	Uniform Resource Locator
USD	Dollar américain
USPTO	United States Patent and Trademark Office
XML	Extensible Markup Language

LISTE DES ANNEXES

ANNEXE A – INSTALLATION ET CONFIGURATION DE NUTCH AVEC MYSQL.....102

ANNEXE B – SCRIPT DE DÉTECTION DE LANGAGE.....106

ANNEXE C - SÉLECTION DE PAGES PERTINENTES 109

ANNEXE D –STRATÉGIE DE RECHERCHE DE MOTS CLÉS DE LA NANOTECHNOLOGIE 114

ANNEXE E – STRATÉGIE DE RECHERCHE PAR INDICATEUR.....115

ANNEXE F - SITE D'ENTREPRISES CANADIENNES CLASSÉES PAR DOMAINE.....117

INTRODUCTION

Depuis le discours annonciateur de Richard Feynman « *There's plenty of room at the bottom* » (Feynman 1960) et le développement des premiers microscopes à effet tunnel au début des années 1980 le domaine de la nanotechnologie n'a pas cessé d'évoluer. Cette nouvelle technologie offre d'énormes promesses économiques et sociétales. Le potentiel de cette technologie permet à la fois des améliorations incrémentales sur des produits existants dans de nombreux domaines, mais aussi la transformation ou la création d'industries entières (McNeil et al. 2007). La diffusion de cette technologie peut jouer un rôle crucial sur l'économie future et agir comme moteur de croissance (Palmberg 2007). Cependant, le développement de nouveaux produits, nanotechnologies ou autre, reste fortement dépendant du passage de l'invention à l'innovation. Un des facteurs clés du succès de la commercialisation est sans doute la collaboration entre firmes et universités de même que le financement public de la recherche universitaire (Lorenzoni et al., 2010). Si la génération de connaissances ainsi que la recherche fondamentale sont généralement bien réalisées par les universités, c'est bien la recherche appliquée qui est cruciale au succès des efforts de commercialisation (Osman *et al.*, 2006; Palmberg 2007). Il semble en effet y avoir un fossé entre la recherche en nanotechnologie générée dans les universités et la R-D conduite dans les entreprises de nanotechnologie. L'arrimage et la collaboration entre les deux parties sont donc essentiels au succès de la commercialisation dans le domaine. En conséquence, augmenter le financement de la recherche universitaire peut contribuer indirectement à des résultats commerciaux à travers la collaboration université-entreprise (Huang *et al.*, 2005; Rasmussen et Anderson, 2009). Svensson (2007) a en revanche montré que les brevets des institutions gouvernementales ont de moins bonnes perspectives de commercialisation (plus faibles taux de commercialisation, plus longue période précédant la commercialisation, profits espérés plus faibles). Par ailleurs, les retombées de la connaissance provenant des universités et « capturées » par les industries peuvent avoir un effet important sur les résultats industriels même sans l'apport d'une collaboration formelle entre l'université et l'industrie.

Dans ce contexte, il n'est pas toujours évident avec les indicateurs d'innovation traditionnels tels que les brevets et le financement de la recherche et du développement de mesurer le niveau de commercialisation et d'innovation du domaine des nanotechnologies. Nous essayons donc d'utiliser des sources de données complémentaires, à savoir les sites Internet des compagnies

travaillant dans le domaine des nanotechnologies, pour analyser grâce à différents indicateurs l'innovation et la commercialisation des produits en nanotechnologie au Canada.

Ce mémoire de recherche est structuré de la manière suivante. Le chapitre 1 recense la littérature portant sur l'innovation, avec sa modélisation et ses indicateurs, et la commercialisation des produits de haute technologie. Dans le chapitre 2, nous nous intéressons plus en particulier à la nanotechnologie, sa définition, ses sous-domaines ainsi que les particularités concernant sa commercialisation. Le chapitre 3 décrit la démarche générale du projet de recherche ainsi que les différentes hypothèses, les données et la méthodologie employée dans notre étude pour répondre aux objectifs. Le chapitre 4 présente les résultats ainsi que les analyses obtenues. Et enfin dans le chapitre 5, nous discutons de manière générale de ces résultats ainsi que de la validation des hypothèses de recherche. Finalement nous concluons le mémoire avec les limites de cette recherche ainsi que les préconisations pour la poursuite éventuelle de ce projet.

CHAPITRE 1 RECENSION DE LA LITTÉRATURE

Le but de ce chapitre est de comprendre les mécanismes d'innovation menant à la commercialisation d'une technologie de pointe, telle que la nanotechnologie, et de cadrer notre étude. Nous discutons premièrement de différents modèles d'innovation afin de mettre en lumière la synergie existant entre les entreprises, les universités et le gouvernement dans le développement des innovations et les moyens possibles pour la mesurer. Enfin nous abordons les mécanismes de commercialisation d'une technologie de pointe pour comprendre les facteurs contribuant à une commercialisation réussie.

1.1 Innovation

Toutes les inventions issues de la recherche et du développement n'obtiennent pas le même succès commercial. Une innovation se distingue d'une invention par une commercialisation réussie (Schumpeter 1934). Dans la troisième édition du Manuel d'Oslo, l'Organisation de Coopération et de Développement Économique (OCDE) définit une innovation comme « la mise en œuvre d'un produit (bien ou service) ou d'un procédé nouveau ou sensiblement amélioré, d'une nouvelle méthode de commercialisation ou d'une nouvelle méthode organisationnelle dans les pratiques de l'entreprise, l'organisation du lieu de travail ou les relations extérieures » (OCDE 2005, p.54)

Cette définition générale prend en compte de nombreux types d'innovations, dont l'innovation organisationnelle et l'innovation de commercialisation. Notre recherche se limitera toutefois à l'étude des innovations technologiques de produits et de procédés (TPP), adaptés au domaine des nanotechnologies.

Nous allons donc commencer par voir quels sont les processus de production de connaissance menant à l'innovation à travers différents modèles et quels sont les mécanismes de transferts de connaissances permettant de faire le passage réussi d'une invention à une innovation. À travers ces modélisations nous serons capables d'identifier les acteurs participant à la commercialisation des nanotechnologies pour cadrer notre recherche.

1.1.1 Production de connaissance

Pour analyser l'innovation dans un domaine de haute technologie, tel que la nanotechnologie, il est important de prendre en compte tous les acteurs présents dans le processus de création de connaissances.

La nanotechnologie étant encore actuellement une technologie embryonnaire, elle réside encore beaucoup au sein de la recherche universitaire (Kim et al. 2008). En effet, la nanotechnologie repose énormément sur la recherche fondamentale, nécessaire à son développement. La recherche fondamentale est couteuse en temps, en argent et en personnel. La nécessité de matériel couteux ainsi que de personnel hautement qualifié fait des universités, ainsi que de tous les autres établissements d'enseignement supérieur, des lieux privilégiés pour la recherche fondamentale. Cette recherche fondamentale est nécessaire au développement des nouvelles technologies dans les domaines de pointe (Anselin et al. 1997). Les établissements d'enseignement supérieur stimulent l'innovation en établissant des liens formels ou informels avec des entreprises de haute technologie (OCDE 1993).

Les liens existant entre le milieu académique et l'innovation sont mis en valeur dans les modèles de production de connaissance ainsi que dans les modèles de processus de transfert technologique.

1.1.1.1 Modèles d'innovation

Le modèle linéaire historique d'innovation, de type « market pull » ou « technology push », proposé par Bush en 1945 (Bush 1945) permet de comprendre rapidement le développement d'une innovation.

Dans le modèle linéaire, Figure 1.1, le processus de développement de produit commence avec la recherche fondamentale qui permet faire de nouvelles découvertes. La recherche fondamentale se poursuit en recherche appliquée et développement avec la conception du produit, la réalisation d'essais pour le valider et la planification du lancement du produit. Enfin la dernière phase, la production et la diffusion de l'innovation permet finalement de lancer le produit sur le marché.

Figure 1.1 : Modèle linéaire

Ce modèle a beaucoup été employé pour des raisons statistiques, car il permet de classer facilement en catégorie les ressources dépensées entre la science et la technologie et a été largement diffusé par l'OCDE pour sa simplicité d'utilisation (Godin 2006). Cependant ce modèle est maintenant désuet, car il ne permet pas, entre autres, de prendre en compte les transferts de technologie ou de connaissance.

Actuellement le processus de création de connaissances connait trois modèles en concurrence pour expliquer les relations existantes d'un point de vue socio-économique entre les acteurs de la création de connaissances: le modèle des systèmes d'innovation nationaux de Dosi et al. (1988), le modèle « Mode 2 » proposé par Gibbons et al. (1994) et le modèle de la « Triple hélice » de Etzkowitz and Leydesdorff (2000).

Dosi et al., proposent dans leur modèle des systèmes d'innovation nationaux que l'innovation soit dirigée en majorité par les entreprises qui ont un rôle de meneur au sein d'un pays. Les connaissances nécessaires à l'innovation ne proviennent pas de la recherche, mais plus probablement du marketing, des utilisateurs et des ingénieurs en production (Lundvall et al. 2002).

Dans leur livre, Gibbons et al. (1994) argumentent que le contexte de création de connaissances est bien plus large que le contexte académique et que les universités n'ont qu'un rôle limité dans la recherche et dans la création de connaissances. Le « Mode 2 » qu'ils proposent est orienté vers la résolution de problèmes pratiques par des équipes de chercheurs pluridisciplinaires. Il est en opposition au « Mode 1 » selon lequel les chercheurs ne produisent que des connaissances fondamentales sans véritables applications pratiques dans des domaines très cloisonnés. Néanmoins, Etzkowitz and Leydesdorff (2000) avancent de leur côté un modèle prenant en compte les interactions existantes entre les institutions majeures de notre société actuelle à différents stades de développement dans le processus de création de connaissances et de transfert technologique. Ces institutions majeures, nommées « dynamiques », sont les universités, le gouvernement et l'industrie. Ce modèle est une évolution des modèles précédents et prend la forme d'une « Triple

hélice » symbolisant l'influence continue qu'exercent entre elles ces trois dynamiques (Etzkowitz and Leydesdorff 2000).

Le modèle de la Triple hélice peut être étendu à un modèle de N-uple hélice (Leydesdorff 2011) pour intégrer d'autres dynamiques. Rieu (2011) préconise par exemple l'ajout d'une quatrième dynamique représentant l'influence de la société sur l'innovation. Cependant pour la suite de notre recherche nous utilisons le modèle de la « Triple Hélice » développée par Etzkowitz and Leydesdorff, car le rôle de la société est difficilement quantifiable.

1.1.1.2 « Triple Hélice »

Dans une société de connaissance telle que la nôtre, la progression de la technologie est un enjeu majeur et seule la synergie entre chaque acteur permet le développement de nouvelles innovations. Ainsi pour les innovations basées sur le savoir, un minimum de trois dynamiques est nécessaire: le milieu académique (production de connaissances), le milieu industriel (diffusion), et le gouvernement (contrôle) (Leydesdorff 2011). La Figure 1.2 montre les interactions entre ces trois dynamiques. Selon Etzkowitz et Leydesdorff, la proximité géographique de ces trois dynamiques est aussi une condition nécessaire, mais non suffisante pour le développement réussi d'une grappe industrielle.

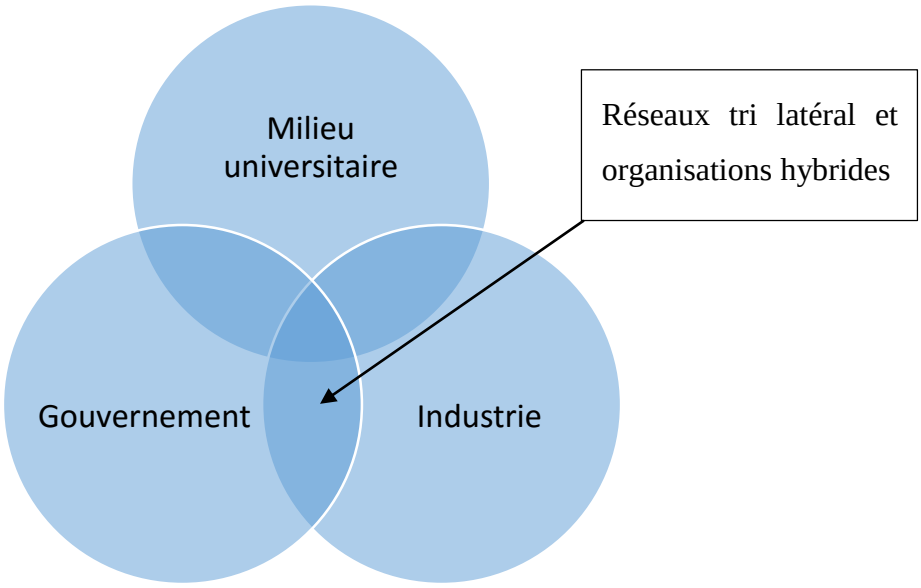

Figure 1.2 : Le modèle « Triple hélice » du milieu universitaire - industrie – gouvernement adapté de Etzkowitz and Leydesdorff (2000)

Le rôle toujours plus important effectué par les universités dans la production de connaissance ainsi que leur rôle d'incubateur d'entreprises de haute technologie leur procurent une place centrale dans ce contexte institutionnel. En plus de synthétiser et recombiner d'anciennes technologies, le milieu universitaire crée de nouvelles technologies et parvient à les utiliser pour résoudre des problèmes concrets. Ce rôle privilégié dans la production d'innovation est mis en valeur par l'augmentation des créations entreprises universitaires dérivées et des collaborations entreprises-universités.

Les entreprises pour leur part, augmentent leur capital intellectuel et leur niveau technologique en interagissant avec les universités et en menant leur propre activité de R-D, en plus de leur rôle de production et de distribution de produits. Les retombées de la recherche universitaire sont indispensables aux entreprises technologiques afin de pouvoir bénéficier de la recherche fondamentale effectuée au sein de ses établissements pour rester compétitive sur un marché ultra concurrentiel et en constante évolution.

Le rôle du gouvernement est principalement de fournir un cadre de travail en modifiant les infrastructures et les réglementations pour s'adapter aux différentes configurations institutionnelles au fil du temps. Pour que la recherche fondamentale puisse être correctement effectuée, les établissements d'enseignement supérieur et les laboratoires gouvernementaux ont besoin de beaucoup de financement de la part du gouvernement. Ce dernier doit donc définir des priorités dans sa politique de financement de la recherche en accord avec les universités et les entreprises. Les politiques d'innovation sont donc issues de l'interaction du gouvernement avec les autres dynamiques.

Ainsi les dynamiques s'ajustent sans cesse les unes par rapport aux autres pour anticiper les opportunités et les résistances (Lamy 2007).

Toutefois, la contribution des trois dynamiques du modèle « Triple hélice » n'est pas forcément égale. Par exemple dans le cas de la recherche en nanotechnologie, la part mondiale universitaire de recherche en nanotechnologie s'élevait à 70,45%, celle des instituts de recherche publique à 22,22% et celle de l'industrie à 7.33% en 2007. Cette forte disparité s'explique par le caractère émergeant de cette technologie et sa faible adoption par l'industrie (Miyazaki and Islam 2007).

Si les lieux de production de connaissance ont tendance à se diversifier, le rôle de l'université est encore incontournable et reconnu pour la création d'innovations. Contrairement à ce que laissent croire Gibbons et al. (1994) en affirmant que l'augmentation des lieux de production de

connaissance fait diminuer fortement la part de recherche des universités, Godin et Gingras (2000) signalent une forte augmentation des collaborations entre les industries et le secteur public au Canada. Cette augmentation des collaborations souligne le renforcement des interactions entre les différentes sources de connaissance et renforce plus encore la position centrale de l'université dans le processus de recherche et développement. La contribution des universités dans le modèle d'innovation est donc primordiale.

En partant de ce modèle, nous mettons en évidence le fait qu'une entreprise ayant des liens avec une université et/ou un organisme gouvernemental est plus à même de bénéficier de la recherche fondamentale effectuée par ces derniers et donc de développer une innovation.

1.1.2 Processus de transfert de technologie

Ayant identifié les acteurs de la production de connaissance, il est maintenant intéressant de voir comment les entreprises parviennent à bénéficier de la recherche fondamentale et appliquée effectuée par les universités et comment les innovations sont diffusées au grand public.

Plusieurs canaux permettent aux entreprises de capturer les connaissances produites par les universités. Ainsi les flux de capitaux humain entre les établissements d'enseignement supérieur et les entreprises, les publications scientifiques, les brevets participent à diffuser les connaissances favorisent l'innovation technologique.

Afin de mettre en valeur les retombées de connaissances acquises par les entreprises et le rôle de la recherche académique, des modélisations de transfert de connaissance ont été effectuées. Le processus de transfert de technologie est le processus qui permet de transférer les connaissances acquises par la recherche, publique ou privée, dans le domaine de l'industrie pour des fins de commercialisation.

La modélisation de ce processus continue d'évoluer et de se complexifier pour tenir compte de tous les facteurs contribuant aux transferts technologiques universitaires. Au début des années 2000, le modèle généralement acquis est un modèle linéaire. Le modèle commence par la découverte effectuée par un chercheur universitaire pour aboutir à l'utilisation d'une nouvelle technologie par une entreprise existante ou à la création d'une jeune pousse (« start-up ») ou d'une entreprise essaimée (« spin off ») (Siegel et al. 2004). Le modèle proposé par Bradley et al. (2013) Figure

1.3, quitte le schéma linéaire pour tenir compte de plusieurs facteurs supplémentaires existant en parallèle contrairement au modèle linéaire classique.

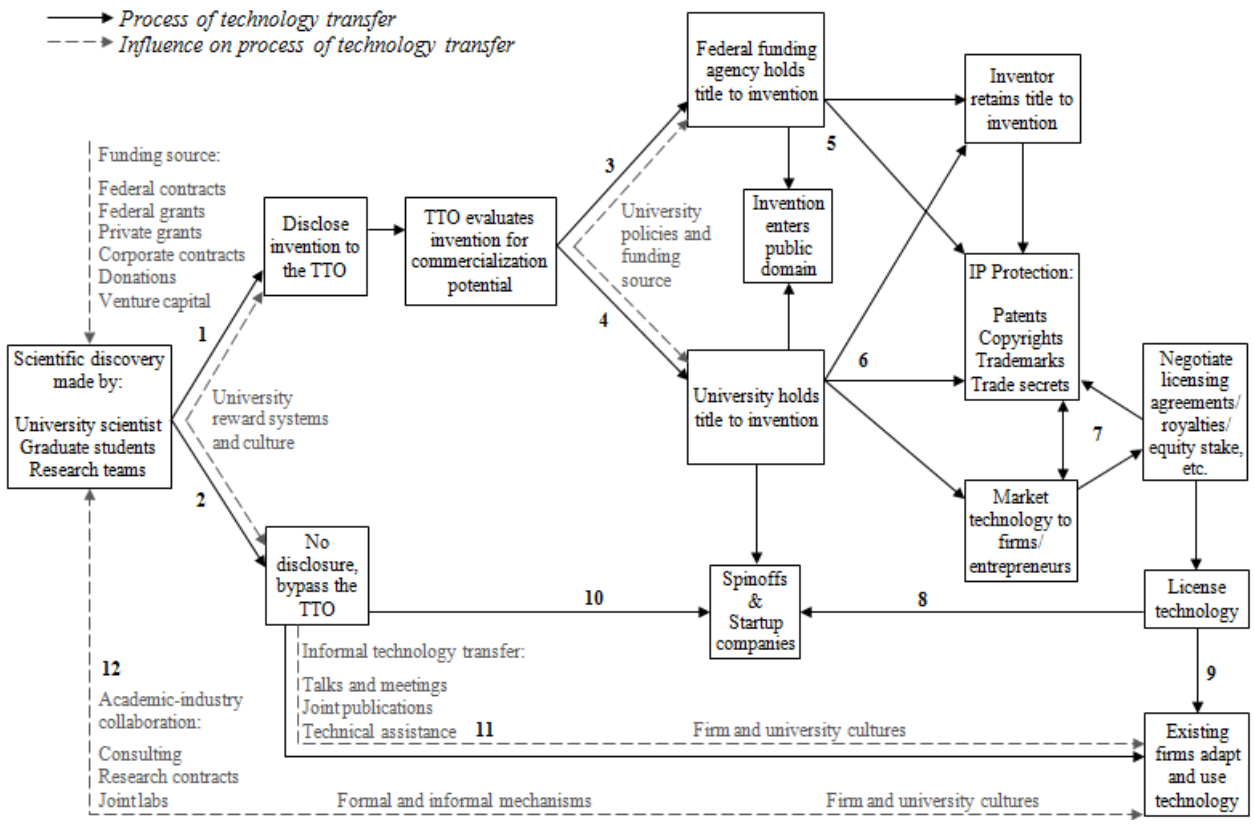

Figure 1.3 : Modèle de transfert universitaire des technologies (Bradley et al. 2013)

Ce modèle nous donne beaucoup d'informations sur les circuits empruntés par les technologies, leur moyen de financement, les acteurs prenant part aux transferts, la protection intellectuelle, et l'utilisation finale de la technologie par une entreprise déjà existante ou par la création d'une nouvelle entreprise. Nous verrons que beaucoup de ces informations sont utilisées pour évaluer le niveau d'innovation.

Nous avons maintenant identifié les différentes parties prenantes grâce aux modèles d'innovation et de transferts de technologie ainsi que leurs rôles effectués dans le processus de création d'innovations. Il est intéressant de voir comment quantifier le travail effectué par ces acteurs et quels sont les indicateurs employés pour mesurer l'innovation.

1.1.3 Mesure de l'innovation

Les indicateurs d'innovation sont nécessaires pour prendre des décisions stratégiques dans la politique d'innovation d'une entreprise ou d'un pays. C'est pour comprendre plus en détail les répercussions, principalement économiques, que peut avoir une politique d'innovation, que de nombreux indicateurs ont été élaborés. Toutes ces informations sont employées par les analystes et les décideurs politiques.

Historiquement, les États-Unis ont commencé les premiers à mesurer le développement des sciences et des technologies pour gérer les laboratoires industriels ainsi que pour planifier les activités de recherche gouvernementale dès les années 1930. La mesure des sciences et technologies se base alors sur un modèle simple d'intrant et d'extrant (Figure 1.4) de l'activité de recherche effectuée par les entreprises (Godin 2004).

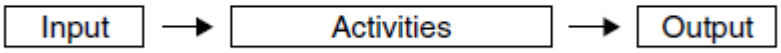

Figure 1.4 : Modèle intrant/extrant (Godin 2004)

Afin de comprendre plus en détail le processus d'innovation, les données doivent être récoltées de façon systématique et uniforme pour construire les indicateurs. Cette récolte systématique des données doit obéir à certaines règles afin de pouvoir effectuer des comparaisons à l'échelle internationale. C'est dans cet objectif précis que l'ODCE édite le Manuel d'Oslo pour étudier les innovations TPP dans le secteur des entreprises en fournissant une méthodologie pour constituer les indicateurs d'innovation. (OCDE and Eurostat 2005)

Les intrants de la recherche sont généralement plus employés pour les études d'innovation, car la plupart des entreprises, en particulier celle possédant de grands pôles de recherche, tiennent à jour leurs dépenses effectuées en R-D. Ces états financiers facilitent la récupération des données à des fins d'analyse pour les études. Généralement ce type d'étude se base sur des questionnaires envoyés aux entreprises. Beaucoup d'études sont donc encore basées sur des indicateurs concernant les intrants de la recherche particulièrement pour des raisons de facilité.

Cependant le processus d'innovation est bien plus vaste que la simple activité de R-D et l'existence d'écarts entre les mesures des intrants et l'impact réel sur l'économie ont incité le développement de nouveaux indicateurs basés sur les extrants de la recherche(Godin 2004).

1.1.3.1 Recherche et développement

D'après le Manuel de Frascati édité par l'OCDE *« la recherche et le développement expérimental (R-D) englobent les travaux de création entrepris de façon systématique en vue d'accroître la somme des connaissances, y compris la connaissance de l'homme, de la culture et de la société, ainsi que l'utilisation de cette somme de connaissances pour de nouvelles applications»* (OCDE 1994, p. 34).

La mesure de l'activité et des ressources de R-D est un indicateur d'innovation, car il est directement lié au progrès technologique. La R-D est un indicateur populaire, car depuis les années 1950 les données sur les ressources consacrées à la R-D au sein des entreprises sont collectées de façon systématique. Généralement, la mesure des ressources consacrées à la R-D se fait grâce à des études nationales menées selon le Manuel de Frascati afin de comparer les efforts en R-D entre pays. Pour la comparaison entre entreprises, il est d'usage de comparer le pourcentage d'investissement en R-D par rapport aux ventes ou le pourcentage de personnes travaillants en R-D par rapport au nombre d'employés total de l'entreprise (Kleinknecht et al. 2002).

La R-D reste cependant un indicateur limité et possède aussi de nombreux inconvénients. Premièrement, la R-D est un intrant de l'innovation. Ainsi, sa mesure ne permet pas de mesurer l'innovation, au sens de mise sur le marché d'un nouveau processus, produit ou service, qui est un extrant de la recherche. La R-D n'est pas le seul intrant de l'innovation et ne concerne donc pas toutes les sources d'innovation possibles. Par exemple l'apprentissage par la pratique n'est pas pris en compte (OCDE and Eurostat 2005). De plus, les mesures officielles de la R-D tendent à sous-estimer les efforts de R-D effectuée par les petites et moyennes entreprises (PMEs) qui peinent à répondre aux questionnaires trop élaborés.

La R-D est donc très fortement liée à l'innovation, mais ne permet pas de la mesurer directement. Connaitre des activités de R-D d'une entreprise et leur financement, public ou privé, montre les

efforts effectués par l'entreprise dans le but de commercialiser de nouveau produit sans pour autant pouvoir le quantifier.

1.1.3.2 Brevets

Les brevets font partie des propriétés intellectuelles accordées aux entreprises. La propriété intellectuelle correspond à l'ensemble des droits accordés sur les créations intellectuelles. On distingue deux grandes parties : la propriété industrielle et le droit d'auteur. Le droit d'auteur s'appliquant majoritairement pour les créations artistiques, il ne nous intéresse pas dans le cadre de notre étude. La propriété industrielle a pour but de favoriser la création d'innovations en offrant une protection juridique pour encourager les investissements et permettre aux chercheurs d'obtenir le monopole de la commercialisation de leur invention (Besen and Raskind 1991).

La propriété intellectuelle, et plus particulièrement les brevets, est un output de la R-D et donc un indicateur intermédiaire de l'innovation. Les brevets sont souvent employés comme indicateur d'innovation dans les études scientométriques (Pavitt 1985). L'utilisation des bases de données de brevet i.e. des offices européens, américains, japonais, répertoriant les brevets par catégories rend ce type d'analyse abordable.

Cependant, le brevetage semble négligé par les PME qui emploient d'autres types de PI moins onéreux comme le secret industriel. Les entreprises invoquent plusieurs raisons pour expliquer cette démarche:

- Le coût du brevetage
- Le faible (voire négatif) retour sur investissement.
- Les coûts élevés de la surveillance d'utilisations des droits liées aux brevets.
- La complexité et la cherté des actions à mener en cas d'infraction. (Libaers, et al. 2006)

Il est donc évident que l'usage du brevet comme indicateur d'innovation, même s'il est très fréquent, reste limité et ne permet pas de prendre en compte les inventions et innovations non brevetées et non brevetables. Aussi les stratégies de brevetage dépendent, comme nous l'avons déjà vu, du type de technologie.

1.1.3.3 Autre protection de la propriété intellectuelle

Un autre type de protection de la propriété intellectuelle est employé par les entreprises. Les marques déposées sont employées par les entreprises comme protection de la propriété intellectuelle de leur nom de produit. Selon l'Organisation Mondiale de la Propriété Intellectuelle (OMPI), une marque est « *un signe permettant de distinguer les produits ou services d'une entreprise de ceux d'autres entreprises. Les marques sont des droits de propriété intellectuelle protégés* »[1]. Une marque déposée, contrairement aux brevets, peut être utilisée comme protection de la propriété intellectuelle aussi longtemps que le propriétaire paye son enregistrement. Le fait que le propriétaire paye pour employer la marque est une preuve de la valeur commerciale du produit. Une marque non employée étant considérée comme une barrière au développement économique, elle peut être retirée à une entreprise. Un produit ou service protégé par une marque déposée est reconnaissable aux signes suivant : « ™ » ou « ® » (Mendonça et al. 2004)

Une marque déposée est donc un moyen de connaitre la valeur commerciale d'un produit pour une entreprise.

1.1.3.4 Collaboration

La nanotechnologie étant une technologie qui chevauche les autres disciplines sans pour autant les remplacer (Roco 2001), la collaboration au stade de R-D est donc primordiale pour une commercialisation réussie des produits en nanotechnologies.

Comme nous l'avons déjà vu dans le modèle de gestion de l'innovation, la collaboration pour l'innovation est un facteur clé dans le processus d'innovation et en particulier pour une entreprise ancrée dans un domaine scientifique. La collaboration pour l'innovation est définie par Tether (2002) comme l'action de participer de manière conjointe avec une autre entreprise dans un projet d'innovation. La collaboration peut avoir lieu avec différentes entités. Ainsi au Canada, 70% des entreprises en nanotechnologie annoncent avoir une activité de collaboration dans les phases de développement ou de commercialisation de leur produit et parmi les 88 entreprises identifiée au

---

[1] http://www.wipo.int/trademarks/fr/ consulté en mai 2015

cours de l'étude menée courant 2005, 55 d'entre elles ont réalisée 107 ententes de collaboration (McNiven 2007) avec des universités, des instituts gouvernementaux et d'autre entreprises.

De manière générale, la collaboration pour l'innovation est encouragée à travers l'innovation ouverte, un nouveau modèle d'innovation proposé par Chesbrough (2003). Dans ce modèle d'innovation, les entreprises partagent leur effort de recherche et leurs propriétés intellectuelles.

De plus selon Kim (2008), la collaboration entre les universités et les entreprises facilite la commercialisation des nanotechnologies. La recherche scientifique et l'innovation technologique sont fortement liées et les retombées des activités de R-D effectuées par les universités dans l'industrie sont multiples, que la collaboration soit directement établie ou non. De plus, la mobilité des chercheurs échanges d'information et la création de nouvelles « start-ups » (Audretsch and Aldridge 2009) sont autant de retombées de connaissance possibles prouvant le rôle clé joué par l'université dans le processus d'innovation. Aussi, le financement public de la recherche universitaire issue de la collaboration entre le gouvernement et les universités et le financement privé provenant des entreprises auraient un impact positif sur la production scientifique (Blume-Kohout *et al.* 2009) et le travail de recherche des chercheurs subventionnés (Campbell et al. 2010).

1.2 Commercialisation

Le passage de l'invention technologique à l'innovation technologique est caractérisé par une commercialisation réussie de la technologie. Ainsi, avant de pouvoir bénéficier des retombées de la recherche effectuée grâce au financement public, il est nécessaire que les recherches fondamentales et appliquées réalisées dans les laboratoires soient transférées dans les usines de fabrication et dans les magasins.

Les développements effectués par la recherche universitaire aussi bien que par les laboratoires de recherche privés et gouvernementaux doivent être transférés, la plupart du temps sous forme de brevets aux entreprises afin qu'ils les produisent et qu'ils les commercialisent, comme nous avons pu le voir avec le processus de transfert universitaire.

Deux types différents d'entreprises coexistent pour la commercialisation des technologies de pointe. Soit une nouvelle compagnie est fondée pour la commercialisation spécifique d'une nouvelle technologie, soit la technologie est acquise par une compagnie déjà existante.

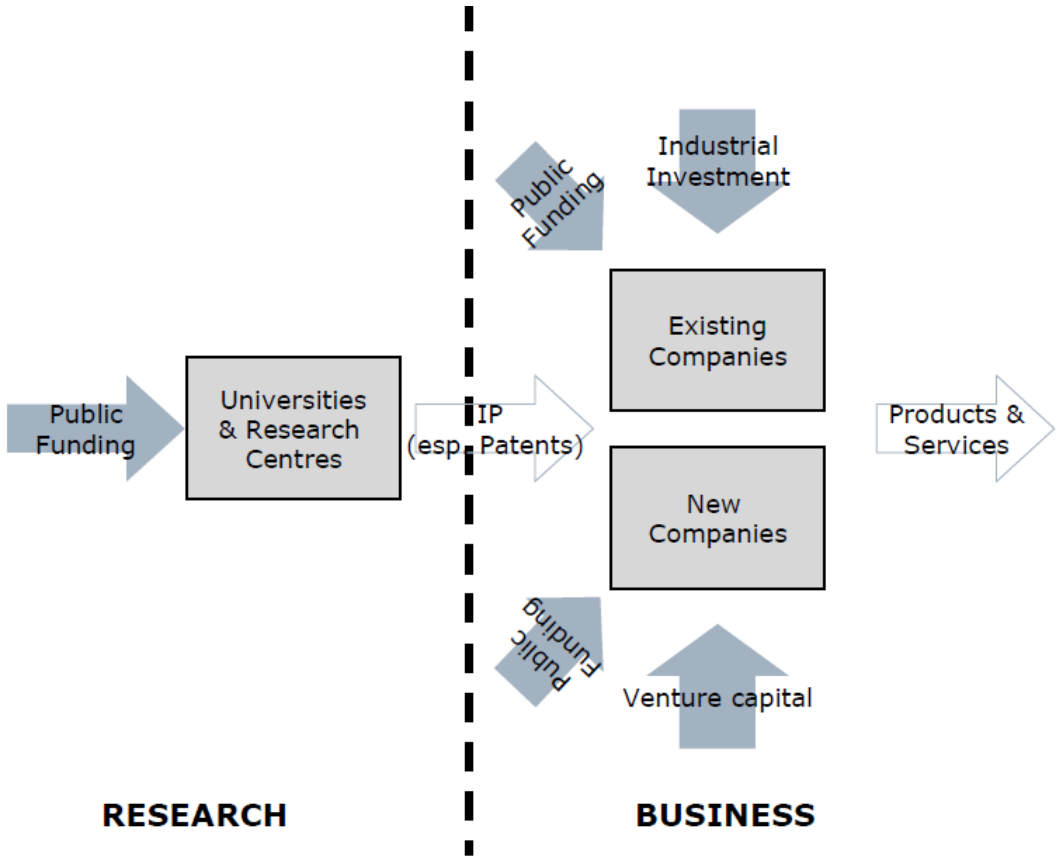

Figure 1.5 : Flux de financement et acteurs clés de la commercialisation des nanotechnologies (Crawley 2007)

Comme le montre la Figure 1.5, le financement public est présent dans tout le processus, passant par le financement de la recherche jusqu'aux aides de financement des nouvelles entreprises et des programmes de développement des entreprises existantes.

Dans le cas d'une nouvelle compagnie, avant de pouvoir générer des revenus de manière indépendante, elle doit obtenir un financement externe sous forme de capital-risque pour débuter. Si le lancement de la compagnie se passe correctement en générant suffisamment de profit pour survivre avec la commercialisation de ces produits et services, la compagnie peut continuer d'évoluer de manière indépendante ou se faire racheter par une plus grosse compagnie. Dans le cas contraire, c'est un échec et la compagnie est amenée à disparaître.

Dans le cas d'une compagnie déjà existante, celle-ci peut acquérir de nouvelles technologies pour développer son activité et gagner en compétitivité ou viser à exploiter de nouvelles opportunités de

marché (Crawley 2007). Des programmes de financement public ainsi qu'un investissement industriel aident les entreprises à continuer à se développer.

Dans le cas de l'adoption de la technologie par une entreprise déjà existante on peut s'intéresser à la taille de l'entreprise en question. En effet, le modèle de transfert technologique employé pour les études en nanotechnologie est souvent basé sur celui d'une autre technologie émergente, la biotechnologie. Dans ce modèle, les PMEs jouent un rôle important dans le transfert de la technologie sur le marché (Darby et Zucker 2003, Rothaermel et Thursby 2007).

Cependant Genet et al. (2012) avancent de leur côté que la nanotechnologie est plus proche de la microélectronique à ses débuts avec le rôle fondamental de grandes compagnies i.e. Fairchild Semiconductors, IBM, Texas Instrument. Seules ces grandes compagnies possèdent les connaissances et les fonds nécessaires aux phases de R-D et de commercialisation laissant un rôle moindre aux PMEs. Libaers et al (2006) estiment aussi que le rôle joué par les entreprises universitaires dérivées (USOs) dans la commercialisation des technologies émergentes reste très limité. En effet, dans le cas de la nanotechnologie, la fabrication de produits à l'échelle nanométrique en grande série demande un investissement impossible à fournir pour une petite entreprise. La taille d'une entreprise est un élément déterminant dans le cas de la commercialisation des technologies de pointe.

Si pour Gans, Hsu, et Stern (2000) la sortie de nouveaux produits technologiques est un facteur de plus en plus important pour garder un avantage compétitif sur le marché, la phase de commercialisation, bien que critique dans le processus d'innovation est souvent considéré comme étant la moins bien gérée pour les innovations technologiques. Selon Cierpicki et al. (2000), le taux d'échec de lancement d'un nouveau produit au cours des vingt dernières années est compris entre 40% à 50%.

Pour le marché de la haute technologie, l'évolution rapide de la technologie, la prolifération de nouveaux produits et le caractère volatile de la technologie (Bayus 1994) réduisent considérablement les opportunités de lancement de produit. Les décisions stratégiques de commercialisation ont donc plus d'impact dans le domaine des hautes technologies que sur le marché traditionnel (Nevens 1990).

1.3 Synthèse

Dans ce chapitre nous avons fait le bilan des définitions possibles pour notre domaine d'étude ainsi que les enjeux économiques et technologiques liés à son développement. Les modèles d'innovations nous aident à mieux comprendre les acteurs présents dans le processus de création des innovations, en passant par la recherche initiale jusqu'à la commercialisation des produits. Nous allons maintenant voir quelle est notre problématique, nos hypothèses ainsi que les objectifs spécifiques pour y répondre.

CHAPITRE 2 LES NANOTECHNOLOGIES

Le présent chapitre a pour but d'offrir une meilleure compréhension de notre cadre d'étude, à savoir la nanotechnologie. Nous effectuons premièrement une mise au point sur l'origine et la définition de la nanotechnologie. Nous examinons ensuite une classification possible de la nanotechnologie en sous-domaine avec les tendances de recherche actuelle de ces sous-domaines. Un état de l'art est dressé pour avoir une vision d'ensemble du développement de la nanotechnologie et des perspectives économiques envisageables. Enfin nous voyons les particularités de la nanotechnologie en ce qui concerne l'innovation et la commercialisation.

2.1 Contexte

La nanotechnologie est l'application des recherches effectuées par les nanosciences. Les nanosciences concernent l'étude, la manipulation ainsi que le développement de procédés de fabrication de structures et de dispositifs dont au moins une des dimensions est à l'échelle nanométrique. Ces dimensions nanométriques donnent à la matière de nouvelles propriétés offrant de très nombreuses applications aux nanotechnologies.

Il est généralement admis que le concept de nanotechnologie provient du discours prononcé par le prix Nobel Richard Feynman en 1959 (Feynman 1960). Le terme de nanotechnologie est introduit quelques années plus tard, en 1974, par le scientifique japonais Norio Taniguchi (Taniguchi et al 1974).

Le développement du microscope à effet tunnel en 1981, suivi par le microscope à force atomique et le microscope optique en champ proche, a permis l'essor des nanosciences en offrant la possibilité aux laboratoires de recherche d'observer et de manipuler directement les molécules et les atomes pour un coût relativement restreint.

La nanotechnologie est souvent proposée comme étant la future technologie à usage général (TUG) ayant la capacité de modifier profondément l'économie actuelle ainsi que notre structure sociétale en agissant comme moteur de croissance ((Palmberg 2007, Mangematin et al. 2011). Selon Lipsey et al. (2005), si la nanotechnologie réalise même une fraction de son potentiel, elle deviendra la TUG la plus importante du XXIe siècle.

Ainsi les nanotechnologies ont la capacité de résoudre des défis industriels actuels, mais aussi de servir de point de convergence à d'autres technologies telles que la biotechnologie, l'électronique, la physique (Linton and Walsh 2004).

La nanotechnologie a la capacité de modifier un grand nombre de secteurs tels que l'électronique, les cosmétiques, les médicaments, l'énergie, les systèmes catalytiques, l'alimentation, l'emballage, les ingrédients dans la formulation de produits domestiques et industriels ou encore les équipements de sport (Maynard 2006). La Figure 2.1 nous donne un aperçu des applications possible des nanoparticules.

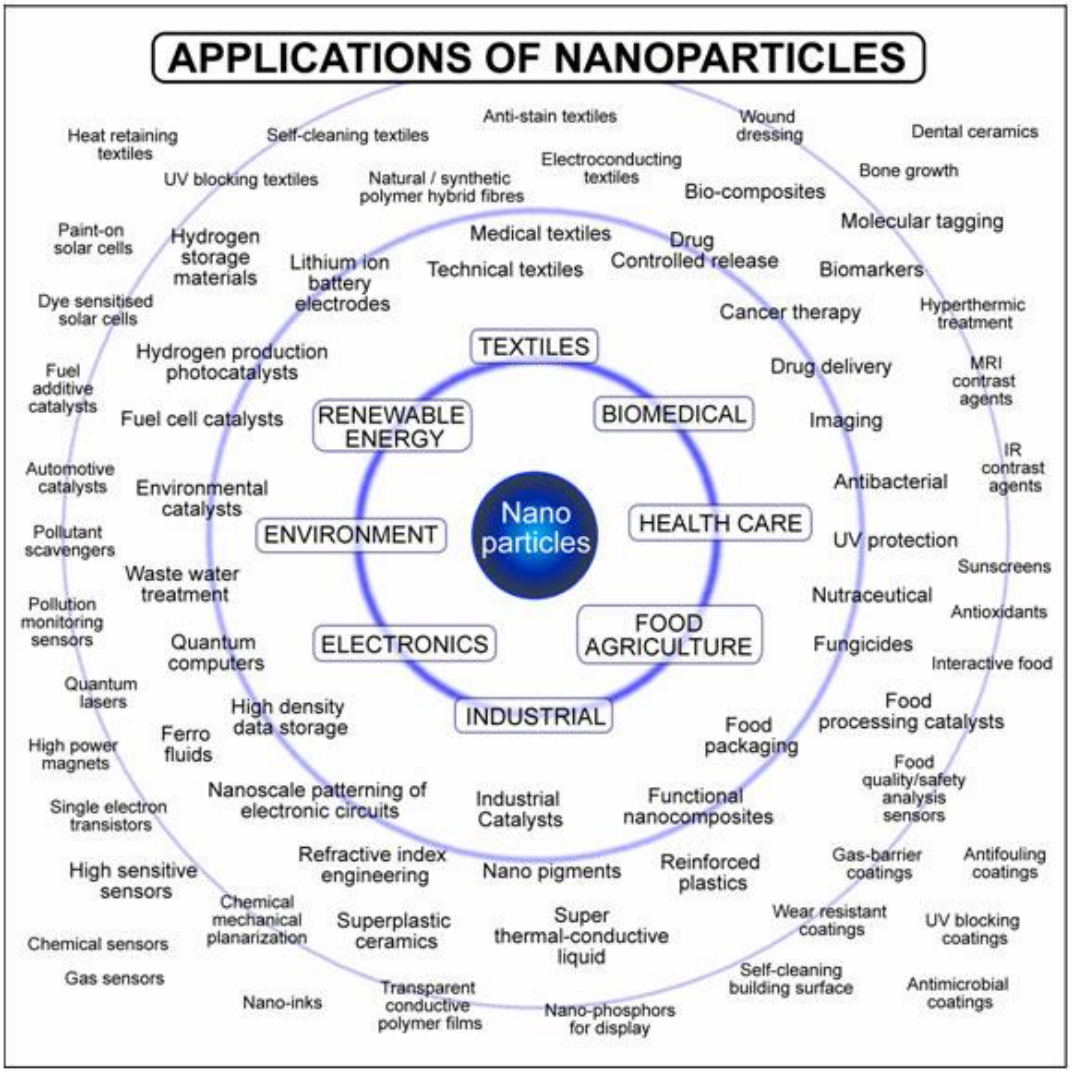

Figure 2.1 : Application des nanoparticules par secteur d'activité (Tsuzuki 2009)

2.2 Définition

La nanotechnologie est une technologie très prometteuse et de nombreux gouvernements, tel que le Canada, ont compris la nécessité d'investir rapidement dans la recherche et le développement de cette technologie à travers des programmes de recherche nationaux.

En raison des multiples programmes de recherche nationaux, beaucoup de définitions coexistent dans la littérature, dont cette liste non exhaustive tirée de sources internationales (voir le Tableau 2.1).

Tableau 2.1 : Définitions générales de la nanotechnologie (OECD 2010)

Source	Définition
US: National Nanotechnology Initiative (2001-	Nanotechnology is the understanding and control of matter at dimensions of roughly 1 to 100 nanometres, where unique phenomena enable novel applications. Encompassing nanoscale science, engineering and technology, nanotechnology involves imaging, measuring, modelling, and manipulating matter at this length scale.
EU: Seventh Framework Programme (2007-2013)	Generating new knowledge on interface and size-dependent phenomena; nano-scale control of material properties for new applications; integration of technologies at the nano-scale; selfassembling properties; nano-motors; machines and systems; methods and tools for characterisation and manipulation at nano dimensions; nano precision technologies in chemistry for the manufacture of basic materials and components; impact on human safety, health and the environment; metrology, monitoring and sensing, nomenclature and standards; exploration of new concepts and approaches for sectoral applications, including the integration and convergence of emerging technologies
Working definition of ISO in 2007	Understanding and control of matter and processes at the nanoscale, typically, but not exclusively, below 100 nanometres in one or more dimensions where the onset of size dependent phenomena usually enables novel applications. Utilising the properties of nanoscale materials that differ from the properties of individual atoms, molecules, and bulk matter, to create improved materials, devices, and systems that exploit these new properties
European Patent Office (EPO) in 2009	The term nanotechnology covers entities with a geometrical size of at least one functional component below 100 nanometres in one or more dimensions susceptible of making physical, chemical or biological effects available which are intrinsic to that size. It covers equipment and methods for controlled analysis, manipulation, processing, fabrication or measurement with a precision below 100 nanometres.

Toutes les définitions existantes de la nanotechnologie ont été créées pour être employées dans un cadre précis. Elles sont toutes construites autour des mêmes points et selon Palmberg et al. (2009), il faut retenir trois aspects communs de ces définitions :

- La volonté affirmée de contrôler et manipuler la matière à l'échelle nanométrique;
- L'importance des phénomènes dépendant de la taille de la matière;
- Le caractère innovant des possibilités offertes par la nanotechnologie dans plusieurs domaines.

En 2010, l'Organisation internationale de normalisation (ISO) a mis au point avec 33 pays, dont le Canada, une définition internationale de la nanotechnologie dans la norme ISO/TS 80004-1:2010 que nous emploierons pour la suite de notre projet :

« application of scientific knowledge to manipulate and control matter in the nanoscale in order to make use of size- and structure-dependent properties and phenomena, as distinct from those associated with individual atoms or molecules or with bulk materials ».

Cette définition met en avant l'importance de l'échelle nanométrique pour au moins une des dimensions des produits en nanotechnologies, mais aussi l'usage des propriétés et phénomènes rendus possibles à cette échelle.

2.3 Classification

Grâce aux nouvelles possibilités offertes par les nanotechnologies, le nombre de domaines d'application est très vaste. Comme pour toute technologie émergente, il n'est pas aisé de classer et organiser les secteurs d'activités impactés par la nanotechnologie. Dans ce but des efforts de sous catégorisation ont été effectués par plusieurs chercheurs afin de mieux définir les activités reliées aux nanotechnologies (Portée et al. 2008, Islam and Miyazaki 2010)

Pour notre étude, nous avons choisi la classification proposée par Portée et al. (2008) et utilisée par la Georgia Tech Technology and Assessment Center (GT CNS-ASU Group). Le choix de cette classification pour la suite de notre analyse tient à sa simplicité ainsi qu'à son nombre réduit de catégories. Le nombre réduit de catégories est un avantage pour la suite de notre projet en particulier pour la classification automatique par apprentissage. En effet, avec un nombre réduit de

catégories, la classification automatique est facilitée et le risque d'erreur de classement fortement diminué. L'inconvénient est d'avoir une précision moindre sur la nature exacte de la technologie en question. Une classification similaire est employée par Islam et Miyazaki (2010) pour leur analyse par domaine des publications en nanotechnologie. Cette classification se compose de quatre sous-domaines qui peuvent se recouper entre eux comme le montre le diagramme Venn de la Figure 2.2.

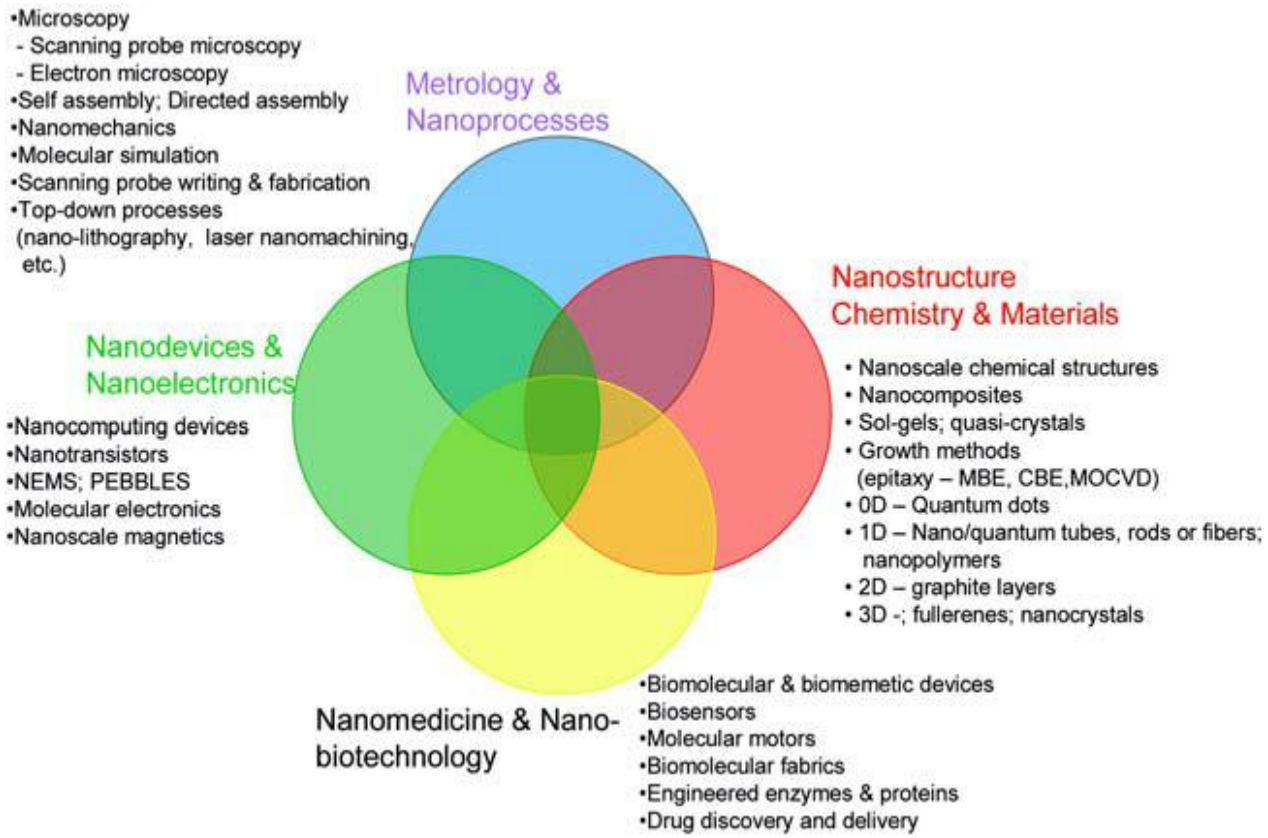

Figure 2.2 : Diagramme Venn des sous-domaines de la nanotechnologie. *Source* : « Field scope of nanotechnology » développé par Georgia Tech Technology and Assessment Center (GT CNS-ASU Group)

Le premier sous domaine, métrologie et nano-processus, concerne l'observation ainsi que la fabrication à l'échelle nanométrique et la nano-mécanique. Plus spécifiquement l'observation est réalisée soit par microscopie à sonde locale soit électronique. La microscopie à sonde locale permet

d’obtenir la représentation d’une surface grâce à une sonde en exploitant la force atomique avec un microscope à force atomique ou l’effet tunnel avec un microscope à effet tunnel. La microscopie électronique est semblable à la microscopie optique, mais elle utilise un faisceau d’électrons pour former l’image, ces derniers ayant une longueur d’onde plus petite que celle des photons, permettant une meilleure résolution. Pour la fabrication elle peut se faire par auto assemblage, par assemblages dirigés, par sonde locale en bougeant les atomes individuellement, par nanolithographie et par nano-usinage laser.

Le sous-domaine nanodispositifs et nanoélectronique fait référence à tous les nanodispositifs informatiques tel que les transistors, aux nanosystèmes électromécaniques (NEMS) i.e. accéléromètre, résonateur, à l’électronique moléculaire ainsi qu’au magnétisme nanométrique.

Dans le sous-domaine nanomédecine et nanobiotechnologie on retrouve les dispositifs et tissus biomoléculaires, les biocapteurs, les moteurs moléculaires d’origine biologique, les enzymes et protéines génétiquement modifiées.

Enfin, le dernier sous domaine, nanostructure et nanomatériaux, concerne les structures chimiques nanométriques, les nanocomposites, les nanotubes, les fullerènes et les nanocristaux. Ces composants servent de briques élémentaires aux constructions nanométriques.

La recherche dans chacun de ces domaines est motivée par des besoins concrets. Par exemple, le besoin constant d’augmenter les performances en électronique pousse à diminuer la taille des composants à l’échelle nanométrique comme le prévoit la loi de Moore (Moore 1965). La recherche de nouvelles propriétés des matériaux ou encore le développement de médicament à administration ciblée sont autant de raisons pour la recherche en nanotechnologie. Néanmoins la recherche dans ces sous-domaines est assez inégale. Comme le montre la Figure 2.3 la recherche en nanoélectronique ainsi que dans les nanomatériaux est très largement supérieure en terme de publication par rapport à la nanobiotechnologie et aux nano-processus qui sont des sous-domaines plus récents (Islam et Miyazaki 2010)

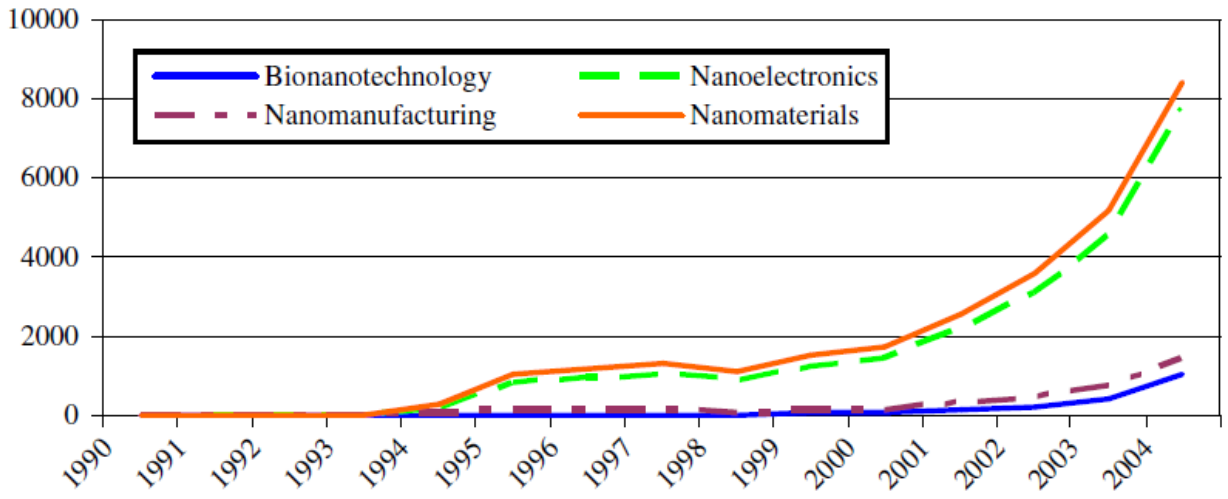

Figure 2.3 : Publication scientifique par domaine de la nanotechnologie (Islam and Miyazaki 2010).

Ces sous-domaines comportent à la fois des technologies complexes et discrètes. Selon Levin et al. (1987) qui ont introduit cette dichotomie dans la technologie, une technologie est complexe si elle nécessite plusieurs brevets complémentaires, souvent possédés par différents titulaires. La nanoélectronique où plusieurs centaines de brevets sont être nécessaire pour un seul produit est un exemple de technologie complexe. Au contraire, pour une technologie discrète telle que la nanobiotechnologie ou la nanochimie, un produit nécessite peu de brevets, voire un seul souvent détenu par un unique titulaire, pour être commercialisé.

Cette différence technologique se traduit par une gestion de la propriété intellectuelle propre à chaque domaine. Les entreprises, selon la complexité de leurs produits, vont avoir des raisons de brevetage ainsi qu'une propension à breveter inégale. Selon Baron et Delcamp (2010), la valeur d'un brevet, suivant qu'il permet directement la commercialisation d'un produit ou qu'il fasse partie d'un réseau de brevet complémentaire pour une seule technologie, va ainsi beaucoup varier. Dans une industrie de produits complexes, les entreprises se partagent des brevets complémentaires pouvant se chevaucher. Ces industries tendent à poser beaucoup de brevets pour bloquer une technologie, mais aussi les concurrents avec des brevets de barrage et renforcer leur position dans les négociations de licences croisées.

Les industries produisant des produits simples tendent à faire un usage des brevets, soit pour produire des licences, soit pour profiter du monopole conféré par les droits de propriété intellectuelle. Cela est possible grâce à la protection suffisante offerte par les brevets pour des produits simples (Cohen, Nelson, and Walsh 2000).

La différence entre technologies complexes et simples a donc une incidence sur le management de l'innovation, et le développement des produits.

2.4 État de l'art

Avec un champ d'application potentielle aussi vaste, et des bouleversements technologiques possibles aussi profonds sur le long terme, la nanotechnologie est une des technologies clés de notre siècle. La nanotechnologie est une technologie critique pour la recherche scientifique, le développement de la technologie et l'innovation de produit et peut être vue comme un domaine à part entier ou une solution générale à divers problème (Mangematin and Walsh 2012).

Dépendamment de la source d'information, de la définition de la nanotechnologie et de la mesure, les prévisions d'évolution de la nanotechnologie varient énormément. En compilant plusieurs sources d'informations, Hullmann (2007) prévoyait un envol du marché des nanotechnologies en 2010. Les estimations, datant du début des années 2000 estimaient que la taille du marché atteindrait entre 150 milliards USD en 2010 (Kamei 2002) et 2.6 billions en 2014 (Lux Research 2007).

Roco (2001) annonçait que dans un horizon de 10 à 15 ans la production mondiale de nanotechnologie dépasserait un billion USD et devrait employer plus de 2 millions de travailleurs pour la production en 2015 en supposant une croissance continue.

Cependant l'estimation d'une croissance quasi exponentielle a été révisée à la baisse à la suite de la crise financière de 2009. En effet, le taux de croissance moyen de 25 % des indicateurs de développement de la nanotechnologie pendant la période de 2000-2008 a été réduit de moitié pendant la crise. Ainsi la production mondiale en 2009 a généré près d'un quart de billion USD dont 91 milliards rien qu'aux États-Unis (Roco 2011).

Avec de telles perspectives économiques, plus de 35 pays ont inscrit les nanotechnologies dans leurs programmes de recherche nationaux (Bensaude-Vincent 2009). Lancée en 2001 aux États-

Unis, la National Nanotechnology Initative (NNI) est considérée comme le point de départ des programmes en nanotechnologie dans les pays développés, bien que plusieurs pays financent déjà de tels programmes depuis le milieu des années 1980. Ainsi, au Canada, l'activité de recherche publique est effectuée par l'Institut national de nanotechnologie (INN) qui est établi en 2001 avec plusieurs organisations provinciales telles que : NanoQuebec, NanoAlberta, Nanotechnology Network of Ontario et British Columbia Nanotechnology Alliance (Saner and Pelley 2009).

D'un point de vue international, la Commission Européenne reste le plus grand investisseur dans la recherche en nanotechnologie en temps qu'agence individuelle avec un budget estimé à plus de 1,3 milliard d'euros entre 2004 et 2006. Au sein des membres européens, l'Allemagne, la France et le Royaume-Uni sont les pays ayant dépensé le plus. Les États-Unis, le Japon et la Corée du Sud ont un budget de recherche public équivalent à ces pays. Cependant, la dépense totale des États-Unis au niveau fédéral dépasse largement ces chiffres avec un total de plus de 1,7 milliard d'euros rien que pour l'année 2007. Parmi les 3,85 milliards d'euros de dépense publique mondiale en R-D pour l'année 2005, le Canada occupe le 16e rang avec 37.9 millions (Hullmann 2007). Le principal investisseur de R-D au Canada est le gouvernement fédéral à travers ses organisations de recherche publique telles que le Conseil de recherches en sciences naturelles et en génie (CRSNG), les Instituts de recherche en santé du Canada (IRSC), la Fondation canadienne pour l'innovation (FCI) et le Conseil national de recherches Canada (CNRC) (Fitzgibbons and McNiven 2006).

La situation actuelle de la R-D reliée à la nanotechnologie au Canada a fait l'objet de plusieurs études. Au niveau de la recherche scientifique, plus précisément de la production d'articles scientifiques, le Canada se place au 13e rang avec 1579 articles publiés en 2005 selon Kostoff et al. (2007). Tahmooresnejad et al. (2015) ont par ailleurs démontré un impact positif linéaire significatif du financement public canadien sur la publication d'articles scientifiques ainsi que sur leur nombre de citations, prouvant leur valeur. Concernant la production de propriété intellectuelle (PI), le Canada se positionne au 5e rang en termes de cessionnaires de brevets associés à la nanotechnologie (Li et al. 2007). Beaudry et Schiffauerova (2011) montrent cependant qu'une grande partie de la PI produite par des chercheurs canadiens en nanotechnologie est absorbée par des compagnies privées américaines via leurs filiales canadiennes; i.e. Xerox Corporation, une compagnie américaine qui détient presque un tiers des brevets en nanotechnologies inventés au Canada.

Une enquête menée en par McNiven (2007) sur les biotechnologies et comportant un volet spécifique sur les nanotechnologies dresse un portrait de l'industrie canadienne ayant des activités en nanotechnologie. Selon les résultats de l'enquête menée en 2005, 88 entreprises canadiennes déclarent avoir des activités en nanotechnologie. Parmi ces entreprises, 91% déclarent effectuer des activités de R-D et 27% d'entre elles sont en phase de production ou de commercialisation. Parmi les entreprises ayant répondu à ce questionnaire 90% sont de petite ou moyenne taille.

Au vu des sommes actuellement dépensées, autant d'un point de vue national que mondial, la R-D et la commercialisation en nanotechnologie est donc un enjeu majeur pour espérer profiter des retombées économiques possibles de cette nouvelle technologie dans un futur proche.

2.5 Commercialisation des nanotechnologies

La nanotechnologie promet de bousculer de nombreux secteurs. Mais avant de pouvoir bénéficier des nombreux avantages offerts par la nanotechnologie, les produits doivent atteindre les marchés en passant par une phase production et de commercialisation réussie.

2.5.1 Phases de commercialisation

Les premiers produits issus des nanotechnologies sont apparus sur le marché dès le début des années 2000 sous forme de nano structure passive avec très peu de réglementation spécifique. Renn et Roco (2006) proposent que la commercialisation des nanotechnologies se déroule plusieurs phases distinctes étalées dans le temps. Ils identifient ainsi quatre générations de produits et procédés en nanotechnologies, des plus simples aux plus complexes, correspondant à quatre phases de commercialisation différentes (voir la Figure 2.4).

Les quatre phases de commercialisation s'étalent dans le temps de la façon suivante :

- Les nanostructures passives dès 2000 i.e. les nanomatériaux ou les nanoparticules ayant des propriétés particulières liées à leurs dimensions.
- Les nanostructures actives à partir de 2005 i.e. les dispositifs nanoélectroniques, les systèmes de distribution ciblée de médicaments.
- Les nano systèmes intégrés à partir de 2010 i.e le développement de la nano robotique, les assemblages de nano système et les systèmes évolutifs.

- -Les nanosystèmes moléculaires à partir de 2015-2020 i.e la conception de système moléculaire au niveau atomique (Renn et Roco 2006).

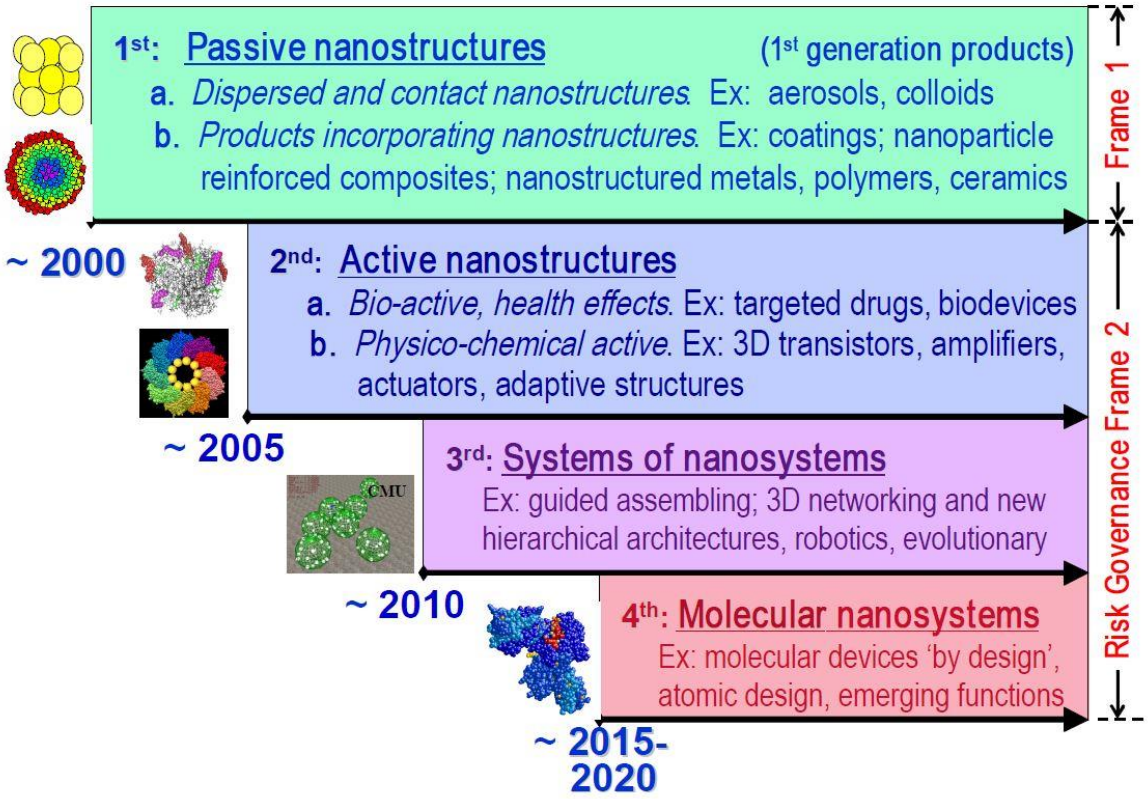

Figure 2.4 : Échéancier de début de prototypage industriel et commercialisation de la nanotechnologie: quatre générations imbriquées de produits et de procédés (Renn et Roco 2006)

2.5.2 Facteurs contribuant à la commercialisation ou à son frein

Malgré les investissements et les efforts de recherche et développement effectués dans le domaine des nanotechnologies au Canada, les retombées économiques ne sont pas aussi élevées qu'attendu. Certains freins à la commercialisation peuvent encore exister, à commencer par l'adoption par le secteur public et les réglementations en cours d'évolution. En effet, avec les rapides progrès de la nanotechnologie, les risques intrinsèques à cette technologie sont encore mal connus. La diffusion de produit de consommation peut donc être entravée, à tort ou à raison, par des facteurs inhérents à cette technologie, mais aussi par plusieurs facteurs externes.

Nous essayons donc de comprendre quels sont les facteurs contribuant à la commercialisation des nanotechnologies, les freins à l'adoption par le grand public et enfin quelles sont les législations en place au Canada pour contrôler la diffusion des nanotechnologies.

Tsuzuki (2013) identifie cinq facteurs critiques pour une commercialisation réussie des nanotechnologies :

- l'orientation du produit (et non l'admiration de la technologie);
- l'interaction continue avec le marché et la sélection d'une application tête de pont;
- l'application de la méthodologie de développement de produits en spirale;
- l'attraction et la rétention de partenaires de commercialisation;
- l'atténuation des risques technologiques propres aux nanotechnologies.

C.-J. Lee et al. (2013) proposent aussi différents facteurs pouvant influencer la commercialisation des nanotechnologies à travers une étude menée sur les entreprises sud-coréennes.

La stratégie de R-D a un fort impact en procurant une supériorité technique aux entreprises facilitant la commercialisation future de leur produit. Cependant, la recherche fondamentale aurait un impact négatif sur la commercialisation en éloignant l'entreprise de son objectif de commercialisation. De plus un financement public de projet de R-D favorise l'entrée de l'entreprise sur le marché et une grande entreprise possédant d'importants moyens financiers a aussi de meilleures chances de commercialisation.

2.5.2.1 Adoption par le grand public

Si la nanotechnologie promet de profonds changements dans notre société dans un futur proche, la majorité des gens n'y connaissent rien et sont même craintifs face à cette nouvelle technologie. L'avis que le public a d'une nouvelle technologie est souvent déterminé par la première image perçue de cette technologie (Grinbaum 2006). La vision drexlerienne (Drexler 1992) de la nanotechnologie, composée de nano robots ou « nanobots » capables de se reproduire de façon autonome, propagée par des œuvres de science-fiction telle que *La Proie* (Crichton and Berthon 2004) a fortement marquée l'imaginaire collectif. L'opinion publique se méfie des nouvelles technologies dont les risques sont encore inconnus et dont les producteurs ne veulent pas communiquer par peur du « syndrome OGM » (Manach 2009).

L'acceptation par le public, nécessaire pour la commercialisation, est particulièrement sensible à l'image véhiculée par la technologie. Selon Grinbaum (2006), un débat public permettant d'informer sur les bénéfices et les risques liés à cette technologie est nécessaire pour améliorer l'image des nanotechnologies et permettre une meilleure commercialisation.

2.5.2.2 Risques liés à la nanotechnologie

Alors que de nombreux produits employant des nanotechnologies sont déjà sur le marché, les risques liés à la nanotechnologie sont encore méconnus. Pour le moment le recul ne permet d'évaluer les risques qui peuvent être de plusieurs natures, tel que les risques liés à la santé des consommateurs et des travailleurs ou les risques liés à l'environnement. Le risque présenté par la nanotechnologie est tel que des compagnies d'assurances classent la nanotechnologie parmi les plus hauts risques émergents[2].

Le nombre de travailleurs fabriquant et utilisant les nanotechnologies augmente avec la demande de l'industrie tout comme le nombre de produits, exposant de plus en plus de personnes à des risques encore mal connus voir inconnus (Colvin 2003). À l'échelle nanométrique, les particules peuvent interagir directement avec le corps humain pouvant causer des dommages internes.

Des études toxicologiques menées sur l'impact des nanotubes de carbones sur le système respiratoire ont par exemple démontré des formations de tumeur pulmonaire dans les poumons de rats (Dreher 2004). Selon Chen et al. (2006) de graves blessures ont été observées sur les reins, le foie et la rate des souris expérimentales exposées à des nanoparticules. Les nanoparticules TiO_2 employées dans les pigments, les cosmétiques, les filtres à air ont aussi un impact sur la santé en s'agglomérant dans le corps humain (Rothen-Rutishauser et al. 2006)

Une certaine prudence est donc justifiée vis-à-vis des nanotechnologies et des réglementations doivent être mises en place pour limiter de tels risques. Tant que les risques ne seront clairement identifiés, les produits considérés comme toxiques à tort ou à raison vont avoir un impact négatif sur la commercialisation.

---

[2] http://www.lloyds.com/~/media/lloyds/reports/emerging%20risk%20reports/er_nanotechnology_report.pdf

2.5.2.3 Normes et réglementations

Les normes et réglementations concernant les nanotechnologies sont mises en place pour limiter les risques et favoriser l'innovation. Il est important de légiférer la commercialisation et la diffusion des nanotechnologies avant que l'industrie ne soit établie pour profiter.

Selon Maniet (2010), il n'existe actuellement aucune réglementation spécifique aux nanotechnologies actuellement au Canada. Un rapport de Santé Canada datant de 2008 estime que la réglementation déjà en place est suffisante pour réguler la nanotechnologie : « The existing Canadian regulatory approaches and risk management strategies are appropriate to this new challenge »[3].

Les nanotechnologies sont donc encadrées par les lois et règlements spécifiques existant déjà dans chaque domaine industriel. Santé Canada liste 6 lois et règlements pouvant intervenir dans la réglementation des nanotechnologies[4].

- Loi sur les aliments et drogues;
- Règlement sur les cosmétiques;
- Règlement sur les additifs alimentaires;
- Règlement sur les aliments et drogues;
- Règlement sur les instruments médicaux;
- Règlement sur les produits de santé naturels.

Le Conseil Canada-États-Unis de coopération en matière de réglementation, créé en 2011, a aussi mis en place une Initiative sur la nanotechnologie dans son plan d'action conjoint avec les États-Unis afin d'harmoniser les réglementations sur les nanomatériaux dans le but « de réduire le risque pour la santé humaine et l'environnement, de promouvoir l'échange d'expertise scientifique et

---

[3] Council of Canadian Academies, *The Small is different*, p.16

[4] http://nanoportal.gc.ca/ consulté en juin 2015

réglementaire et d'encourager l'innovation[5]». Cette initiative « vise le partage d'information et l'harmonisation des approches de réglementation des nanomatériaux, y compris la terminologie et la nomenclature, de même que l'évaluation et la gestion du risque. [3]».

Le site d'Industrie Canada, qui recense les organismes participants aux normes et réglementations des produits en nanotechnologies au Canada et à l'international, liste 9 organismes canadiens et 12 organismes internationaux susceptibles d'intervenir dans la réglementation des nanotechnologies[6].

La nanotechnologie, solution incrémentale à de nombreux problèmes, est susceptible d'intervenir dans de très nombreux produits de consommation. Il n'existe pas encore de réglementation spécifique à la nanotechnologie, mais au contraire un grand nombre d'intervenants, de lois et de réglementations compliquant les démarches de commercialisation de ces produits.

2.5.2.4 Financement public

Le Canada offre de nombreux programmes gouvernementaux et de possibilités de financement aux entreprises pour le développement des nanotechnologies. Nous l'avons vu, le financement public est un élément clé pour le développement d'une entreprise, nouvelle ou déjà existante. Le site d'Industrie Canada[7] recense les financements et programmes suivants :

- Atlantic Canada Opportunities Agency;
- Business Development Bank of Canada;
- Canada Business;
- Canada Economic Development for Québec Regions (CED);
- Canada Revenue Agency—Small and Medium-size Enterprises Zone;
- Canada Small Business Financing Program;

---

[5] http://plandaction.gc.ca/fr/page/rcc-ccr/plan-daction-conjoint-du-conseil-de-cooperation consulté en juin 2015

[6] http://www.ic.gc.ca/eic/site/aimb-dgami.nsf/fra/h_00010.html consulté en juin 2015

[7] http://www.ic.gc.ca/eic/site/aimb-dgami.nsf/eng/h_00005.html consulté en juin 2015

- Federal Economic Development Initiative for Northern Ontario;
- Industrial Technologies Office;
- Industrial Research Assistance Program (IRAP);
- National Research Council Canada;
- National Angel Capital Organization;
- Sustainable Development Technology Canada;
- Western Economic Diversification Canada—Investments in the West.

Ces treize programmes de financement recensés sur le site d'Industrie Canada montrent dans un premier temps l'importance accordée dans le développement des nanotechnologies et les possibilités offertes aux entreprises et aux entrepreneurs de financer leurs projets. Dans un second temps, nous pouvons tirer de ces sources le nom des programmes de financement proposé aux entreprises en nanotechnologie pour identifier des financements en les cherchant sur les sites des entreprises.

2.6 Synthèse

Dans ce chapitre nous avons examiné plus en profondeur notre domaine d'étude. La nanotechnologie est un domaine très vaste composé de plusieurs sous domaines et touchant de nombreux secteurs d'activité. Nous avons pu constater qu'il est encore difficile de définir correctement la nanotechnologie pour cadrer notre recherche. La commercialisation des nanotechnologies est influencée par de nombreux facteurs et les réglementations sont encore susceptibles d'évoluer dans le futur proche, pouvant influencer la recherche, le développement, la production ainsi que la distribution des produits employant des nanotechnologies. Avec un horizon aussi vaste, la nanotechnologie n'est pas un domaine spécifique, mais une solution à de nombreux problèmes, intervenant à de multiples stades de développement des produits.

CHAPITRE 3 SUJET ET MÉTHODOLOGIE DE RECHERCHE

La revue de littérature ainsi que le chapitre sur la nanotechnologie nous ont permis de dresser l'état des connaissances ainsi que le contexte sur lequel s'appuie ce mémoire de recherche.

Ce chapitre présente dans un premier temps le sujet de recherche, la problématique et les objectifs de ce mémoire. Nous aborderons les hypothèses sur lesquelles s'appuie notre recherche, l'objectif principal de ce projet et enfin les objectifs spécifiques réalisés. Dans un second temps nous détaillons notre méthodologie avec les données, les outils et les méthodes d'analyse employées pour mener à bien notre recherche.

3.1 Problématique

Le développement des nanotechnologies offre d'énormes perspectives économiques d'un point de vue mondial et agit comme un générateur de croissance. L'investissement effectué par les gouvernements au travers des initiatives nationales est la preuve de l'intérêt économique porté pour cette technologie. Les initiatives lancées par le gouvernement canadien ainsi que le financement de la recherche au Canada, aussi bien à l'échelle nationale que provinciale, montrent bien l'intérêt et l'implication du Canada pour la recherche et le développement des nanotechnologies. Le programme des chaires de recherche du Canada a investi 290 millions USD dès 2005 dans le but d'attirer les meilleurs chercheurs en nanotechnologie selon un rapport de la Commission Européenne (2005).

Afin que l'investissement public ne soit pas vain, les efforts de recherche doivent être correctement orientés et priorisés par les politiques d'innovation. Pour pouvoir définir et orienter cette vision sur le long terme, il est important de bien comprendre les mécanismes qui amènent le financement public de la recherche à une commercialisation réussie d'une invention technologique.

La revue de littérature identifie beaucoup de facteurs, tels que l'activité de R-D, la collaboration et les sources de financement disponible, qui participent à la réussite commerciale d'une innovation technologique.

Cependant, bien que de nombreux indicateurs existent pour mesurer l'innovation au sens large, il n'est pas aisé de mesurer directement l'impact des innovations sur l'économie.

La mesure de l'innovation est dans ce sens primordial pour rendre compte des efforts déjà effectués, mais aussi des faiblesses qui pourraient apparaître par la politique actuelle.

C'est dans cette optique que de nombreuses études notamment sur le financement de la recherche ont déjà été effectuées au sein de la Chaire de recherche du Canada sur la création, le développement et la commercialisation de l'innovation. Ces études ont été menées avec des indicateurs traditionnels en utilisant des bases de données gouvernementales. Comme nous avons pu le voir, ces bases de données sont issues de questionnaires sur l'innovation avec le faible taux de réponse correspondant à de telles études.

Les grandes compagnies sont généralement bien représentées dans ce type d'études, car elles ont les moyens financiers de publier ou de breveter les résultats de leur recherche, contrairement aux PMEs qui laissent moins de trace de leurs activités.

De nouvelles méthodes d'analyse de l'innovation basées directement sur les sites Internet des entreprises, permettant de s'affranchir des questionnaires, commencent à être utilisées de plus en plus pour l'étude des domaines de haute technologie.

Si de telles études se basant sur les sites Internet des entreprises ont déjà été menées notamment aux États-Unis, en Chine et en Angleterre pour analyser la nanotechnologie (Arora et al. 2013, Gök, Waterworth, and Shapira 2014, Shapira and Youtie 2006, Van der Lei and Cunningham 2006), les sites d'entreprises canadiennes dans le domaine des nanotechnologies sont encore peu, voire pas du tout, exploités pour l'analyse de l'innovation.

Dans notre mémoire nous essayons d'employer ces nouvelles méthodes d'analyses basées directement sur les sites internet des entreprises afin de faire plusieurs regroupement d'entreprise pour donner un aperçu de l'innovation et de la commercialisation des nanotechnologie au Canada.

La problématique principale de ce mémoire est donc la suivante : comment capturer et mesurer l'innovation et l'orientation commerciale des compagnies canadiennes travaillant en nanotechnologie via leurs sites Internet?

3.2 Hypothèses de recherche

L'approche utilisée ainsi que les résultats énoncés dans ce mémoire s'appuient sur deux hypothèses de recherche.

La première hypothèse concerne la répartition des entreprises entre les domaines de la nanotechnologie. Comme nous l'avons dit plus haut, les entreprises profitent de la recherche universitaire pour développer de nouveaux produits. D'après les travaux d'Islam et Miyazaki (2010), les nanomatériaux et la nanoélectronique sont les deux domaines à profiter de la plus forte recherche scientifique. De plus, nous avons vu que selon Renn et Roco (2006) la nanotechnologie est commercialisée par phase correspondant à différentes générations de produits mis sur le marché. Les deux premières phases de la commercialisation concernent les nanostructures passives et actives. Cela suggère que le domaine des nanostructures et nanomatériaux, bénéficiant à la fois d'une forte recherche scientifique et correspondant aux deux premières phases de commercialisation des nanotechnologies, est le domaine le plus représenté au sein des entreprises travaillant en nanotechnologie au Canada. Cela nous permet de poser l'hypothèse suivante :

Hypothèse 1 : a) La répartition des entreprises canadiennes par domaine de la nanotechnologie est inégale et le domaine des nanostructures et des nanomatériaux est le plus représenté. b) Le nombre de produits en nanotechnologies proposés au Canada est plus important dans le domaine des nanomatériaux et nanostructures.

Au Canada, Schiffauerova et Beaudry (2009) ont étudiés la collaboration des inventeurs en nanotechnologie et plus particulièrement l'aspect géographique lié à cette collaboration. En extrayant d'une base de données de brevets tous les inventeurs canadiens avec leur adresse, Schiffauerova et Beaudry montrent que la majorité d'entre eux résident au sein de grappes en nanotechnologies. Nous savons déjà que les brevets procurent un avantage concurrentiel pour une entreprise. Nous proposons donc que les entreprises canadiennes soient localisées dans les mêmes regroupements que les inventeurs en nanotechnologie donnant l'hypothèse suivante :

Hypothèse 2 : Les entreprises canadiennes ayant des activités de nanotechnologie sont localisées dans les mêmes grappes que les inventeurs œuvrant dans le domaine de la nanotechnologie et que le nombre d'entreprises au sein de ces grappes est lié au nombre de brevets par grappe.

Enfin, nous étudions les facteurs influant positivement la commercialisation des nanotechnologies. Dans ces facteurs, Lee et al. (2013) proposent aussi un impact positif de la taille des entreprises, du financement, de l'intensité de la recherche appliquée ainsi qu'une propriété intellectuelle

importante procure un avantage concurrentiel. Toujours selon Lee et al. (2013), la recherche fondamentale en entreprise aurait un impact négatif sur la commercialisation des nanotechnologies. Nous proposons donc de vérifier ces facteurs sur les entreprises canadiennes en employant les données directement présentes dans les sites d'entreprises. Notre dernière hypothèse est donc la suivante :

Hypothèse 3 : a) Les entreprises réussissant le mieux la commercialisation des nanotechnologies sont des grandes entreprises, ayant des sources financement, une forte activité de R-D, une propriété intellectuelle importante ainsi que des activités de collaboration et b) la recherche fondamentale influence négativement la commercialisation des nanotechnologies.

3.3 Objectifs de recherche

L'objectif des travaux effectués dans ce mémoire est d'analyser l'innovation et la commercialisation de la nanotechnologie au Canada via les sites Internet des entreprises.

Pour atteindre cet objectif principal, nous avons répondu à plusieurs objectifs spécifiques :

- Mettre en place un processus de récupération et de nettoyage des données présentes dans les sites Internet des entreprises;
- Identifier et classer les entreprises canadiennes en nanotechnologie selon la littérature;
- Identifier les domaines de la nanotechnologie les plus actifs dans la commercialisation au Canada et estimer le nombre de produits par domaine;
- Construire des regroupements géographiques à partir des informations contenues dans les sites Internet;
- Créer une matrice d'indicateurs d'innovation basés dans les sites Internet pour identifier les entreprises les plus innovantes et les plus à même de réussir la commercialisation de produits en nanotechnologies.

3.4 Données

Pour notre étude nous avons employé des données de plusieurs sources. Nous verrons quelle a été leur utilité et d'où elles proviennent.

3.4.1 Entreprises

Notre étude porte sur la mise en place d'une méthode pour classer les entreprises selon leur degré d'innovation dans le domaine des nanotechnologies via leur site Internet public. Nous avons donc cherché à obtenir le maximum de sites d'entreprise travaillant dans le domaine de la nanotechnologie. Pour cela nous avons employé plusieurs sources répertoriant des compagnies canadiennes dans ce domaine. Plusieurs sites Internet recensent ce type de compagnie au niveau fédéral ou provincial. Le choix des compagnies pour notre étude répond à deux critères. La compagnie doit faire partie d'une base de données canadienne sur les nanotechnologies et la compagnie doit avoir un site Internet accessible au public.

Nous avons répertorié plusieurs sources listant des entreprises pour notre étude. Après avoir récupéré l'ensemble des sites Internet des différentes sources, nous obtenons un total de 880 sites d'entreprises répertoriées en nanotechnologie. Nous avons effectué un premier tri d'entreprises pour éliminer les doublons et les sites n'existant plus ou n'étant pas accessibles. Au total nous avons donc 809 sites d'entreprises différentes et accessibles. Le Tableau 3.1 donne la répartition des entreprises selon les différentes sources employées.

Tableau 3.1 : Sources des entreprises en nanotechnologie

Source	**Nombre d'entreprises répertoriées**
NanoAlberta (http://www.albertatechfutures.ca/nanoAlberta.aspx)	162
NanoOntario (http://nanoontario.ca/)	89
NanoWerk (http://www.nanowerk.com/nanotechnology/nanomaterial/commercial_country.php?country=Canada)	69
Industrie Canada (https://www.ic.gc.ca/eic/site/aimb-dgami.nsf/eng/03503.html)	264
British Columbia Nanotechnology Alliance (http://www.nems.ca/nanobc/#home)	12
AGY Consulting	284
Total :	880
Total accessible et dé-doublonné :	673

Une des limitations de l'emploi de ces sources est la difficulté de capturer efficacement l'ensemble des entreprises canadiennes travaillant effectivement dans le domaine des nanotechnologies. L'inscription étant généralement basée sur le volontariat des entreprises, n'importe qui peut inscrire son entreprise pour bénéficier d'une meilleure visibilité sans pour autant travailler dans le domaine de la nanotechnologie. Inversement, il est difficile de savoir si toutes les entreprises travaillant dans

les nanotechnologies sont effectivement inscrites dans au moins une des sources. Internet étant continuellement en mouvement, les sites listés peuvent aussi avoir changé de nom ou avoir été supprimés, réduisant le nombre de sites exploitable pour notre recherche.

3.4.2 Source de financement

Pour identifier les sources de financement des entreprises, nous essayons de trouver les noms des programmes de financement public disponibles pour la nanotechnologie sur les sites des entreprises afin de voir si elles ont bénéficié d'une aide publique. Ces programmes mis en place pour aider les entreprises à financer de nouveaux programmes pour leurs activités de R-D et pour le lancement de nouveaux produits sont donnés dans la revue de littérature. Nous cherchons donc sur les sites Internet des entreprises le nom des programmes de financement pour savoir si l'entreprise a bénéficié d'une aide spécifique.

3.5 Méthodologie de recherche

Premièrement nous étudions la littérature pour établir la faisabilité d'une analyse portant sur l'innovation et la commercialisation via les données extraites de sites Internet d'entreprises. Une fois avoir validé la possibilité d'une telle étude nous décrivons les deux grands axes de notre méthodologie de recherche, l'acquisition des données par un robot et le traitement de ces données.

La Figure 3.1 décrit la démarche générale suivie au cours de ce projet de recherche pour répondre aux objectifs de recherche. Les données employées au cours de ce mémoire proviennent exclusivement des sites internet des entreprises. Le point de départ de notre recherche est la récupération des données sur les sites des entreprises qui seront employées pour mener nos analyses. Notre analyse se divise ensuite en trois grands axes.

Un premier axe se propose de classer les sites d'entreprises par domaine de nanotechnologie avec des techniques d'apprentissage machine. Un second axe de recherche concerne les regroupements des entreprises par grappe géographique et dans le troisième axe nous regroupons les sites d'entreprises avec des indicateurs construits à partir de la littérature sur l'innovation et la commercialisation.

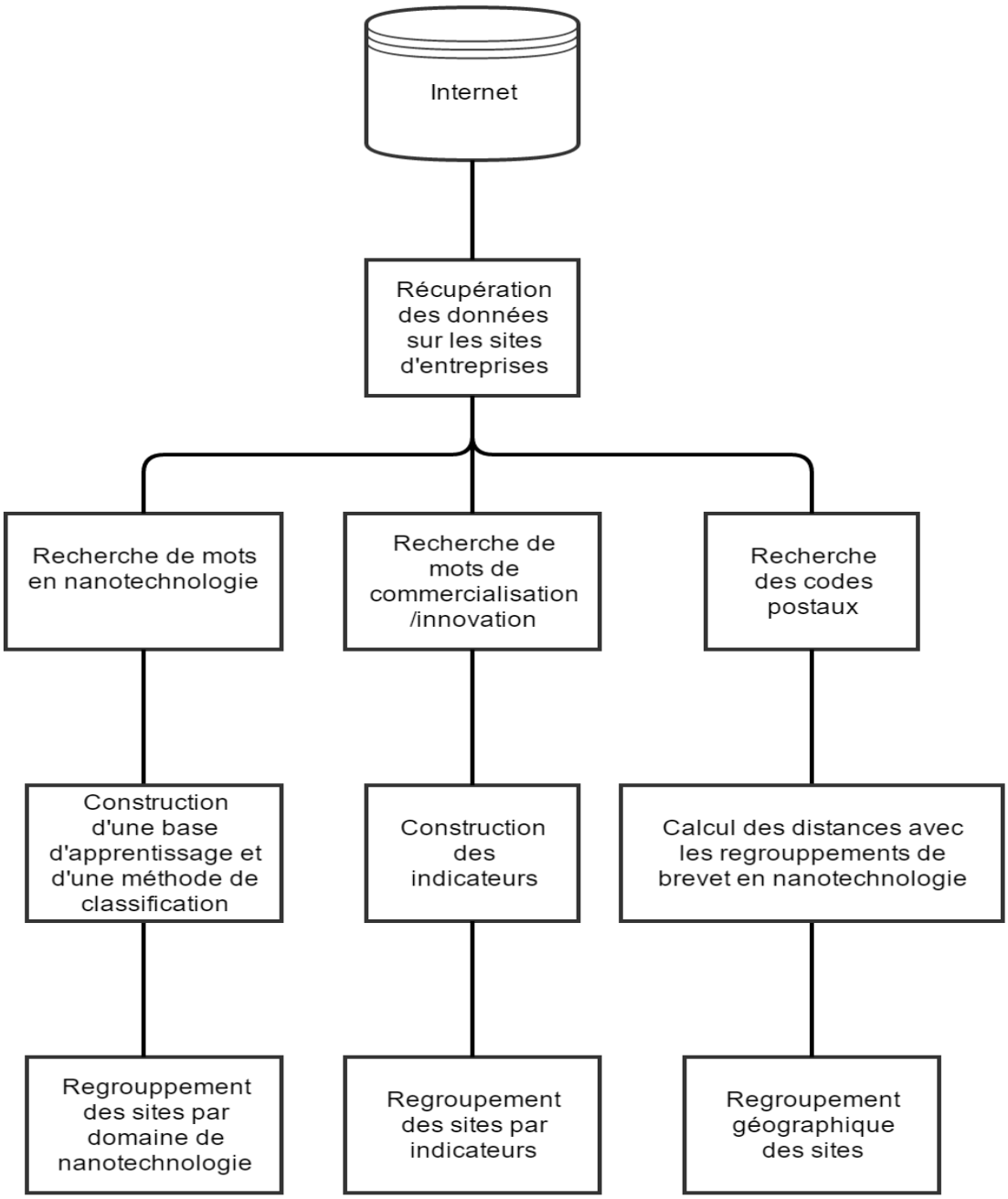

Figure 3.1 : Méthodologie générale de recherche

3.5.1 L'exploration de données sur Internet

La tendance des entreprises à donner toujours plus d'informations sur leurs sites Internet publics favorise les techniques d'analyse non intrusive telle que l'exploration de données Internet. Les sites Internet sont des sources d'information complémentaires pour les études sur les nouvelles

technologies. En effet, les entreprises se servent de leurs sites Internet pour promouvoir leurs identités visuelles ainsi que les valeurs qu'elles souhaitent nous transmettre. En plus des informations sur la localisation, les produits et services proposés, les compétences associées à l'entreprise, il est courant d'avoir des informations financières ainsi que les partenariats avec d'autres compagnies, universités et organismes gouvernementaux.

L'analyse non intrusive telle que l'exploration de données Internet a pour avantage de ne pas être en contact directement avec le sujet et reste à une certaine distance d'observation. Ainsi le sujet étudié ne peut pas adapter son comportement à l'étude comme cela peut être observé avec des questionnaires ou des entrevues. Généralement ce type d'étude est bien moins onéreux par rapport aux études dites intrusives, comme les questionnaires et les entrevues, nécessitant des chercheurs pour récolter les données (Lee 2000).

De plus les études d'innovation ont tendance actuellement à s'appuyer de plus de plus sur des questionnaires en ligne que les entreprises doivent compléter elles-mêmes. Selon Sauermann (2013), les études d'innovations basées sur ces sondages Internet obtiennent généralement de faibles taux de réponses, entre 10 et 25%, affectant les résultats des analyses par un biais de non-réponse. La complexité ainsi que le temps nécessaire pour répondre aux questionnaires en ligne étant les principales raisons de non-réponse de la part des entreprises.

Le concept de l'exploration de données Internet consiste à récupérer des informations Internet des compagnies pour les convertir en données d'analyse. Une très large majorité des compagnies travaillant dans des domaines technologiques tiennent à jour un site Internet pour informer les potentiels clients et investisseurs. Bien que les informations soient mises en ligne par les entreprises elles-mêmes, cette source d'information libre d'accès est adaptée à l'étude des technologies émergentes telles que la nanotechnologie (Gök et al. 2014). De plus Youtie et al. (2012) notent que les petites entreprises ont tendance à avoir des sites plus restreints, facilitant la manipulation des données. Il est évident que les entreprises ne dévoilent pas toutes leurs données stratégiques et commerciales dans leurs sites Internet, mais ce n'est pas non plus le cas au travers des autres sources de données telles les publications scientifiques ou les brevets.

Les analyses d'innovation employant l'exploration des données Internet comme données d'analyses ont donc plusieurs avantages sur les analyses par questionnaire. L'ensemble des données est présent sur le site Internet, libre d'accès, mis à jour régulièrement et le taux de réponse

est bien plus élevé, car la majorité des entreprises tiennent à jour un site Internet potentiellement utilisable par ce type d'analyse.

L'exploration de données Internet est composée en plusieurs sous domaines permettant d'étudier différents aspects des données présentes sur Internet. Actuellement on dénombre trois catégories distinctes (Miner et al. 2012).

L'exploration des données d'utilisation Internet (*web usage mining)* est consacrée à l'étude de l'utilisation des sites Internet en analysant la géolocalisation des visiteurs, les journaux de connexion des serveurs ainsi que les clics sur la page.

L'exploration de structure Internet (*web structure mining)* est une méthode d'analyse des réseaux utilisant les liens hypertextes reliant entre eux plusieurs sites Internet.

Les analyses dites webométriques, mesurant les activités scientifiques présentes sur Internet, emploient une combinaison des deux précédentes analyses.

Enfin l'exploration de contenu Internet (*web content mining)*, plus récente que les précédentes, s'intéresse aux informations brutes directement contenues sur les pages des sites telles que le texte, les images et les chiffres, pour tenter d'extraire des données structurées.

Van der Lei et Cunningham (2006) ont établi un réseau d'activités entre les portails Internet en nanotechnologie afin identifier les différents acteurs du milieu en menant une analyse de type exploration de structure Internet. Dans leur article, Gök et al. (2014) ont utilisé l'exploration de contenus Internet pour mesurer les activités de R-D de 296 entreprises du Royaume-Uni.

En analysant les sites Internet avec une liste de mots clés et en comparant les résultats des indicateurs obtenus grâce aux sites Internet avec d'autres indicateurs issus de sources de données conventionnelles, ils ont trouvé des activités de R-D non repérables par les données conventionnelles.

Comme nous l'avons vu, l'intérêt d'utiliser les données d'Internet est multiple. La plupart des entreprises, indépendamment de leur taille, possèdent et maintiennent à jour un site Internet. La population couverte par une étude employant la fouille du Web (web mining) est donc très élevée dans un domaine où les études par questionnaires obtiennent peu de retours et plus particulièrement dans le domaine des nouvelles technologies. L'accès à l'information est grandement facilité et

contrairement aux données gouvernementales, la fréquence de mises à jour est élevée, voire quotidienne, dans la plupart des cas.

Ainsi l'information contenue dans les sites Internet est parfaitement adaptée à plusieurs types d'études possibles dans le domaine des nouvelles technologies. Le principal inconvénient réside en la difficulté à structurer et interpréter les données, car chaque site possède des informations ainsi qu'une organisation différente.

3.5.2 Mise en place d'un protocole d'analyse

Avant de pouvoir commencer à analyser les sites Internet des compagnies travaillant dans le domaine de la nanotechnologie nous avons mis au point un processus de récupération et de nettoyage des données non structurées présentes sur Internet.

Notre processus, décrit dans la Figure 3.2 comprend quatre étapes distinctes entre les données brutes d'Internet et notre analyse. Dans la première étape, l'acquisition des données, nous commençons par extraire le texte des sites Internet grâce à un logiciel de moissonnage de données. La deuxième étape correspond au stockage de ces données dans un système de gestion de base de données (SGBD). Dans la troisième étape, nous utilisons des scripts pour sélectionner et traduire si besoin les données pertinentes pour notre analyse à partir du SGBD. Enfin, la dernière étape consiste à traiter ces données avec un logiciel de forage de données.

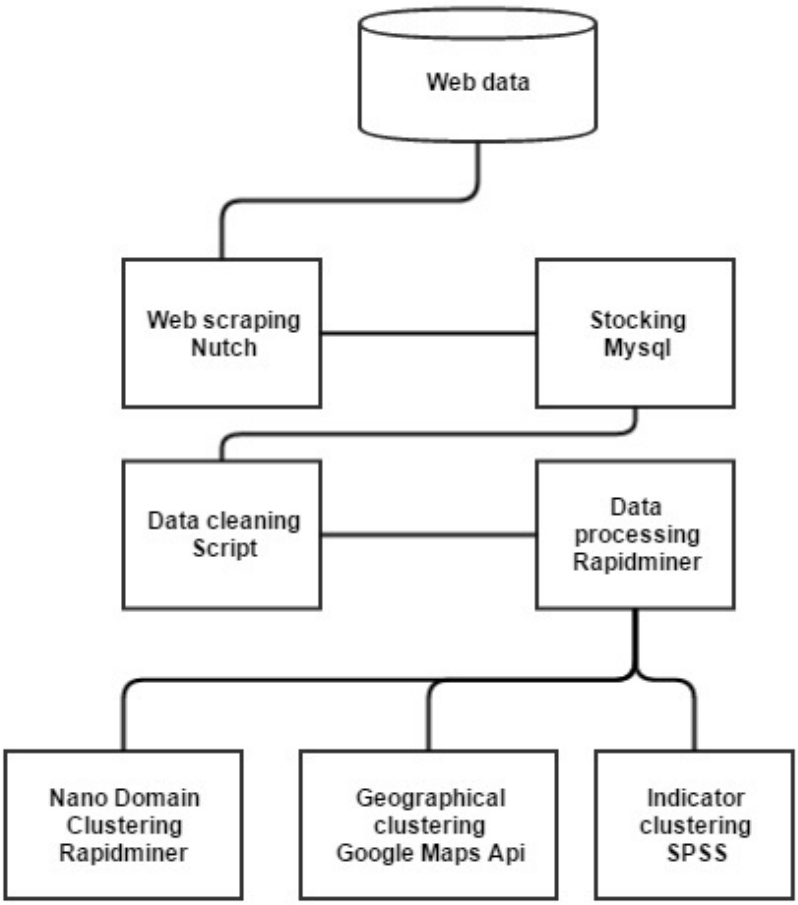

Figure 3.2 : Processus d'acquisition et de traitement des données

Nous avons réalisé la majorité de notre étude sur un serveur Linux vingtras 3.13.0-35-generic. Travailler sur un serveur possède plusieurs avantages. Le serveur offre des performances de capacité de calcul et de stockage bien supérieur aux ordinateurs personnels. De plus, les données ainsi que les logiciels sont accessibles par plusieurs utilisateurs simultanément et à distance via un terminal.

3.5.3 Récupération des données

La récupération des données est la première étape du processus et l'une des plus importantes. Le processus de récupération des données doit répondre à plusieurs critères. Il doit être automatique, adapté à un volume important de données, rapide et configurable selon nos besoins. Le but principal

de cette étape est de récolter toutes les données nécessaires à l'étude et de les stocker dans un même endroit pour faciliter la suite du processus. L'information contenue sur Internet est susceptible d'être modifiée pendant la durée de notre étude d'où l'intérêt de stocker toutes les données pendant le processus d'analyse.

3.5.3.1 Choix du robot d'indexage

Un robot d'indexage est un logiciel qui se comporte comme un utilisateur testant systématiquement tous les liens présents sur la page qu'il parcourt en enregistrant les informations. Il va ainsi progresser lien par lien en indexant tout le texte disponible sur les pages parcourues.

Il existe de nombreux robots d'indexage disponibles offrant tous des prestations similaires. Nos critères de choix sont d'avoir un logiciel libre de droit, modifiable facilement selon nos besoins et supporté par une communauté active pour avoir un support en cas de problème.

Voici une liste non exhaustive de robots d'indexage répondant à nos critères :

- Apache Nutch;
- Heritrix;
- Crawler4j;
- Scrapy.

Notre choix c'est porté sur le logiciel Nutch dans sa version 2.2.1 de la Fondation Apache. Nutch est un outil puissant avec de nombreux développeurs y contribuant, une forte communauté, codé en Java et entièrement paramétrable. Nutch supporte beaucoup de SGBD et permet de récupérer de grandes quantités de données.

3.5.3.2 Fonctionnement

Comme nous l'avons vu, un robot d'indexage est un logiciel capable de suivre des liens comme un utilisateur lambda et de récupérer le contenu des pages parcourues pour les stocker. On peut dès à présent annoncer deux limitations au robot d'indexage. Premièrement, le robot se comportant comme un utilisateur non identifié, il ne peut parcourir que les pages qui sont accessibles au public. Enfin, le robot indexe les pages en suivant des liens. Une page indépendante dans l'indexation d'un site ne pourra pas non plus être parcourue.

Dans la Figure 3.3, nous décrivons le processus de récolte des données de Nutch.

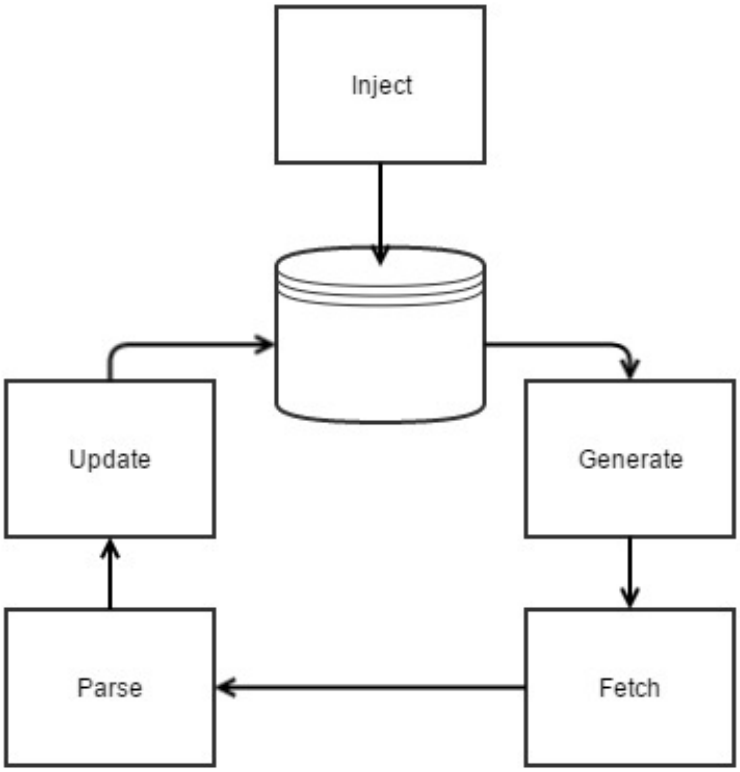

Figure 3.3 : Processus de récolte de donnée du logiciel Nutch

Pour initialiser le robot d'indexage, nous devons lui fournir une liste de sites initiale. Cette liste constitue la première profondeur d'indexage parcourue par le robot pour récupérer les liens des futures pages à parcourir et le texte. Pour débuter le processus de récolte des données, on écrit dans un fichier texte les liens *uniform resource locator* (URL) permettant d'accéder aux sites Internet des entreprises. Les URL sont ensuite injectées dans la base de données grâce à la commande suivante: `bin/nutch inject urls`

Une fois la base de données renseignée des sites à parcourir, on génère le parcourt du robot en lui indiquant le nombre maximum de liens par page N que l'on souhaite récolter. Le robot classe les pages avec un score qu'il calcule pour prioriser la collecte du texte. Une page aura un score élevé si elle est pointée par beaucoup d'autres pages. Plus les pages pointant vers une autre auront des scores élevés, plus le score de la page pointée sera élevé. Nutch sélectionne alors les liens des pages

ayant le meilleur score à récolter en premier. La commande suivante permet de générer le parcourt du robot : `bin/nutch generate -topN N`

Lorsque les liens à parcourir sont générés, Nutch va récolter tout le code HTML des pages ainsi que les liens pointant vers les autres pages avec la commande : `bin/nutch fetch –all`

Il existe principalement trois types de liens différents qui permettent de structurer le réseau Internet. Les liens en amont partent d'une page pour atteindre une ou plusieurs pages non nécessairement sur le même site. Les liens en aval correspondent à tous les liens pointant vers une même page. Enfin les liens verticaux sont tous les liens au sein d'un même site Internet (Lei et Cunningham 2006).

Comme nous l'avons vu, le robot d'indexage parcourt tous les liens sans faire de distinction de domaine et les classe selon un score pour parcourir en premier les liens qu'il juge plus intéressants.

Ce comportement du robot n'est pas en accord avec notre étude, car il va indexer d'énormes quantités de données sans rapport avec notre étude. Les pages d'un site d'une encyclopédie en ligne avec de très nombreux liens obtiennent par exemple de très bons scores sans fournir de données pertinentes pour notre étude. Nous avons modifié le comportement du robot pour nous limiter aux liens verticaux des domaines présents dans la liste initiale. Ceci est possible en modifiant le fichier de configuration de Nutch limitant le robot à la liste de domaines initiale. Ainsi, le robot ne va indexer que les données présentes dans les sites de la liste d'origine.

Par défaut, Nutch indexe uniquement le texte des pages au format *HyperText Markup Language* (HTML). Cependant des informations pertinentes sont aussi stockées sous d'autres formats numériques. Nous utilisons le *plug-in* Tika pour parcourir et indexer le texte des documents au format Portable Document Files (.pdf) et les documents Word (.doc).

Une fois toutes les pages analysées avec le texte et les liens URL, nous pouvons mettre à jour la base de données avec la commande :`bin/nutch updatedb`.

Les liens indexés lors de cette étape serviront de départ pour le prochain niveau de moissonnage.

3.5.3.3 SGBD

Pour le stockage des données, nous avons comme critère de sélection un SGBD compatible avec notre robot d'indexage, libre de droits, permettant le stockage de grandes bases de données et bien documenté.

Voici une liste des principaux SGBD répondant à nos critères :

- MySQL;
- HBase;
- Cassandra;
- Accumulo;
- Avro;
- MongoDB.

Nous avons sélectionné Mysql qui est le SGBD libre le plus populaire et le plus simple d'utilisation[8]. Il offre une capacité de stockage suffisante pour notre étude et est libre de droits.

Le stockage de nos données est important, car le support d'information qu'est Internet est de moins en moins stable. Les pages sont régulièrement créées, mises à jour ou supprimées. L'intérêt de stocker les informations est de s'assurer que les données ne vont pas changer pendant la durée de l'étude et de pouvoir les manipuler plus facilement en connectant les autres logiciels directement à MySql.

3.5.4 Traitement des données

Le nettoyage préliminaire des données est une des étapes principales de la fouille de donnée. Cette étape permet de supprimer les données non pertinentes pour notre étude. Nous effectuons un premier nettoyage grâce à des procédures écrites en langage Python. Le rôle de ces scripts est de créer une nouvelle table dans notre SGBD exploitable pour la suite de l'analyse avec l'ensemble de nos données pertinentes.

---

[8] Source : http://db-engines.com/en/ranking

Les critères de pertinence sont d'avoir une page Internet correctement indexée par le robot, d'avoir du texte sur la page et d'être écrite en français ou en anglais.

En effet, un problème pour analyser les sites des compagnies canadiennes est la gestion de la langue. Notre étude portant sur l'analyse de la nanotechnologie du Canada nous devons tenir compte du fait que l'information peut être donnée soit en français soit en anglais.

Notre analyse étant effectuée uniquement avec des mots anglais, il existe plusieurs cas différents à traiter. Pour cela nous avons fait plusieurs hypothèses. Soit le site est entièrement en anglais, ce qui correspond à une majorité de sites de sociétés canadiennes et nous gardons alors toutes les données du site. Soit le site est disponible en plusieurs langues, dont l'anglais. On ne prend alors en compte que les pages écrites en anglais. Soit le site est exclusivement disponible en français et il est nécessaire de traduire le texte. Et enfin, le dernier cas est une page dont la langue n'est ni en français ni en anglais. Une telle page n'est pas considérée comme intéressante pour notre étude est n'est pas conservée pour la suite de l'étude.

La détection de la langue est effectuée par le logiciel Compact Language Detector tandis que la traduction est effectuée avec l'interface de programmation de Google, Goslate. La traduction avec Goslate est aussi employée par Arora et al. (2013) pour analyser le contenu des sites d'entreprises chinoises fabriquant du graphène.

Le script va donc affecter une information sur la pertinence ainsi que sur la langue de chaque page stockée dans la base de données.

Pour les premiers essais, nous avons affiché la langue de chaque texte et comparé les résultats du script avec la page en question pour gérer les aberrations. Le script renvoie la probabilité que la page soit dans une certaine langue. Étonnamment de nombreuses pages ont été affectées de la langue latine, ce qui a attiré notre attention. Cela s'explique par l'habitude des développeurs d'afficher un texte en latin (« lorem ipsum ») en attendant le texte définitif de la page.

Nous avons finalement choisi de n'afficher que « français », « anglais » et « autre ».

Un deuxième script utilise cette information pour stocker les pages pertinentes dans une deuxième table en traduisant les pages françaises en anglais. Nous avons décidé de stocker les données après traitement dans une deuxième table de notre SGBD pour garder les données brutes de l'étape de récupération.

Enfin un troisième script permet de créer une nouvelle table en concaténant l'ensemble du texte récupéré par domaine Internet. Une nouvelle colonne est aussi ajoutée afin de connaitre le nombre de pages récoltées par domaine.

Le Tableau 3.2 donne un aperçu des données employées pour l'analyse. En moyenne, le robot a récupéré 25 pages de texte pertinentes par site avec un maximum de 105 pages. Chaque page est composée en moyenne de 520 mots. Au total il s'agit donc d'analyser plus de 6 millions de mots.

Tableau 3.2 : Données récoltées

Nombre de sites récupérés	Nombre total de pages pertinentes	Moyenne de page par site	Moyenne de mot par page
487	12278	25	520

Comme nous pouvons le constater, le nombre de pages par domaine varie énormément. Si certains sites se contentent de quelque page, d'autres sont beaucoup plus étendus. Généralement, comme le font remarquer Youtie et al. (2012), les petites entreprises, ayant peu de produits ou de services à proposer, ont tendance à avoir des plus petits sites.

La différence entre le nombre de sites récupérés, 487, et le nombre de sites présents dans nos sources, 809, correspond aux sites inaccessibles par le robot. Un site inaccessible par le robot est un site ayant été supprimé, ayant changé de nom ou ayant bloqué les robots d'indexage.

3.6 Forage de données

Pour analyser toutes les données récoltées par le robot d'indexation, nous utilisons des techniques de forage de données. Le forage de données regroupe tous les processus d'analyse de grandes quantités d'information pour tenter de découvrir des corrélations entre plusieurs documents de façon automatisée. Cette technique est adaptée à l'analyse des domaines de la haute technologie (Shapira and Youtie 2006, Fall et al. 2003, Gök et al. 2014, Domeniconi et al. 2005, Dong et al. 2004, Arora et al. 2013).

Pour analyser les données récoltées, nous avons utilisé un logiciel de forage de données, Rapidminer Community Edition. Ce logiciel a été choisi pour sa puissance de calcul et son

utilisation intuitive. Nos données étant en format texte nous utilisons le module spécifique Text Mining de Rapidminer dédié au forage de texte.

3.6.1 Classification automatique par apprentissage

Nous désirons classer les sites Internet stockés dans notre base de données suivant la classification de notre revue de littérature. Notre objectif est de répartir chaque site Internet présent dans notre panel d'entreprises canadiennes travaillant dans le domaine des nanotechnologies dans un sous domaine de la nanotechnologie. Il existe plusieurs moyens d'obtenir ce résultat. Pour un petit nombre de sites à classer, effectuer manuellement la classification en ouvrant chaque site Internet et en lisant son contenu est surement la meilleure solution. Dans le cas de notre étude nous tentons de classer plus de 400 entreprises il est donc préférable de trouver un moyen d'automatiser la tâche.

3.6.1.1 Traitement préliminaire des données

Dans un premier temps nous devons récupérer le texte stocké et le transformer avec le logiciel Rapidminer pour appliquer des algorithmes de forage de donnée.

Pour cela nous devons transformer notre texte en vecteur numérique. Plusieurs créations de vecteurs sont possibles telles que la fréquence d'apparition des mots, leur présence-absence, leur nombre d'occurrences ou encore le vecteur Term Frequency – Inverse Document Frequency (TF-IDF).

Pour la classification, certains mots ont plus d'importance que d'autres et il est important de refléter cette importance relative pour distinguer les documents. Le vecteur TF-IDF permet de visualiser l'importance d'un mot au sein d'un corpus de documents en lui donnant un poids. Cette méthode est beaucoup employée notamment par les moteurs de recherche Internet. Si un mot apparait de manière fréquente dans un document il aura un poids élevé. Ce poids est pondéré par le nombre de documents du corpus contenant ce mot, car si un mot se répète dans tous les documents il devient inutile pour la classification (Ramos 2003).

Le vecteur TF-IDF se calcule de la manière suivante:

$$w_{i,j} = tf_{i,j} \times \log(\frac{N}{df_i})$$

Avec $tf_{i,j}$= nombre total de termes i dans j

df_i = nombre total de documents contenant i

N = nombre total de documents

Afin de créer la matrice des vecteurs TF-IDF pour procéder à la classification automatique des documents, nous devons procéder à une phase de préparation des données. Le traitement se décompose en plusieurs sous étapes :

- Transformer le texte en minuscule;
- Séparer le texte en unité lexicale (token). Jusqu'à présent le logiciel traite le texte dans son ensemble comme un flux;
- Filtrer les mots vides pour alléger la suite du traitement. Beaucoup de mots présents dans les textes des sites Internet sont inutiles pour la classification. Il s'agit de l'ensemble des mots trop fréquents. Pour alléger la suite du processus et pour une meilleure visibilité des résultats, ces mots vides, contenus dans une base de données de mots anglais, sont éliminés. Ces mots n'ont aucun rapport avec notre étude des nanotechnologies;
- Raciniser les unités lexicales avec un algorithme adapté à l'anglais. La racinisation permet de transformer les tokens pour ne garder que la racine d'un mot. Cette opération permet aussi d'alléger le processus en combinant plusieurs mots ayant la même racine. Nous utilisons l'algorithme de Porter car il est adapté à la langue anglaise;
- Filtrer les tokens suivant leurs nombres de caractères et ne garder que les mots entre 3 et 99 caractères pour éviter les erreurs.

Nous obtenons à cette étape une matrice de vecteurs TF-IDF correspondants à l'ensemble des mots présent dans tous les documents récupérés sur Internet. Cette matrice de vecteurs nous permet d'utiliser les algorithmes de forage de données pour effectuer notre classification automatique.

3.6.1.2 Méthodes d'apprentissage

Une fois le traitement de texte effectué nous pouvons passer à la classification à proprement parler.

Il existe deux types majeurs de classifications différentes, à savoir l'apprentissage non supervisé et l'apprentissage supervisé. L'apprentissage non supervisé est employé pour effectuer des

regroupements de documents sans connaitre le nombre de classes ni leur définition. L'algorithme va lui-même devoir déterminer les classes en regroupant les documents ayant des caractéristiques communes. L'apprentissage supervisé permet au contraire de classer des documents dans des classes connues d'avance en apprenant les classes au logiciel.

Pour classer les sites Internet par domaine de la nanotechnologie, nous employons donc l'apprentissage supervisé. Le processus de classification automatique est décrit dans la Figure 3.4.

Afin de pouvoir classer les documents suivant des classes prédéfinies il est nécessaire de les apprendre au logiciel grâce à une base d'apprentissage. Le processus de classification se divise donc en deux parties; l'apprentissage et la classification. Pour l'apprentissage des classes, nous devons créer une base d'apprentissage qui va permettre au logiciel de créer des règles de classification à appliquer au reste des documents.

La base d'apprentissage est créée manuellement en classant un échantillon de cinquante sites Internet (10% du nombre total de sites) prenant en compte tous les domaines de la nanotechnologie. L'identification des sites se fait manuellement grâce à une lecture rapide du site Internet.

Il existe de nombreuses méthodes d'apprentissages possibles pour la comparaison des documents, chaque méthode étant plus ou moins précise et robuste. Kotsiantis et al. (2007) évaluent plusieurs caractéristiques de six méthodes d'apprentissage, toutes disponibles dans Rapidminer, à savoir Decison Trees, Neural Networks, Naïve Bayes, kNN, Support Vector Machin et Rule learner. De plus, afin d'augmenter la précision Kotsiantis préconise l'emploi d'une méthode de combinaison par vote qui attribue la classe par rapport à la majorité des prédictions données par les méthodes de classification.

Pour notre projet nous utilisons Decision Tree, Naïve Bayes et kNN pour leur rapidité d'apprentissage, leur précision et leur capacité à gérer plusieurs catégories. Ces trois méthodes sont combinées par vote comme le propose Kotsiantis.

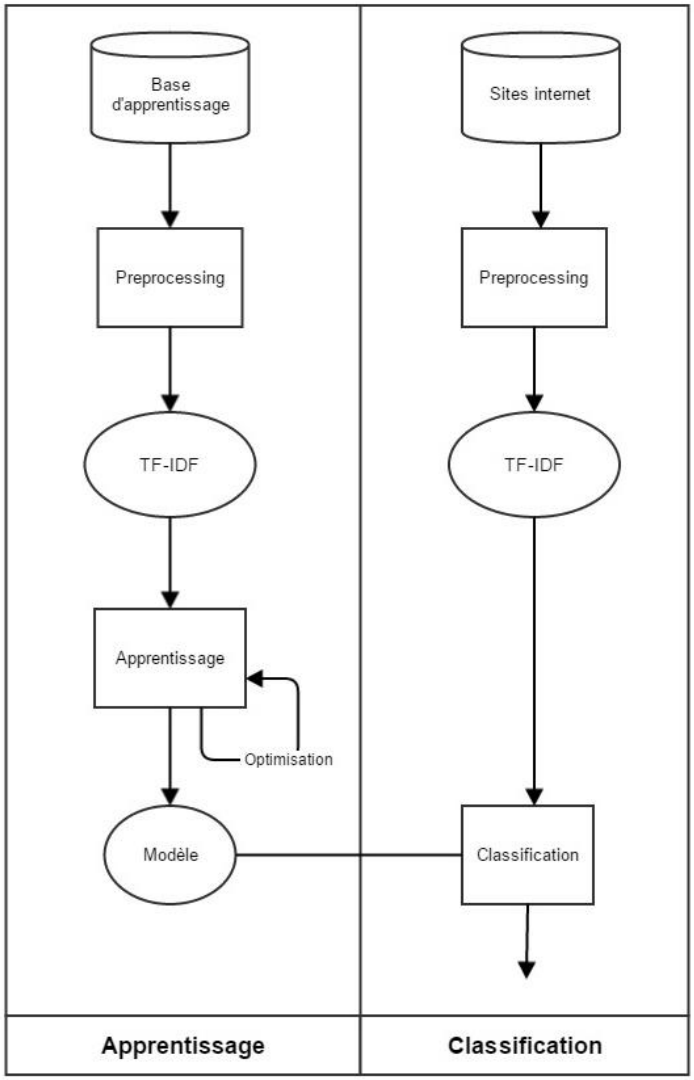

Figure 3.4 : Processus de classification automatique

Les méthodes d'apprentissages possèdent un ou plusieurs paramètres permettant d'améliorer la précision de la classification. Nous utilisions une boucle d'optimisation pour optimiser la précision moyenne de l'apprentissage en réalisant des combinaisons entre les paramètres des modules d'apprentissage. Les paramètres ajustés sont : le paramètre K pour la méthode kNN, la correction Laplace pour Naive Bayes et le nombre minimum d'instances par feuille pour la méthode Decision Tree.

Pour effectuer la validation du modèle de classification, nous utilisons un opérateur composé de deux sous-processus. L'opérateur divise notre base de données d'apprentissage en deux échantillons de façon aléatoire et l'apprentissage est effectué sur un échantillon et validé sur l'autre.

L'opérateur répète les mêmes opérations en inversant l'ordre des échantillons puis retourne la matrice de confusion moyenne des deux modèles.

Une fois l'apprentissage effectué et validé, on applique le modèle aux sites Internet non classés après avoir effectué le même traitement de texte que pour la base d'apprentissage.

Les résultats obtenus lors de la validation du modèle de classification avec cette méthode sont cependant très bas avec une précision générale de 22%.

Nous avons constaté que le logiciel ne fait que des associations en rapport avec le vocabulaire commun des sites Internet, la plupart des sites adoptant le même langage pour la navigation, le menu, les copyrights, etc. L'analyse avec Rapidminer est donc biaisée par tout ce bruit. Peu importe la méthode d'apprentissage employée, le logiciel n'arrive pas à créer un modèle capable de reclasser correctement les sites Internet de notre base d'apprentissage.

3.6.1.3 Classification par mots clés

Nous avons donc testé une autre méthodologie pour effectuer notre classification en ne prenant en compte que des mots clés, présents dans le Tableau 3.3, reliés à la nanotechnologie et capables de nous renseigner sur le domaine de la nanotechnologie du site issu des travaux effectués par Islam et Miyazaki (2010) et Porter et al. (2008).

Le regroupement est effectué en calculant la similarité entre les vecteurs mots de chaque site. Nous employons la similarité cosinus, qui utilise le cosinus de l'angle formé entre deux vecteurs de mots pour mesurer la divergence entre deux vecteurs, car mieux adapté à la fouille de texte (Singhal 2001).

Tableau 3.3 : Mots clés selon le domaine de la nanotechnologie

Domaine	Mots-clés
Bio nanotechnologie	bionanotechnology, nanobiotechnology, biological nanosensor, nanobiomagnetic, nanocantelevers, targeted nano-therapeutic, nanoreplication, nanoencapsulation, nanocontainer, nanomedicine, nanocomplex, biosensor, molecular motor, biomolecular fabric, engineered enzyme/protein, nano-biocomposite
Métrologie et nano processus	Scanning tunneling microscopy, atomic force microscopy, scanning probe microscopy, nanoprototyping, nanofabrication, nanolithography, nanomachining, nanomanufacturing, electron-beam lithography, nanimprint lithography, self-assembly, directed assembly, nanomechanic, AFM, focus ion beam, FIB
Nanoélectronique	Nanoelectronic, nanodevice, nanosensor, nanocrystal memory, molecular electronic, nanorobotic, nems, nanoswitch, nanocomputing, nanotransisator, nanoscale magnetic
Nanomatériaux	Nanomaterial, fullerene, nanocomposite, nanofilm, nanotube, quantum tube, nanoparticle, nanostructure, nanopowder, nanocoating, sol gel, graphite, growth methode, graphene, quantum dot

Cependant cette méthode de classification non supervisée n’est pas capable d’effectuer les regroupements selon les domaines de la nanotechnologie au sein des 487 sites. Le logiciel n’arrive pas à regrouper le grand nombre de sites dans des classes différentes.

3.6.1.4 Classification par mots clés avec base d'apprentissage

Finalement, nous testons une méthode hybride entre les deux méthodes précédentes. Nous recherchons les mots clés de chaque domaine dans le texte des sites de la base d'apprentissage en leur assignant une nouvelle étiquette et nous effectuons un regroupement à partir de ces mots en donnant au logiciel le nombre de classes à effectuer.

Le logiciel Rapidminer effectue les regroupements entre les sites de notre base d'apprentissage à partir de ces mots et nous retrouvons les sous-domaines de la nanotechnologie de la littérature dans les regroupements. La Figure 3.5 correspond au regroupement des sites en fonction des mots clés correspondant aux domaines de la nanotechnologie. Comme nous l'avons vu plus haut avec le diagramme Venn, les domaines se chevauchent. Ce chevauchement est aussi visible sur la Figure 3.5. Les grappes 1 à 4 correspondent chacune à un domaine de la nanotechnologie tandis que la grappe 5, non visible sur la figure, correspond aux sites n'ayant pas pu être classés dans un de ces quatre domaines.

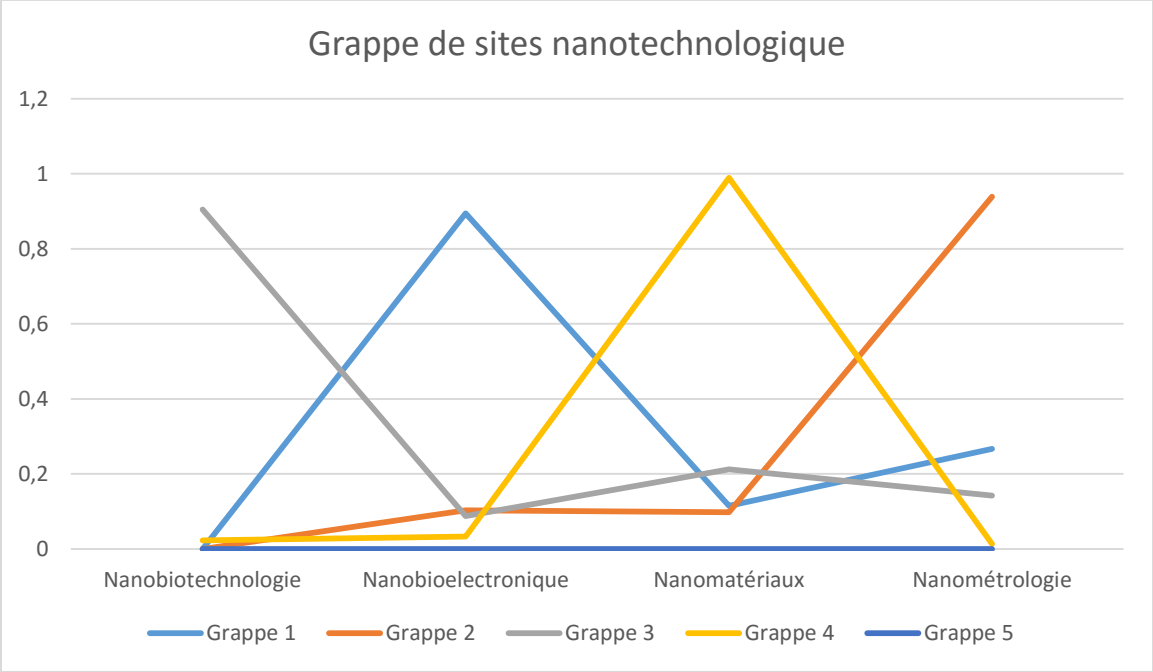

Figure 3.5 : Regroupement des sites Internet par domaine

Avec cette nouvelle méthodologie, les résultats obtenus par l'apprentissage sont plus pertinents. Nous évaluons l'apprentissage grâce à une matrice de confusion, Tableau 3.4. La matrice de confusion est construite à partir de la base d'apprentissage en testant le modèle pour confronter la valeur de la prédiction avec la valeur réelle.

La matrice montre qu'il y a eu une confusion entre deux classes. Un site a été classé dans une mauvaise classe par le modèle, avec une confusion entre le domaine nanoélectronique et nanoprocessus (cluster_0 et cluster_1). Le niveau de fiabilité, correspondant à la proportion du nombre total de prédictions correctement effectuées, est de 98,08%. De plus, l'indice Kappa, permettant d'évaluer le degré de concordance dans la manière de classer, est de 0,973 pour notre matrice de confusion. Landis and Koch (1977) interprètent un coefficient de Kappa compris entre 0.81 et 1 comme étant un accord parfait.

Tableau 3.4 : Matrice de confusion

	Vrai. Nanobiotechnolog.	**Vrai. Nanoélectronique**	**Vrai. Nanomatériaux**	**Vrai. Nanométrologie**	**Vrai. Autre**	**Précision de la classe**
Pred. Nanobiotechnologie	8	0	0	0	0	100%
Pred. Nanoélectronique	0	6	0	1	0	85.71%
Pred. Nanomatériaux	0	0	21	0	0	100%
Pred. Nanométrologie	0	0	0	5	0	100%
Pred. Autre	0	0	0	0	11	100%
Rappel de la classe	100%	100%	100%	83.33%	100%	

Notre modèle de classification est donc capable de reclasser les sites Internet de notre base d'apprentissage précisément. Pour classer le reste des sites Internet, nous effectuons la même recherche de mots clés avant d'appliquer le modèle.

3.7 Regroupement géographique

Pour effectuer les regroupements géographiques des entreprises, nous calculons la distance de l'emplacement géographique des entreprises avec les grappes en nanotechnologie identifiées par Schiffauerova et Beaudry (2009).

En pratique l'emplacement géographique des entreprises est déterminé grâce aux codes postaux présents dans les sites des entreprises. Pour identifier les codes postaux nous recherchons l'expression régulière suivant dans le texte : \b[aA-zZ][0-9][aA-zZ]\ ?[0-9][aA-zZ][0-9]\b.

Les regroupements d'entreprises correspondent à toutes les entreprises distantes de moins de 100 km d'une grappe en nanotechnologie, correspondant aux villes de Toronto, Montréal, Ottawa, Vancouver, Edmonton, Québec, Kingston et Calgary.

3.8 Indicateur d'innovation et de commercialisation dans les sites Internet

Nous cherchons à travers les sites Internet des entreprises en nanotechnologie des indicateurs d'innovation et de commercialisation. La revue de littérature nous donne plusieurs indicateurs permettant d'évaluer l'innovation et la commercialisation.

3.8.1 Création des indicateurs

Pour classer les entreprises via leurs sites Internet nous avons mis au point une liste d'indicateurs inspirés par la littérature sur l'innovation et la commercialisation des technologies de pointe.

Comme nous avons pu le voir dans la revue de littérature, plusieurs indicateurs pour l'innovation et la commercialisation existent déjà. Cependant, tous ne sont pas nécessairement adaptés à une étude par analyse de mots clés sur Internet. Nous avons donc sélectionné six indicateurs portant sur l'innovation et la commercialisation.

Nous allons appeler indicateur dans la suite de ce mémoire l'ensemble des mots clés regroupés par concepts. Les concepts sont les suivants.

- Recherche et développement;
- Propriété intellectuelle;
- Nombre de produits;
- Collaboration;
- Financement;
- Taille de l'entreprise.

Pour chaque indicateur d'innovation employé pour construire la matrice, une liste de mots clés a été établie. Tous ces mots clés sont ensuite cherchés dans le texte des sites Internet publics des entreprises, récupérés au préalable par notre moissonneur. Le nombre total de mots par site ainsi que le nombre de pages récoltées sont aussi comptabilisés pour effectuer une normalisation des indicateurs.

3.8.1.1 Recherche et développement

Avec notre indicateur « recherche et développement », nous cherchons comme son nom l'indique des mots relatifs à la recherche et au développement. Comme nous l'avons vu dans la recension de la littérature, la R-D est un intrant de l'innovation. Nous cherchons donc à déterminer avec cet indicateur si l'entreprise effectue des activités de R-D. Nous cherchons dans un premier temps à savoir si l'entreprise effectue une activité de R-D. Dans un second temps nous cherchons à savoir l'orientation de la R-D de l'entreprise en effectuant une dichotomie entre la recherche et le développement afin de distinguer la recherche orientée recherche fondamentale et recherche appliquée.

Pour cet indicateur nous nous basons sur les travaux effectués par Gök et al. (2014) qui ont développé une stratégie de recherche pour les activités de R-D.

Nous recherchons l'ensemble des mots suivants, permettant de mettre en avant les entreprises ayant une activité de R-D :

> *research and development, r&d, laboratories, researcher, scientist, product development, technology development, development phase, technical development, development program, development process, development project, development cent, development*

facility, technological development, development effort, development cycle, development research, research & development, development activity, fundamental research, basic research (Gök et al. 2014)

Pour préciser le type d'activité de recherche, nous effectuons une distinction entre la recherche fondamentale et la recherche appliquée. Cette dichotomie dans les activités de recherche est effectuée en séparant la liste de mots dans ces deux catégories.

- Recherche orientée fondamentale :

  Laboratories, scientist, fundamental research, basic research;

- Recherche orientée appliquée :

  product development, technology development, development phase, technical development, development program, development process, development project, development center, development facility, technological development, development effort, development cycle, development research, development activity.

3.8.1.2 Propriété intellectuelle

Nous savons déjà, grâce à la recension de littérature, que la propriété intellectuelle et à plus forte raison les brevets sont des extrants de la recherche procurant à l'entreprise un avantage compétitif sur la concurrence en procurant l'exclusivité des produits de la recherche pour des fins de commercialisation. Pour une entreprise, faire breveter une technologie implique qu'elle espère un retour sur investissement en commercialisant la technologie, en vendant des licences ou en revendant le brevet. Dans tous les cas, pour une entreprise le fait de mentionner une technologie brevetée sur son site Internet est donc un bon indicateur d'innovation et de commercialisation. Nous avons aussi vu que les PME emploient le secret commercial comme protection de leur propriété intellectuelle pour des raisons économiques.

Nous cherchons donc pour cet indicateur les mots clés suivants :

- *Patent*, intellectual property, trade secret, industrial design*

3.8.1.3 Produits

Nous cherchons à connaître le nombre de produits différents proposés par chaque site pour évaluer la commercialisation de la nanotechnologie par domaine de la nanotechnologie. Les sites Internet servent à promouvoir et à vendre les produits de l'entreprise et sont donc de bonnes sources de données pour connaître le nombre de produits d'une entreprise.

Pour connaître le nombre de produits différents existant sur chaque site Internet nous pouvons lire la totalité des sites et compter le nombre de produits proposés manuellement. Cependant cette méthode est très couteuse en temps, car il faudrait lire plusieurs milliers de pages. Nous recherchons donc les produits identifiés par les symboles de marque « ™ » et de marque déposée « ® ». Nous devons identifier correctement le nom de chaque produit pour les comptabiliser, car les noms de produit sont souvent répétés plusieurs fois dans les sites et la recherche des symboles n'est pas suffisante.

L'identification des noms de marque se fait au moyen d'une expression régulière. Une expression régulière est un une chaine de caractère décrivant un motif respectant une certaine syntaxe. Ainsi l'expression régulière suivante : \b(((([A-Z-a-z-]{1,}))(\w{1,})\ ?[™®]) nous permet de capturer toutes les chaines de caractères commençant par un espace suivi d'un des deux symboles de marque.

Cette méthode nous limite donc aux produits ayant été identifiés au préalable par l'entreprise avec un des symboles.

3.8.1.4 Collaboration

Conformément à ce qui a été dit dans la recension de la littérature, la collaboration est un facteur clé de l'innovation. Avec l'indicateur de collaboration, nous cherchons donc à savoir si l'entreprise a des partenaires. D'après la recension de la littérature, une entreprise et *a fortiori* une jeune entreprise a plus de chances de développer une innovation si elle est en relation avec une université profitant de sa recherche fondamentale, de ses installations couteuses et de son personnel qualifié.

Nous recherchons les entreprises ayant des activités de collaboration en récupérant tous les mots en rapport avec la collaboration, issus des travaux de Ramdani et al. (2014) *(affiliation, collaboration, cooperation, partners, partnership).*

3.8.1.5 Financement

L'indicateur de financement nous aide à connaître la stratégie de financement de l'entreprise. Nous cherchons savoir si l'entreprise a recours à des investisseurs externes de type capitaux risque ou des subventions publiques. Les entreprises de type « jeune pousse » ou « entreprise essaimée » font généralement plus appel aux investisseurs externes de type capitaux de risque que les grandes entreprises générant déjà des bénéfices (Gök et al. 2014). Nous recherchons donc des sources de financement privé de types investisseurs en capital-risque, des sources publiques en cherchant les programmes de financement de la nanotechnologie d'Industrie Canada et les subventions. Grâce aux programmes de financement de la nanotechnologie proposé par Industrie Canada, nous avons identifié plusieurs bourses et programmes de financement que nous recherchons dans les sites.

Nous cherchons donc les mots clés suivants :

- *atlantic canada opportunities agency, business development bank of canada, sustainaible development technology, venture capital, atlantic innovation fund,nrc-irap, fednor, industrial research assistance program, grants, privat invest*.*

3.8.1.6 Taille de l'entreprise

La taille de l'entreprise étant un facteur important pour l'entrée sur le marché de produits en nanotechnologie nous évaluons la taille des entreprises suivant le nombre de pages récupéré. En effet Youtie et al.(2012) notent une relation entre la taille des entreprises, son nombre d'employés, et la taille des sites Internet, le nombre de pages. Comme nous l'avons vu, la taille des sites Internet varie énormément. En effet une grande entreprise a plus de chance de proposer un nombre de produits plus important, ainsi que davantage d'information supplémentaire. Un site Internet fourni est donc un facteur favorisant la commercialisation. Notre indicateur de taille d'entreprise correspond donc au nombre de pages par site.

3.8.2 Méthodologie de récupération des mots clés

Pour récupérer nos indicateurs, nous devons à la fois chercher des mots simples et des couples de mots ou des symboles. Nous testons différentes manières de récupérer les mots clés avec des avantages et inconvénients pour chaque méthode.

3.8.2.1 Recherche par mots simples

Dans un premier temps nous cherchons tous les mots simples de chaque indicateur après avoir effectué le traitement du texte. Le traitement du texte est semblable à celui effectué pour la classification automatique des documents :

- Récupération des données;
- Transformation du texte en minuscule;
- Séparation des mots en unité lexicale;
- Filtrage de la longueur des unités lexicales pour n'utiliser que les unités lexicales entre 4 et 25 caractères;
- Filtrage les mots d'arrêt qui ralentissent les temps de traitement sans être utile à l'étude.

Ce traitement nous renvoie une matrice de tous les mots présents dans les sites ainsi que leurs nombres d'occurrences. L'avantage de cette méthode de récupération est sa rapidité et sa simplicité.

Cependant l'observation des résultats obtenus, nous a amené à changer de méthode pour enlever les limitations liées au traitement du texte. Avec la tokenisation, la séparation du texte en unité lexicale, nécessaire à la construction de la matrice, la recherche de mots clés composés est impossible. Avec cette méthode de comptage, par exemple chercher le mot « recherche » sans contexte renvoie un grand nombre de faux positifs pour l'indicateur de recherche et développement. De plus, les caractères spéciaux tels que &, ®, ™ utilisés dans plusieurs indicateurs, i. e. « R&D », sont aussi retirés par le traitement de texte.

Ces limitations nous ont obligés à changer notre méthode de récupération pour obtenir des résultats plus cohérents par rapport au contexte dans lequel ces mots sont employés.

3.8.2.2 Recherche par mots composés

Afin de pouvoir capturer des mots clés composés ainsi que des caractères spéciaux, nous avons recherché les indicateurs dans le texte brut de chaque page Internet. Pour cela nous avons créé des expressions rationnelles, ou expressions régulières, pour chaque indicateur.

Grâce à un ensemble de caractères, nous pouvons construire des expressions régulières correspondant à plusieurs mots clés. La chaine de caractères « ([Pp]atent)\w* » correspond à tous les mots tels que « Patent, Patents, Patented, patent,…. ». Les expressions régulières nous permettent aussi de rechercher des mots composés et des couples de mots présents dans un intervalle donné.

Cette méthode de récupération des indicateurs est plus robuste que la première, mais nécessite de devoir écrire une expression régulière pour tous les indicateurs recherchés. Nous favorisons donc cette méthode de recherche pour préciser nos indicateurs. La stratégie de recherche avec les expressions régulières de chaque indicateur est donnée en Annexe 6.

3.8.3 Construction des indicateurs

Après avoir récupéré nos mots clés, nous pouvons construire nos indicateurs d'innovation et de commercialisation. Les indicateurs sont donc construits à partir de plusieurs variables, correspondants au nombre d'occurrence des mots clés par site Internet.

Chaque indicateur est construit en sommant les variables qui le composent.

Afin de comparer les indicateurs entre chaque site, nous devons donc les normaliser. Pour prendre en compte la différence de taille existant entre les sites, nous capturons donc le nombre de pages ainsi que le nombre total de mots récoltés par page. Une fois les indicateurs construits nous effectuons une pondération par le nombre total de mots par site et une normalisation statistique afin de pouvoir comparer les sites entre eux.

3.8.4 Statistique descriptive des variables par indicateurs

3.8.4.1 Financement

Notre indicateur de financement, Tableau 3.5, nous renseigne sur le financement de l'entreprise. Dans notre étude, 55% des entreprises font mention d'une source de financement sur leur site dont une grande majorité de subventions. Les variables concernant les programmes de financement des nanotechnologies proposé par Industrie Canada n'ont pas apporté beaucoup de résultats pour cet indicateur. Soit les entreprises n'en ont pas profité soit elles ne le mentionnent pas sur leur site. Dans ces deux cas, nous ne pouvons pas faire de distinction précise dans les sources de financement.

Tableau 3.5 : Description statistique des variables pour l'indicateur de financement

	Occurrence	Moyenne	Variance	Écart type	Min	Max
busines development bank of canada	2	0,02	0,024	0,15	0	1
grant	231	2,38	45,259	6,72	0	48
industrial research assistance program	14	0,14	0,3956	0,62	0	4
private investment	3	0,03	0,03028	0,17	0	1
private investor	1	0,01	0,01	0,1	0	1
public fund	1	0,01	0,0103	0,1	0	1
venture capital	32	0,33	1,3275	1,15	0	7

3.8.4.2 R-D

L'indicateur de R-D, Tableau 3.6, nous permet d'évaluer les activités de R-D effectuées par les entreprises par domaine. Nous avons trouvé que 90 % des entreprises annoncent effectuer des activités de recherche et de développement, dont 71 % de la recherche fondamentale et 86 % de la recherche appliquée et du développement. D'après les résultats de l'enquête par questionnaire menée en 2005 par McNiven (2007), 91 % des 88 entreprises étudiées déclarant des activités de nanotechnologie effectuent de la R-D.

Tableau 3.6 : Description statistique des variables pour l'indicateur de R-D

	Occurrence	Moyenne	Écart type	Min	Max
basic research	27	0,28	1,55	0	15
development center	1	0,01	0,1	0	1
development cycle	6	0,06	0,34	0	3
development effort	16	0,16	0,51	0	3
development facility	1	0,01	0,1	0	1
development phase	4	0,04	0,19	0	1
development process	1	0,01	0,1	0	1
development program	17	0,18	0,62	0	4
development project	36	0,37	1,3	0	8
fundamental research	5	0,05	0,22	0	1
laboratories	311	3,21	17,07	0	167
laboratory	498	5,13	10,81	0	67
product development	650	6,70	30,86	0	255
r&d	466	4,80	8,48	0	53
research & development	83	0,86	2,88	0	25

Tableau 3.7 (suite et fin): Description statistique des variables pour l'indicateur de R-D

research and development	330	3,40	10,65	0	97
researcher	307	3,16	6,9	0	42
scientist	427	4,40	12,04	0	98
technical development	6	0,06	0,37	0	3
technological development	4	0,04	0,19	0	1
technology development	47	0,48	1,11	0	6

3.8.4.3 Propriété intellectuelle

Selon notre indicateur sur la propriété intellectuelle, Tableau 3.8, 68 % des entreprises font mention de propriété intellectuelle dans leur site, dont 60 % uniquement pour les brevets. Comme nous l'avons vu dans la littérature, les brevets restent la protection intellectuelle la plus commune dans l'industrie. Notre variable concernant le secret industriel n'obtient pas de bon résultat. Cependant il peut être absurde de mentionner un secret industriel sur un site Internet.

Tableau 3.8 : Description statistique des variables pour l'indicateur de protection intellectuelle

	Occurrence	**Moyenne**	**Écart type**	**Min**	**Max**
intellectual property	209	2,15	5,67	0	36
patent	844	8,70	23,31	0	182
patentability	4	0,04	0,4	0	4
patentable	9	0,09	0,52	0	4
patenting	7	0,07	0,52	0	5
patentlicense	2	0,02	0,2	0	2
patentpending	17	0,18	1,02	0	7
patentprotect	2	0,02	0,2	0	2
trade secret	13	0,13	0,6	0	4

3.8.4.4 Collaboration

Avec notre indicateur de collaboration, Tableau 3.9, nous pouvons constater que le pourcentage d'entreprises faisant mention de collaboration, alliance ou de partenariat sur leur site Internet est de 83,5 %.

La collaboration, universitaire ou autre permet aux entreprises de partager les connaissances, les infrastructures et les réseaux de distribution.

Tableau 3.9 : Description statistique des variables pour l'indicateur de collaboration

	Occurrence	Moyenne	Écart type	Min	Max
alliance	117	1,21	6,54	0	63
collaborating	20	0,21	0,53	0	3
collaboration	244	2,52	4,64	0	21
collaborative	96	0,99	2,04	0	8
collaboratively	7	0,07	0,5	0	4
cooperating	5	0,05	0,22	0	1
cooperation	32	0,33	0,98	0	6
cooperative	5	0,05	0,26	0	2
cooperatively	2	0,02	0,2	0	2
partner	965	9,95	21,15	0	121
partnering	59	0,61	2,2	0	20
partnership	336	3,46	7,83	0	59

3.8.4.5 Produits

L'indicateur de produit, Tableau 3.10, correspond au nombre de produits que nous avons pu identifier sur les sites des entreprises ayant des activités en nanotechnologie. Nous trouvons au total 504 produits différents soit un peu plus de 5 produits par site. Sur certains sites nous n'avons pas été capables d'identifier un seul produit tandis que d'autres sites font présence de 26 produits différents. Cet indicateur bien qu'essentielle dans la suite de notre étude est cependant faussé par les marques déposées présentes sur les sites internet, mais pas développées ou commercialisées par l'entreprise. Par exemple un site fait mention de la technologie « Bluetooth® » qui est comptabilisée dans notre indicateur. Il pourrait être possible de chercher dans les registres de marque déposée l'appartenance de chaque marque en fonction des entreprises. Cependant il faudrait s'appuyer sur des données externes aux sites Internet. Nous choisissons de garder ces faux positifs pour le reste de notre étude.

Tableau 3.10 : Description statistique des variables pour l'indicateur de produit

	Occurrence	Moyenne	Écart type	Min	Max
Produits	504	5,19	6,03	0	26

3.8.4.6 Pages

Le nombre de pages, Tableau 3.11, nous sert d'indicateur pour évaluer la taille des entreprises. Selon Youtie et al. (2012) une grande entreprise aura tendance à avoir un site Internet comportant un plus grand nombre de pages, les PMEs ayant des sites plus modestes en nombre de pages.

Si la moyenne des pages par site identifié comme ayant des activités nanotechnologiques est de 33 pages par site, certains sites comportent plus d'une centaine de pages.

Tableau 3.11 : Description statistique des variables pour l'indicateur de page

	Occurrence	**Moyenne**	**Écart type**	**Min**	**Max**
Nombre de pages	3261	33,61856	19,48	1	105

3.8.5 Statistique descriptive des indicateurs

Les indicateurs sont construits en sommant les variables qui les composent. Une fois les indicateurs construits, nous les pondérons par le nombre total de mots par site avant de les normaliser. Le nombre d'occurrences des indicateurs varie beaucoup. Notre indicateur de R-D apparait le plus de fois. Cela est dû à un grand nombre de variables le composant, mais aussi à la volonté des entreprises de faire apparaitre leurs activités de recherche et de développement sur leur site Internet. Notre indicateur de financement en revanche possède peu d'occurrences relativement aux autres. Les entreprises ne trouvent peut-être pas pertinent de mentionner cette information sur leur site ou notre indicateur ne possède pas assez de variables. le nombre d'occurrences des indicateurs varie beaucoup. Notre indicateur de R-D apparait le plus de fois. Cela est dû à un grand nombre de variables le composant, mais aussi à la volonté des entreprises de faire apparaitre leurs activités de recherche et de développement sur leur site Internet. Notre indicateur de financement en revanche possède peu d'occurrences relativement aux autres. Les entreprises ne trouvent peut-être pas pertinent de mentionner cette information sur leur site ou notre indicateur ne possède pas assez de variables.

Tableau 3.12 nous donne le nombre d'occurrence des indicateurs, correspondant à la somme des variables composant l'indicateur ainsi que la moyenne et l'écart type pour chaque indicateur une fois normalisées.

Les occurrences des indicateurs étant faibles par rapport au nombre de mots par site, les moyennes pondérées par indicateur sont très faibles. Le logiciel Rapidminer prend en compte jusqu'à 20 décimales ne faussant pas les résultats par un tronquage non désiré des indicateurs.

Nous pouvons constater que le nombre d'occurrences des indicateurs varie beaucoup. Notre indicateur de R-D apparait le plus de fois. Cela est dû à un grand nombre de variables le composant, mais aussi à la volonté des entreprises de faire apparaitre leurs activités de recherche et de développement sur leur site Internet. Notre indicateur de financement en revanche possède peu d'occurrences relativement aux autres. Les entreprises ne trouvent peut-être pas pertinent de mentionner cette information sur leur site ou notre indicateur ne possède pas assez de variables.

Tableau 3.12 : Description statistique des indicateurs

	Occurrence	**Moyenne**	**Écart type**
Financement	284	2,16 x10^{-4}	1,24
R-D	3243	3,62 x10^{-3}	1,41
-R-D fondamentale	1575	1,33 x10^{-3}	2,52
-R-D appliquée	1673	2,28 x10^{-3}	1,05
Propriété intellectuelle	1107	8,42 x10^{-4}	1,35
Collaboration	1888	1,85 x10^{-3}	2,24
Produits	504	5,5 x10^{-4}	2,42
Pages	3261	33,61	19,47
Mots	1226150	12640	1,40

3.8.6 Stockage des indicateurs

Pour regrouper toutes les occurrences des indicateurs que nous avons construits au moyen de Rapidminer, nous avons créé une table Mysql par indicateur avec un identifiant unique par site Internet servant de clé primaire pour toutes les tables d'indicateurs. Nous rajoutons, en plus des indicateurs construits à partir des mots clés, les regroupements par domaine de la nanotechnologie, les regroupements géographiques ainsi que l'URL des sites Internet.

Une nouvelle table Mysql est créée en joignant toutes les tables d'indicateurs avec la même clé primaire. Cette dernière table correspond à la matrice d'indicateurs que nous employons pour effectuer le reste des analyses en l'exportant vers le logiciel d'analyse statistique SPSS.

3.8.7 Construction des regroupements suivant les indicateurs

À partir de nos indicateurs, nous effectuons des regroupements de sites Internet pour analyser les facteurs influant sur la commercialisation des nanotechnologies. Chaque site internet est représenté par les 6 indicateurs construits à partir des mots clé.

L'analyse de regroupement est effectuée avec le logiciel d'analyse statistique SPSS.

Dans un premier temps nous désirons connaitre le nombre optimal de grappe par regroupement. Pour cela nous effectuons une analyse TwoStep. L'analyse TwoStep forme des regroupements avec différent nombre de grappe et calcule pour chaque nombre de grappe différent le critère d'information bayésien (BIC).

Le meilleur regroupement est le regroupement qui minimise le critère BIC. Le Tableau 3.13 donne le critère BIC pour les regroupements de 2 à 10 grappes. Dans le tableau, les regroupements à 2 et 3 grappes ont un BIC très proche l'un de l'autre avec une modification de 1.9 entre les deux. Nous choisissons de pour la suite de l'étude d'effectuer les regroupements avec 2 et 3 grappes.

Tableau 3.13 : Critère BIC en fonction du nombre de grappe

Nombre de clusters	**Critère bayésien de Schwartz (BIC)**	**Modification BIC**	**Rapport des modifications BIC**	**Rapport des mesures de distance**
1	379,417			
2	**340,802**	**-38,615**	**1,000**	**1,924**
3	**342,697**	**1,896**	**-,049**	**1,655**
4	361,951	19,254	-,499	1,102
5	383,652	21,700	-,562	1,255
6	410,233	26,581	-,688	1,398
7	442,266	32,034	-,830	1,282
8	477,318	35,052	-,908	1,287
9	514,753	37,435	-,969	1,008
10	552,258	37,505	-,971	1,007

Dans un deuxième temps, ayant identifié le nombre de grappe nous formons des regroupements hiérarchique en imposant le nombre de grappe par regroupement et en faisant varier la mesure et la méthode d'optimisation. Nous employons comme mesure Seuclid, Chebychev, Block et Cosine et comme méthode d'agrégation Ward, Complete, Centroïde Median et Baverage.

Pour choisir la meilleure méthode de regroupement nous cherchons le nombre d'individu par grappe en combinant les méthodes de mesure et d'agrégation. Le Tableau 3.14 donne la répartition des entreprises dans les grappes en fonction de la méthode d'agrégation et de la mesure. Un regroupement est considéré comme intéressant si les grappes ont un nombre d'individu équilibré et facilement discriminable.

Tableau 3.14 : Répartition des individus dans les grappes suivant la méthode employée

Méthode d'agrégation	**Méthode de mesure**	**Nombre individu grappe 1**	**Nombre individu grappe 2**	**Nombre individu grappe 3**
Ward	Seuclid	35	45	17
Ward	Block	17	65	15
Ward	Chebychev	35	26	36
Ward	Cosine	44	28	15
Complete	Cosine	30	42	25
Centroid	Cosine	46	14	37
Median	Cosine	15	71	11
Baverage	Cosine	35	45	17

Dans notre cas le regroupement le plus intéressant est obtenu avec la mesure Cosine et l'optimisation Complete car les grappes sont équilibrées et que l'information obtenue est pertinente.

Avec cette analyse de regroupement nous pouvons établir que les entreprises canadiennes, dépendamment de leur site Internet, se regroupent de façon optimale en deux et trois grappes différentes. Chacune de ces grappes sera nommée et décrite dans le chapitre suivant.

3.9 Synthèse

Avec la méthodologie que nous avons développée pour notre étude nanotechnologies et grâce à différents outils nous avons récolté et stocké le texte des sites Internet, compté le nombre

d’occurrence de mots clé, retrouvé des informations comme les noms de produit et les adresses postal des entreprises. Nous avons ainsi construit des indicateurs, ici basés sur la commercialisation et l’innovation dans le contexte des à partir de mots clés présent dans le texte des sites des entreprises pour construire des regroupements nous aidant à mieux comprendre le contexte actuelle de la commercialisation des nanotechnologies au Canada.

CHAPITRE 4 RÉSULTATS ET ANALYSE

Dans ce chapitre nous analysons premièrement les résultats de la classification des entreprises par domaine de la nanotechnologie pour dresser un état général des nanotechnologies au Canada. A partir des sites d'entreprises identifiés comme ayant des activités en nanotechnologie, nous faisons des regroupements géographiques. Enfin nous analysons les résultats issus de la récupération de nos mots clés pour construire nos différents indicateurs ainsi que les regroupements des entreprises que nous avons pu classer correctement dans un domaine de la nanotechnologie basé sur nos indicateurs.

4.1 Commercialisation des produits en nanotechnologie par domaine

La classification automatique des entreprises nous permet d'avoir un aperçu de l'industrie des nanotechnologies au Canada en nous donnant des informations sur le nombre d'entreprises dans ce domaine ainsi que sur le type de nanotechnologie le plus rependu au Canada.

Notre méthodologie de classification automatique des sites Internet nous a permis de classer les entreprises recensées pas diverses sources selon la classification de notre revue de la littérature. Nous donnons en Annexe 7 la liste des entreprises par domaine.

La répartition des entreprises en nanotechnologie par domaine, Figure 4.1, est la suivante :

- Métrologie et nanoprocessus : 13.4%
- Nanodispositifs et nanoélectronique : 9.2%
- Nanomédecine et nanobiotechnologie : 13.4 %
- Nanostructures et nanomatériaux : 63.9%

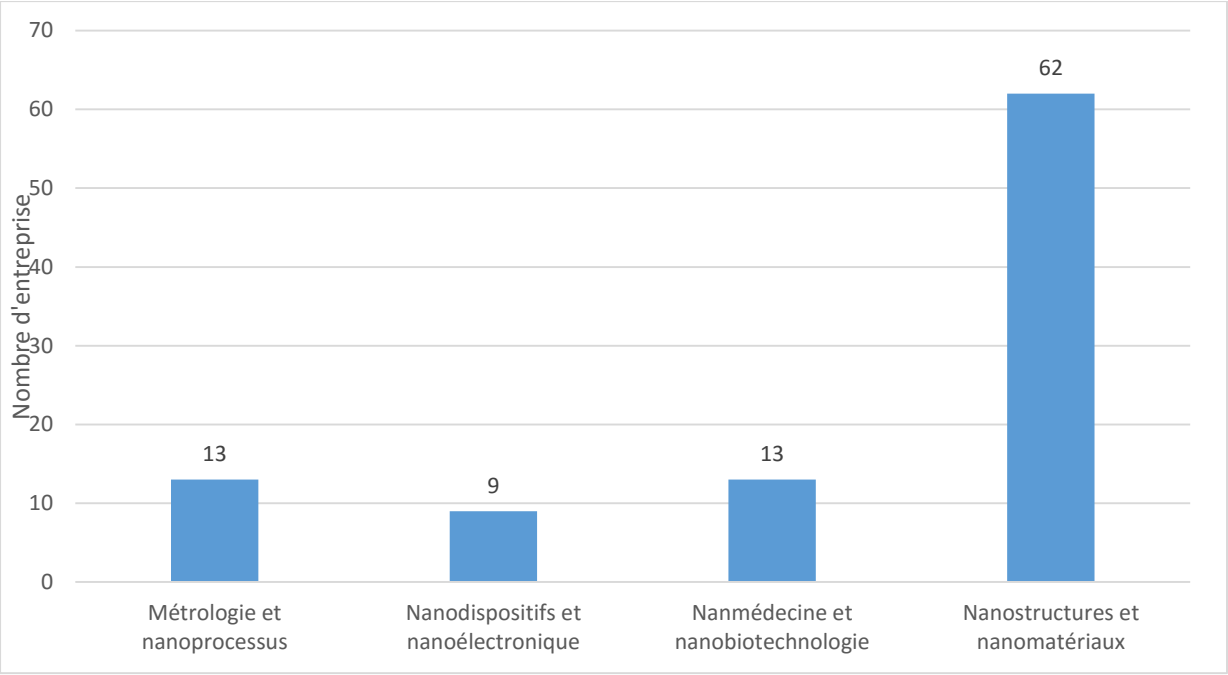

Figure 4.1 : Classification des entreprises

La répartition des entreprises par domaine confirme donc notre première hypothèse selon laquelle le domaine « nanostructures et nanomatériaux » est le mieux représenté dans l'industrie canadienne avec presque 64 % des entreprises en nanotechnologie identifiées.

Néanmoins, nous constatons que le nombre total d'entreprises ayant été classées comme ayant des activités en nanotechnologie, 97, ne correspond qu'à 20% de l'échantillon initial.

Le nombre important de sites inclassables permet de se poser la question quant à notre méthodologie. On peut penser en effet que plus de 97 entreprises utilisent des nanotechnologies dans leurs produits ou procédés au Canada. On rappelle que les sites Internet ont tous été choisis dans des sources recensant les entreprises dans le domaine des nanotechnologies au Canada.

Pour les compagnies n'ayant pas pu être classées dans un domaine de la nanotechnologie, il est intéressant de se demander comment elles ont pu être répertoriées dans un premier lieu dans une base de données d'entreprises en nanotechnologie canadiennes.

Les entreprises peuvent être répertoriées pour employer des nanotechnologies dans leurs produits sans pour autant le mentionner dans leur site pour diverses raisons commerciales voire sans même être au courant de leur utilisation (Maniet 2010). Ces entreprises sont donc à juste titre répertoriées,

mais il nous est impossible avec notre méthodologie de pouvoir affirmer qu'elles font usage de nanotechnologie dans leurs produits.

Une autre possibilité est l'usage d'une mauvaise méthodologie de recensement de la part des sources que nous avons employées. Le site d'Industrie Canada précise par exemple que « les renseignements contenus dans ces répertoires sont directement fournis par les entreprises »[9]. Il est donc probable que des entreprises se soient inscrites à tort ou qu'elles n'aient pas fourni de bonnes informations, faussant les bases de données.

Enfin une autre explication pour ce nombre important de sites inclassables est une erreur dans notre méthodologie. Une des causes peut être que malgré une liste de mots clés aussi exhaustive que possible nous n'avons pas pu trouver de mots reliés à la nanotechnologie dans ces sites Internet.

Cependant avec 809 entreprises au total recensées grâce à diverses sources, nous sommes très au-dessus des recensements des entreprises en nanotechnologie canadiennes. En effet, l'OCDE recense dans un document[10] datant de 2015 sur les indicateurs en nanotechnologie, 41 entreprises actives travaillant dans le domaine des nanotechnologies en 2013. À titre de comparaison, le même document recense 10 341 entreprises en nanotechnologie aux États-Unis en 2014. McNiven (2007) annonce de son côté 88 entreprises canadiennes déclarant travailler avec des nanotechnologies.

Nous pouvons donc supposer que notre classement des entreprises canadiennes est représentatif de l'industrie de la nanotechnologie canadienne actuelle avec 97 entreprises faisant explicitement référence aux nanotechnologies sur leur site Internet.

Pour la suite de notre étude, le sujet d'étude portant sur l'innovation et la commercialisation des nanotechnologies au Canada, nous choisissons de ne pas prendre en compte ces entreprises, car rien ne nous permet d'affirmer leur lien avec la nanotechnologie.

---

[9] https://www.ic.gc.ca/eic/site/aimb-dgami.nsf/fra/03503.html

[10] Source: OECD, Key Nanotechnology Indicators, http://oe.cd/kni, July 2015.

Nous avons cherché à identifier le nombre de produits proposés par les sites identifiés grâce à la classification. L'identification des produits est faite grâce aux symboles des marques de commerce et des marques déposées.

Nous avons comptabilisé les produits ainsi identifiés pour chacun des sites Internet. Les nanostructures et nanomatériaux sont, comme le montre le Tableau 4.1, les domaines proposant le plus grand nombre de produits différents avec plus de 326 produits soit 65 % de produits canadiens en nanotechnologie. Notre hypothèse 1 b) selon laquelle les produits en nanostructure et nanomatériaux sont les plus nombreux est donc vérifiée.

Tableau 4.1 : Produits en nanotechnologie

Domaine	**Nombre total de produits**	**Nombre de produits en %**	**Nombre de produits par entreprise**
Métrologie et nanprocessus	50	10	3.84
Nanodispositifs et nanoélectronique	37	7	4.11
Nanomédecine et nanobiotechnologie	91	18	7
Nanostructures et nanomatériaux	326	65	5.25

Néanmoins, ce nombre est à relativiser par rapport au nombre d'entreprises identifiées dans ce domaine. Proportionnellement au nombre d'entreprises, le domaine proposant le plus grand nombre de produits est le domaine de la nanomédecine et des nanobiotechnologies avec en moyenne 7 produits par entreprise tandis que les nanostructures et nanomatériaux n'en proposent qu'en moyenne 5,25 par entreprise. Le domaine « nanométrologie et nanoprocessus » et le domaine « nanodispositifs et nanoelectronique » n'en proposent respectivement que 3,84 et 4,11 par entreprise.

4.2 Regroupement géographique

Avec les données présentes dans les sites Internet, nous avons identifié les codes postaux des entreprises afin de les localiser.

Les regroupements géographiques des entreprises nous montrent que 83 % des entreprises identifiées par notre classification automatique font partie d'un regroupement d'inventeurs en nanotechnologie.

Tableau 4.2 : Description statistique des regroupements en nanotechnologie

Cluster	Nombre d'entreprises	Entreprises en %	Brevet en % (Schiffauerova and Beaudry 2009)
Toronto	27	28 %	12 %
Montréal	25	26 %	11 %
Ottawa	11	11 %	7 %
Vancouver	10	10 %	7 %
Edmonton	10	10 %	4 %
Calgary	4	4 %	2 %
Québec	3	3 %	2 %
Kinston	1	1 %	1 %
Autre	6	6 %	54 %
Total	97	100 %	100 %

Comme nous pouvons le constater, la répartition des entreprises à l'intérieur des regroupements est assez inégale. Les deux premiers regroupements totalisent près de la moitié (48 %) des entreprises du Canada et seuls 6 % des entreprises ne sont pas localisées au sein d'une grappe. De plus, le nombre d'entreprises est proportionnel au nombre de brevets par grappe.

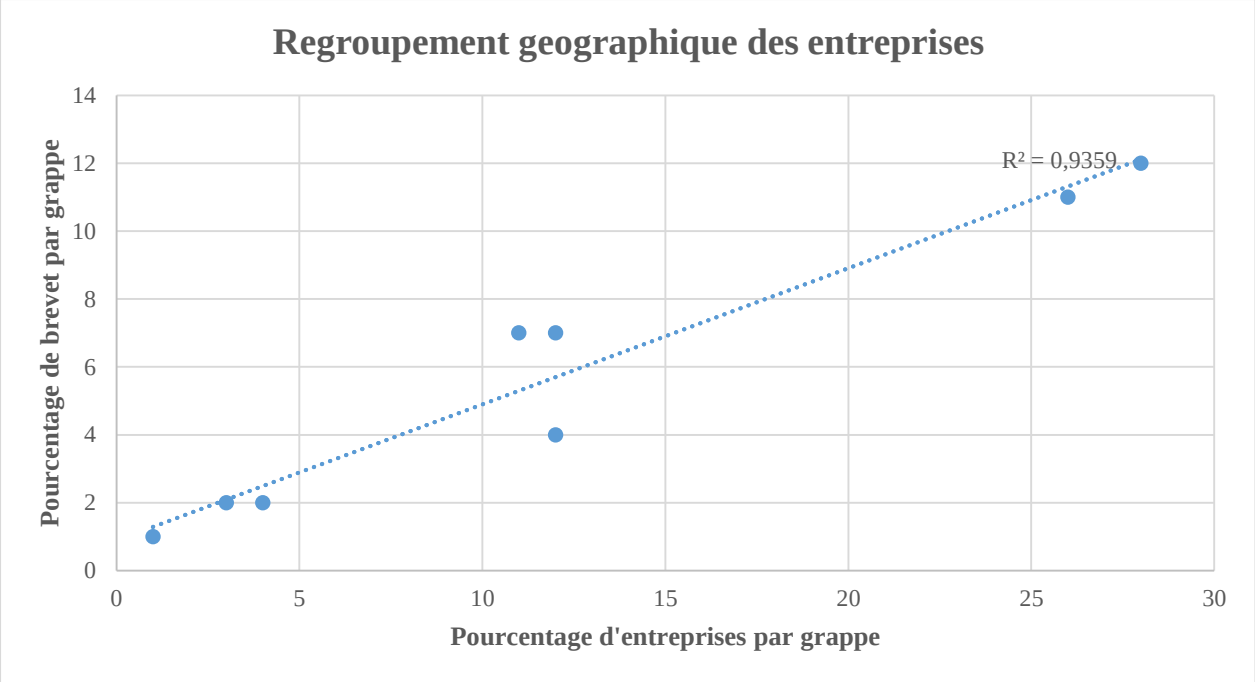

Figure 4.2 : Régression linéaire des regroupements géographiques

Une régression simple linéaire nous donne un coefficient de détermination R^2 de 0,93, indiquant une forte relation entre le pourcentage de brevets et le pourcentage d'entreprises par grappe de nanotechnologie.

Krugman (1991) explique la présence de ces regroupements géographiques d'entreprises par la proximité géographique avec des universités pour profiter des retombées de connaissance, un bassin de travailleurs qualifiés et d'une industrie spécialisée dans les biens intermédiaires. On peut remarquer que les grappes identifiées correspondent en majorité aux villes canadiennes les plus peuplées.

4.3 Innovation et commercialisation

Pour notre analyse de l'innovation et de la commercialisation, nous essayons de dresser un portrait de la nanotechnologie à travers différents aspects tels que les activités de R-D ou la collaboration avec des universités dans les différents domaines de la nanotechnologie. Nous commençons par regarder les corrélations pouvant exister entre nos indicateurs. Enfin nous effectuons des

regroupements des entreprises suivant les indicateurs simples dans un premier temps et les sous indicateurs de R-D et de financement.

4.3.1 Corrélation entre les indicateurs

Dans un premier temps nous effectuons des regroupements d'entreprises suivant les indicateurs de financement, de l'activité de R-D, de la propriété intellectuelle, de la taille de l'entreprise donnée par le nombre de pages, de la collaboration, du nombre de produits, de l'emplacement géographique et du domaine de la nanotechnologie. Le nombre d'occurrences des indicateurs de financement, R-D, PI, collaboration et produit sont pondéré par le nombre de mots par site.

Tableau 4.3 : Matrice de corrélation des indicateurs

	Financement	**R-D**	**PI**	**Page**	**Collaboration**	**Produit**	**Emplacement**	**Domaine**
Financement	1							
R-D	-0,012	1						
PI	0,287	0,093	1					
Page	-0.075	0.123	-0.061	1				
Collaboration	0,212	0,173	0,263	0,023	1			
Produit	0,065	0,133	0,280	0.082	0,091	1		
Emplacement	0,047	-0,113	-0,099	-0,002	-0,004	-0,159	1	
Domaine	0,042	-0,082	-0,057	-0,061	-0,047	0,033	0,253	1

Tableau 4.3 nous renseigne sur les corrélations existant entre nos indicateurs. Pour Rodgers and Nicewander (1988), un coefficient compris entre 0.5 et 1 indique une forte corrélation positive et une faible corrélation entre 0 et 0.5. On remarque premièrement qu'il n'existe aucune forte corrélation dans la matrice. Pour plus de lisibilité, les coefficients supérieurs 0,1 apparaissent en vert dans le tableau.

Les plus fortes corrélations dans notre matrice semblent pertinentes. D'après la matrice de corrélation, la PI est corrélée avec le financement, la collaboration et le nombre de produits. Cela semble indiquer que les entreprises proposant le plus de produits sont aussi celles effectuant le plus d'activité de collaboration, ayant le plus de financement et de PI.

On peut aussi constater la très faible corrélation existant entre la R-D et la PI. Alors que la PI est un extrant de la R-D, on pourrait s'attendre à une corrélation plus forte entre ces deux indicateurs. La propriété intellectuelle est cependant bien corrélée avec le nombre de produits pouvant indiquer qu'une entreprise ayant un nombre élevé de brevets, la principale variable de notre indicateur, propose un plus grand nombre de produits.

Enfin, la corrélation existante entre le domaine de la nanotechnologie et l'emplacement peut correspondre à une spécialisation des grappes en nanotechnologie dans un domaine en particulier.

4.3.2 Regroupement des entreprises par indicateur

À partir des indicateurs nous effectuons des regroupements d'entreprises pour mettre en valeur les caractéristiques communes à plusieurs sites ainsi que l'influence des indicateurs. Comme nous l'avons vu dans notre méthodologie, les regroupements ont été effectués avec la mesure Cosine et la méthode d'agrégation Centroide. Des regroupements sont effectués à la fois avec deux et trois grappes.

4.3.2.1 Regroupement en deux grappes

Le Tableau 4.4 donne les regroupements des sites Internet en deux grappes avec la valeur-p entre les deux grappes. Pour comparer les grappes, nous effectuons le test non paramétrique de Mann-Whitney (M-W). Une valeur-p inférieur à 0.1 nous permet de rejeter l'hypothèse nulle selon laquelle les distributions des deux grappes sont identiques de telle sorte qu'il existe une probabilité de 50% qu'une valeur choisie au hasard dans une population dépasse une observation sélectionnée de façon aléatoire à partir de l'autre population (Mann et Whitney, 1947).

Tableau 4.4 : Regroupement en deux grappes

	Grappe 1	**Grappe 2**	**M-W**
	Moyenne	**Moyenne**	**Valeur-p**
	N_1=55	N_2=42	two-tail
Pages	0,045	-0,059	0,0034
Financement	0,333	-0,435	0,0001
R-D	0,461	-0,603	0,0000
PI	0,324	-0,425	0,0023
Collaboration	0,440	-0,576	0,0000
Nb produits	0,304	-0,398	0,0006

Les valeurs-p entre chaque indicateur des deux grappes sont toutes inférieures à 0.1, ce qui nous permet de dire que les grappes sont bien discriminées.

Les regroupements des sites suivant les indicateurs confirment les hypothèses émises lors de l'interprétation de la matrice de corrélation. La Figure 4.3 correspond à une représentation graphique du regroupement des entreprises en deux grappes. La première grappe en bleu sur la figure contient 55 sites d'entreprises tandis que la grappe en rouge en contient 42.

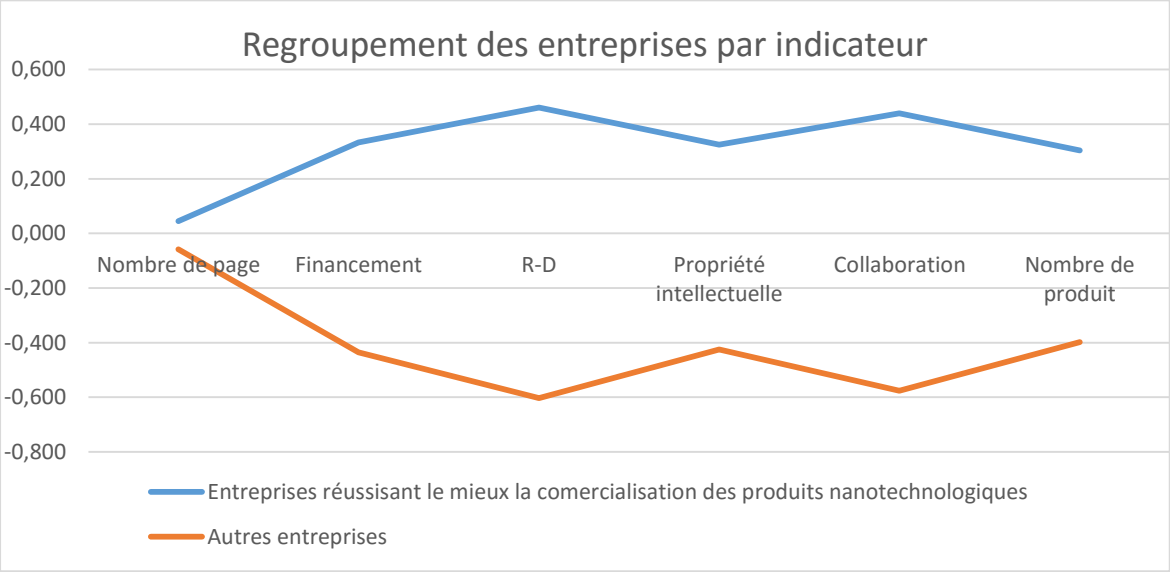

Figure 4.3 : Regroupement des entreprises en deux grappes

Ce regroupement d'entreprise nous permet de faire plusieurs remarques. En lien avec la littérature, les entreprises commercialisant le plus grand nombre de produits correspondent aux entreprises regroupées comme ayant les plus hautes valeurs dans nos indicateurs. Les indicateurs que nous avons choisis ont donc une influence positive sur la commercialisation des nanotechnologies.

L'indicateur de financement montre que les entreprises réussissent à obtenir des financements publique et/ou privé pour les aider à se développer et à créer de nouvelles innovations. Cet indicateur démontre aussi que la politique de financement du gouvernement des entreprises en nanotechnologie semble favorable à leur développement.

L'indicateur de R-D prouve que les entreprises en nanotechnologie effectuent beaucoup de recherche pour produire de nouveaux produits et procédés. De plus, l'indicateur de PI démontre que les entreprises connaissent et utilisent les PI pour protéger les fruits de leur R-D.

L'indicateur de collaboration quant à lui atteste que les entreprises ont des activités de collaboration et ne compte pas uniquement sur leurs ressources internes pour le développement d'innovation.

Le regroupement des entreprises réussissant le mieux la commercialisation des nanotechnologies est aussi le regroupement le plus important en terme d'individu ce qui tend à montrer que la majorité des entreprises canadiennes dans le domaine des nanotechnologies effectuent des activités favorable à la commercialisation des produits en nanotechnologie.

4.3.2.2 Regroupement en trois grappes

Comme nous l'avons vu dans notre analyse de grappe, un regroupement des entreprises en trois grappes est aussi intéressant pour notre étude. Ce regroupement, donné dans le Tableau, regroupe les sites d'entreprises en trois grappes distinctes ayant des tailles comparables. Les valeurs-p sont calculées entre les grappes deux à deux.

Tableau 4.5 : Regroupement en trois grappes

	Grappe1	**Grappe 2**	**Grappe 3**	**M-W**	**M-W**	**M-W**
	Moy	**Moy**	**Moy**	**Valeur-p**	**Valeur-p**	**Valeur-p**
	N_1=30	**N_2=42**	**N_3=25**	**1 vs 2**	**1 vs 3**	**2 vs 3**
Pages	-0,047	-0,059	0,155	0,0867	0,0630	0,0011
Financement	0,919	-0,435	-0,371	0,0000	0,0000	0,7272
R-D	0,049	-0,603	0,955	0,0000	0,0305	0,0000
PI	0,912	-0,425	-0,380	0,0000	0,0005	0,5747
Collaboration	0,334	-0,576	0,567	0,0000	0,7100	0,0000
Nb produits	0,548	-0,398	0,012	0,0002	0,1073	0,0457

Ce regroupement nous donne trois grappes avec respectivement 30, 42 et 25 individus. Par rapport au regroupement précédent, nous retrouvons le même regroupement de 42 individus n'ayant pas d'activité remarquable en nanotechnologie. Les deux autre grappes ont des activités plus ou moins élevé selon les indicateurs et sont de façon générale bien discriminé. Cependant pour les indicateurs de financement et de PI entre les grappes 2 et 3 ainsi que la collaboration entre la grappe 1 et 3 la valeur-p ne nous permet pas de les discriminer correctement.

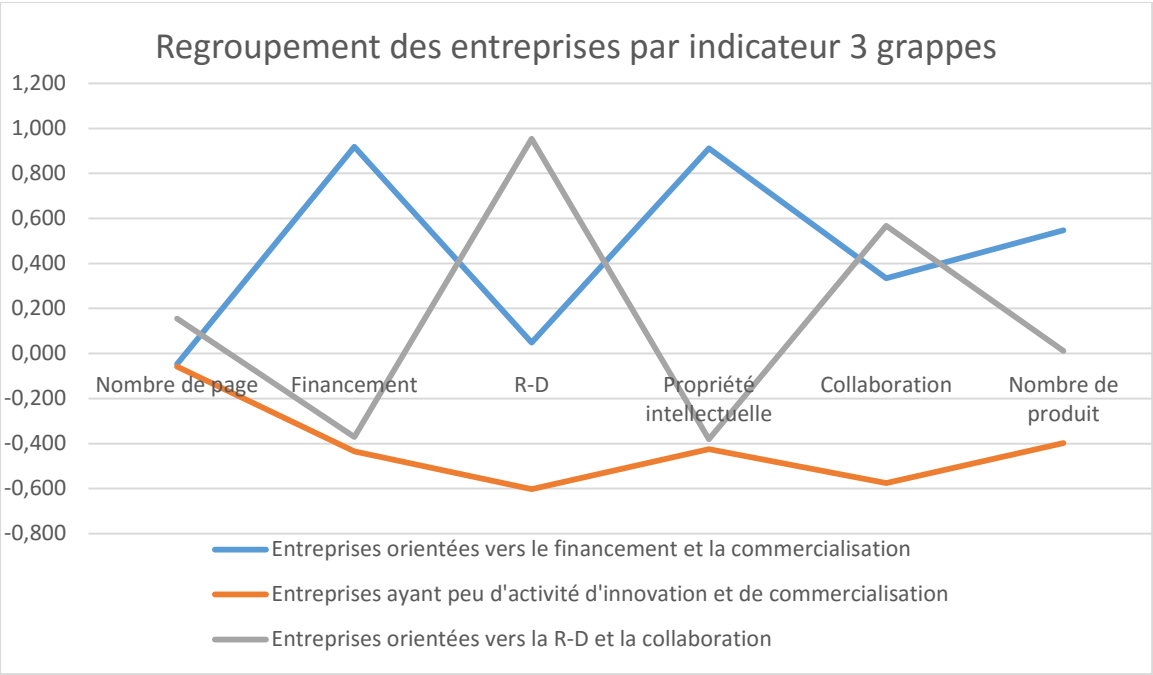

Figure 4.4 : Regroupement des entreprises en trois grappes

La Figure 4.4 nous montre le regroupement des sites d'entreprise en trois grappes. Parmi les deux autres grappes, nous pouvons constater que tous les indicateurs ne semblent pas avoir la même influence sur la commercialisation des produits nanotechnologiques. Dans les deux regroupements, les entreprises proposant le plus de produits sont celles ayant le plus fort indicateur de financement, de R-D, de PI et de collaboration.

Le regroupement en trois grappes nous montre que certain facteur ont plus d'influence sur le nombre de produit proposé sur le site des entreprises. Ainsi les entreprises proposant le plus grand nombre de produits différents semblent plus orientés vers le financement et la propriété intellectuelle.

En accord avec la littérature, les entreprises démontrant la plus forte activité de R-D tendent à collaborer plus. Cela confirme la tendance des entreprises technologique à partager leurs ressources ainsi que leurs connaissances et PI pour le développement d'innovation.

Les entreprises faisant le plus mention de PI correspondent aux entreprises proposant le plus de produit et ayant le plus haut indicateur de financement. Dans la description statistique de l'indicateur de PI, les brevets comptent pour une très grande majorité de la protection de la PI pour

les entreprises en nanotechnologie. Or, nous l'avons dit plus haut un brevet coute cher pour une entreprise, pouvant l'amener à chercher du financement à travers des subventions.

Nous pouvons remarquer cependant que les entreprises ayant d'importantes activités de R-D ne correspondent pas nécessairement aux entreprises ayant le plus de PI. Il peut être donc intéressant de voir l'impact de la recherche appliquée et de la recherche fondamentale sur les autres indicateurs.

4.3.2.3 Influence de la recherche fondamentale

Pour évaluer l'influence du type de recherche, nous regardons comment se comporte les regroupements en effectuant la distinction entre la recherche entre la recherche orientée fondamentale et la recherche appliquée.

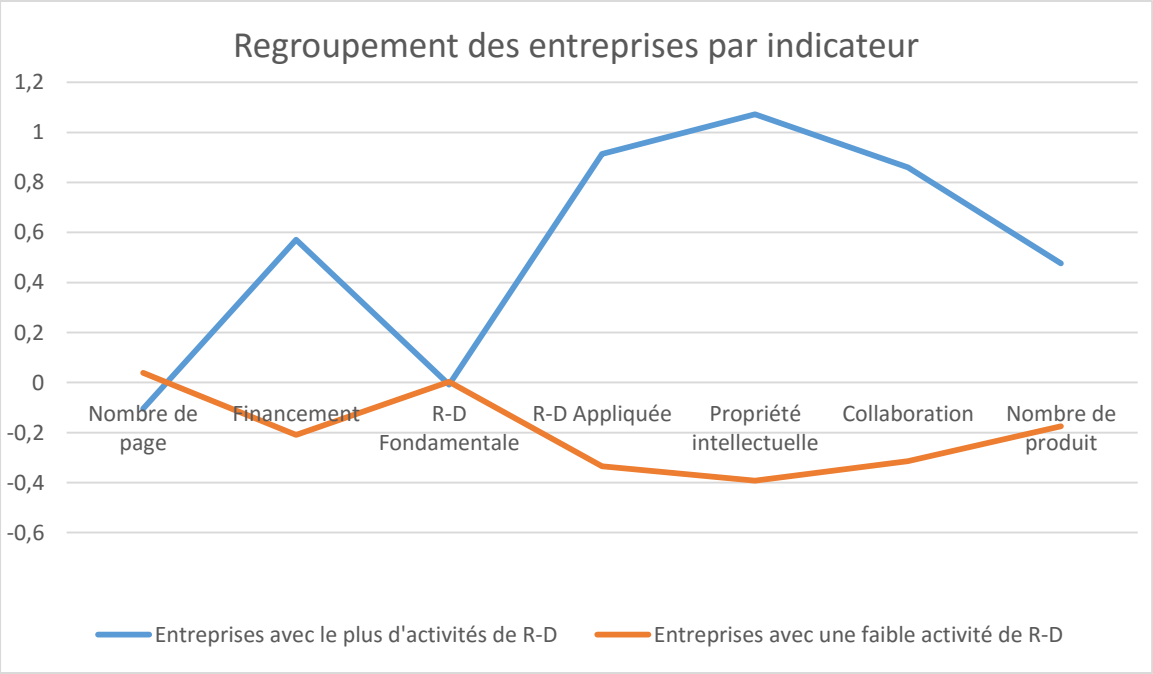

Figure 4.5 : Regroupement des entreprises avec distinction de la recherche en deux grappes

La Figure 4.5 nous montre le regroupement avec les nouveaux indicateurs. Ce nouveau regroupement classe 21 entreprises parmi les entreprises ayant une forte activité de R-D et 76 dans celles ayant une faible activité de R-D.

Comme nous pouvons constater, la grappe bleue comportant les entreprises commercialisant le plus grand nombre de produits effectue peu de recherche orientée fondamentale et inversement, les entreprises effectuant le plus de recherche fondamentale commercialisent le moins de produits comme le proposent C.-J. Lee et al. (2013)

Notre hypothèse 3 b) concernant l'impact négatif, de la recherche orientée fondamentale sur la commercialisation semble donc vérifiée.

La Figure 4.6 donne un regroupement des entreprises avec 3 grappes. Dans ce regroupement, nous retrouvons une grappe d'entreprises de 68 individus ayant peu d'activité. Une grappe comportant 14 entreprises propose le plus grand nombre de produit mais démontre peu d'activité de R-D, appliquée ou fondamentale. À l'opposé les sites mentionnant le plus d'activité de recherche, 15 entreprise effectuent presque autant de recherche fondamentale que de recherche appliquée. Cependant ces entreprises proposent peu de produit.

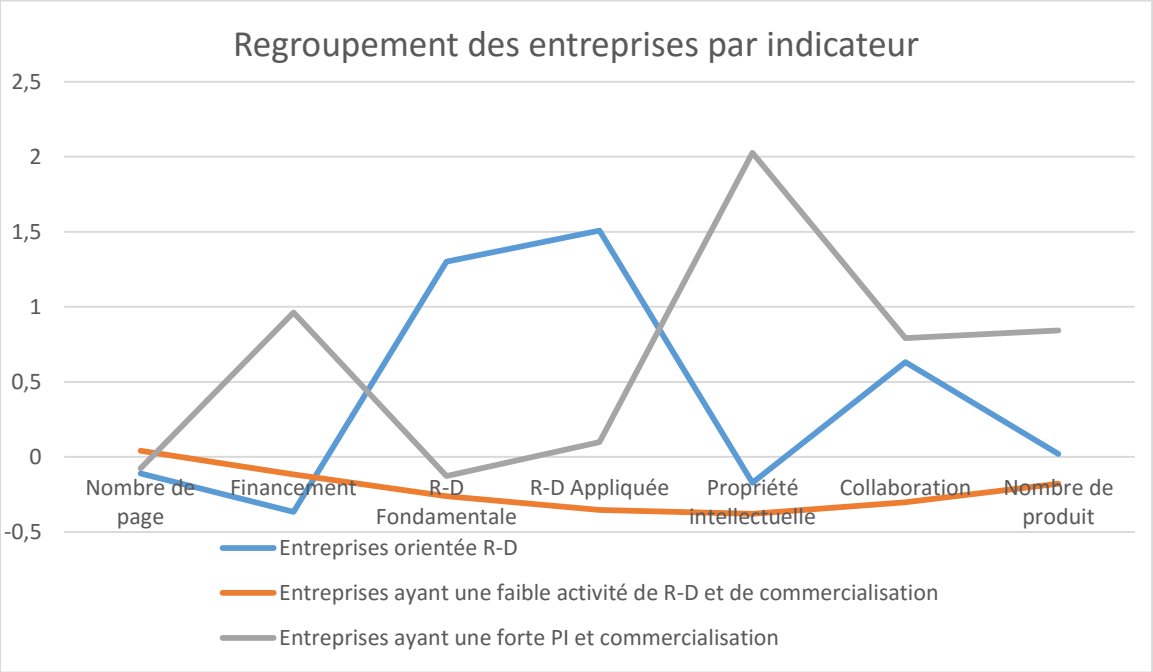

Figure 4.6 : Regroupement des entreprises avec distinction de la recherche en trois grappes

Un rapprochement avec l'article d'Autio (1997) peut être effectué pour expliquer ces deux catégories d'entreprise. Autio classe les entreprises basées sur les nouvelles technologies en deux catégories : les entreprises axées sur la science et les entreprises axées sur l'ingénierie.

Les entreprises axées sur les sciences se concentrent sur la production de technologies de base et sur des applications spécifiques. Ces entreprises ont tendance à effectuer beaucoup de recherche, fondamentale et appliqué et proposer peut de produit directement commercialisable pour les consommateurs.

À l'inverse, les entreprises axées sur l'ingénierie se concentrent principalement sur la production de produits à but commercial. Les firmes d'ingénierie sont plus proches des besoins du marché tandis que les autres sont plus motivés par la science (Autio 1997). Cette distinction entre les types d'entreprises est aussi effectuée dans l'étude d'Arora et al. (2013) dans leur étude des entreprises produisant du graphème. Les entreprises produisant du graphème et collaborant avec les universités ont plus tendance à être orienté vers la science tandis que les entreprises collaborant en elles sont orientés vers l'ingénierie. Il serait intéressant dans la suite du projet de regarder plus en détails, les type de collaboration effectué par les entreprises en fonction de leur profils orienté science ou ingénierie.

Avec notre méthodologie, il n'existe pas de distinction entre la recherche orientée fondamentale et appliquée. Les résultats des regroupements avec la distinction du type de recherche ne différent pas des regroupements précédents. On peut toutefois noté que dans les entreprises proposant le plus de produit, la recherche appliquée est plus importante que la recherche fondamentale.

Pour les entreprises ne donnant pas beaucoup d'information avec notre méthodologie, plusieurs raisons peuvent exister. Pour les plus petits sites, il est possible d'expliquer le manque d'information par un problème lié à notre récolte de donnée. Il est en effet possible que le robot n'ai par parcouru la totalité d'un site Internet, passant à côté de donnée pertinente pour notre étude. Pour les sites plus vastes, l'information pertinente est au contraire noyée par le grand nombre de données récolté et ne ressort pas suffisamment avec notre méthodologie.

CHAPITRE 5 DISCUSSION ET CONCLUSION

Ce mémoire présente une méthodologie visant à l'évaluation de l'innovation et de la commercialisation des nanotechnologies en employant les données présentes directement dans les sites Internet des entreprises. Nous avons établi une méthodologie de classification automatique des sites Internet selon la littérature portant sur la nanotechnologie et avons créé une matrice d'indicateurs basés sur l'innovation et la commercialisation. Avec cette matrice nous avons effectué des regroupements d'entreprises selon les indicateurs.

La nanotechnologie, de par son aspect multidisciplinaire et sa croissance rapide, est à l'aube de bouleverser de nombreuses industries déjà existantes. Considérée comme la nouvelle technologie à usage générale, la nanotechnologie a le potentiel de modifier de façon incrémentale une grande majorité de produits dans de très nombreux secteurs, mais aussi de créer des innovations plus radicales pouvant changer jusqu'à notre comportement social. De nombreux facteurs influence l'innovation et la commercialisation de cette technologie et c'est ce que nous avons essayé de mettre en avant dans cette recherche.

La nanotechnologie est encore en grande partie une technologie de laboratoire existant dans la recherche fondamentale effectuée par les universités. Si les efforts de recherche fondamentale se traduisent par de nombreuses publications, cette technologie semble connaitre plus de difficulté à générer des profits grâce à sa commercialisation au Canada, créant un décalage entre la recherche fondamentale des universités et la recherche appliquée effectuée en entreprise. À travers le modèle de la « Triple Hélice », nous avons constaté l'importance du gouvernement et de l'université pour l'amener la nanotechnologie à une commercialisation réussie. Le financement public, présent à tous les stades dans le processus d'innovation est dépendant de la politique d'innovation du gouvernement tout comme la capacité des universités à mener des recherches fondamentales.

De plus, la littérature concernant le type de développement suivi par les nanotechnologies semble encore hésiter entre deux modèles de développement. Un premier modèle propose un développement similaire à la biotechnologie avec le rôle des PMEs pour effectuer la recherche appliquée et permettre à la technologie d'arriver sur le marché (Darby et Zucker 2003, Rothaermel et Thursby 2007). Plus récemment Genet et al. (2012) proposent au contraire que les grandes entreprises mieux équipées et disposant de plus de fond pour la recherche et le développement

vont, comme ce fut le cas de la microélectronique, répandre les nanotechnologies auprès du grand public. Notre recherche permet en employant des sources de donnée complémentaire de donner un aperçu de l'industrie de la nanotechnologie canadienne à travers plusieurs indicateurs issus de la littérature, tels que le financement, les efforts de R-D et la taille des entreprises.

Notre recherche basée sur les sites Internet est encore peu documentée dans la littérature. Plusieurs études ont déjà permis de démontrer la faisabilité de ce type de recherche, mais le caractère de notre recherche reste en grande partie exploratoire. À la différence de Youtie et al. (2012) qui mènent une étude longitudinale sur la commercialisation des nanotechnologies en employant les archives du Web, notre recherche est transversale et représente l'activité nanotechnologique canadienne à un moment précis. Par rapport à la littérature, notre recherche se distingue par les indicateurs employés et par le sujet étudié. Aucune étude basée sur les sites Internet des entreprises canadiennes ayant des activités en nanotechnologie n'a encore été effectuée.

Notre méthode d'analyse permet de capturer les données brutes dans les sites Internet des entreprises pour extraire des indicateurs construits à grâce à des mots clés. Plusieurs limitations existent rien que par le support employé. Contrairement aux publications et aux brevets, l'information contenue dans les sites Internet n'est pas pérenne dans le temps, ce qui peut rendre délicat la répétabilité des études d'exploration de données Internet. Les sites Internet sont très différents les uns des autres de par leur structure, leur dimension et leur utilisation faite par les entreprises rendant leur analyse plus complexe.

L'analyse fréquentielle des mots clés est aussi limitée par les mots clés eux-mêmes. Nos indicateurs étant principalement basés sur la présence ou non de mots clés dans les sites étudiés, la stratégie de recherche de la liste de ces mots clés doit être la plus exhaustive et la plus robuste possible pour bien représenter l'indicateur. De plus le contexte dans lequel les mots sont employés est difficilement capturable pouvant mener à de nombreux faux positifs.

Nous avons choisi de récupérer nos données avec un robot pour les stocker dans une base de données. Au vu des performances de récupération des sites, 72% des sites initialement injectés ont été récupérés, notre infrastructure matérielle, ici notre serveur, n'est peut-être pas correctement adapté pour un aussi gros projet de moissonnage, malgré notre méthode de récupération par lot. Cependant, ce taux de récupération des données est à confronter aux taux de réponse des études menées avec questionnaire en ligne atteignant des taux de réponse entre 10% et 25% (Sauermann

and Roach 2013). Notre récupération de données connait donc des limitations, mais les taux de réponse sont très supérieurs aux études par questionnaire en ligne. Enfin, notre analyse est basée sur le texte auto reportée par les entreprises présentant les informations qu'elles veulent bien divulguer. Cependant cette limitation s'applique aussi aux études par questionnaire.

À partir des données récupérées, notre processus de classification nous a permis d'identifier les sites d'entreprises ayant des activités en nanotechnologie. L'identification des entreprises est effectuée à partir d'une liste de mots clés issue de la littérature. De par son caractère multidisciplinaire, la nanotechnologie fait l'objet de plusieurs types de classifications différentes avec un niveau de précision plus ou moins élevé. La classification choisie pour notre recherche à l'avantage de comporter peu de classes différentes, facilitant la répartition. Notre méthodologie nous a permis d'identifier 97 entreprises ayant des activités de nanotechnologie, réparties de manière inégale dans les domaines de la nanotechnologie.

Au final, la classification des entreprises nous a permis de vérifier notre hypothèse concernant le domaine le plus représenté au Canada. Comme le propose Roco (2001), les premiers produits nanotechnologiques commercialisés sont les nanomatériaux, car ils concernent les produits les moins complexes de la nanotechnologie. Les nanomatériaux et nanostructures passives et actives sont les premiers produits nanotechnologie à avoir été commercialisé à cause de leur simplicité relative et d'une bonne compréhension de la part des industries grâce à la recherche effectuée dans ce domaine. Il n'est donc pas surprenant de constater que 63% des entreprises canadiennes appartiennent à ce domaine de la nanotechnologie. Le grand pourcentage d'entreprise dans ce domaine explique aussi que le nombre de produits dans ce domaine soit le plus élevé, validant ainsi notre hypothèse.

Le regroupement géographique des entreprises permet de montrer l'incidence de la recherche et la présence d'inventeur sur la commercialisation des nanotechnologies. Même si les grappes en nanotechnologie correspondent en majorité aux grandes villes canadiennes, cela montre l'importance des retombées de la recherche effectuée dans ces grandes villes, de la présence de main-d'œuvre qualifiée et des inventeurs ayant des activités en nanotechnologie pour créer des grappes en nanotechnologie. Un contre-exemple peut être illustré avec la ville de Winnipeg, 7e ville canadienne de par sa population n'apparaissant ni dans les grappes de brevet en nanotechnologie

identifiées par Schiffauerova and Beaudry (2009) ni dans les grappes d'entreprise identifiées par notre recherche.

Enfin les regroupements d'entreprises suivant les indicateurs nous ont permis de valider nos hypothèses concernant les facteurs influant la commercialisation des produits en nanotechnologie. Les résultats de notre étude semblent confirmer que les entreprises participant le plus à la commercialisation des nanotechnologies sont ont un profil type d'entreprise de taille moyenne, disposant de financement, propriété intellectuelle et effectuant des collaborations avec des universités ou d'autres entreprises. Notre recherche a mis en avant le fait que le financement participe bien à amener des produits sur le marché. Cependant contrairement à ce qu'on pourrait penser, l'activité de recherche et plus particulièrement la recherche orientée vers la science pure ne favorise pas la commercialisation des nanotechnologies, mais semble distraire au contraire les entreprises de leur objectif de commercialiser des produits comme le suggèrent C.-J. Lee et al. (2013). Ainsi, les entreprises réussissant le mieux le passage de la nanotechnologie des laboratoires vers le marché public s'attardent moins sur la science derrière la nanotechnologie, mais se concentrent au contraire sur les applications commercialisables, validant les recommandations de Tsuzuki (2013).

Au final, si le Canada est bien placé au niveau mondial pour la recherche effectuée dans le domaine de la nanotechnologie, il semble qu'il n'y est pas encore suffisamment d'entreprises pour appliquer cette recherche sur des produits commercialisables et générer des revenus.

En conclusion, notre mémoire de recherche présente les possibilités et les limitations d'une étude d'innovation basée sur les informations présentes librement sur internet. Notre méthodologie permet de nous affranchir des bases de données difficilement accessibles, des enquêtes téléphoniques couteuses et des questionnaires Internet peu considérés par les entreprises. Beaucoup de possibilités et d'amélioration restent encore possibles pour cette recherche concernant les nanotechnologies au Canada. Enfin une validation de notre recherche pour vérifier nos résultats pourrait être effectuée en se fiant à des données externes. La masse d'information présente sur internet offrant beaucoup d'opportunité de recherche, pour peu que l'on puisse interpréter ces données, ce type d'étude basée sur internet apporte une véritable valeur ajoutée et offre une source de donnée complémentaire intéressante par rapport aux sources conventionnelles.

BIBLIOGRAPHIE

Anselin, Luc, Attila Varga, and Zoltan Acs. 1997. "Local Geographic Spillovers between University Research and High Technology Innovations." *Journal of Urban Economics* 42 (3): 422–48. doi:10.1006/juec.1997.2032.

Arora, Sanjay K., Jan Youtie, Philip Shapira, Lidan Gao, and TingTing Ma. 2013. "Entry Strategies in an Emerging Technology: A Pilot Web-Based Study of Graphene Firms." *Scientometrics* 95 (3): 1189–1207. doi:10.1007/s11192-013-0950-7.

Audretsch, David B., and T. Taylor Aldridge. 2009. "Scientist Commercialization as Conduit of Knowledge Spillovers." *The Annals of Regional Science* 43 (4): 897–905. doi:10.1007/s00168-009-0297-4.

Autio, E. 1997. "New, Technology-Based Firms in Innovation Networks Symplectic and Generative Impacts." *Research Policy* 26 (3): 263–81. doi:10.1016/S0048-7333(96)00906-7.

Baron, Justus, and Henry Delcamp. 2010. "Assessing Indicators of Patent Quality: Complex vs. Discrete Technologies." *HAL Post-Print*. http://core.kmi.open.ac.uk/download/pdf/6507555.pdf.

Bayus, Barry L. 1994. "Are Product Life Cycles Really Getting Shorter?" *Journal of Product Innovation Management* 11 (4): 300–308.

Beaudry, Catherine, and Andrea Schiffauerova. 2011. "Is Canadian Intellectual Property Leaving Canada? A Study of Nanotechnology Patenting." *The Journal of Technology Transfer* 36 (6): 665–79. doi:10.1007/s10961-011-9211-1.

Bensaude-Vincent, Bernadette. 2009. "Nanotechnologies : une révolution annoncée." *Études* 411 (12): 605–16.

Besen, Stanley M., and Leo J. Raskind. 1991. "An Introduction to the Law and Economics of Intellectual Property." *The Journal of Economic Perspectives*, 3–27.

Blume-Kohout, Margaret E., Krishna B. Kumar, and Neeraj Sood. 2009. "Federal Life Sciences Funding and University R&D." National Bureau of Economic Research. http://www.nber.org/papers/w15146.

Bradley, Samantha R., Christopher S. Hayter, and Albert Link. 2013. "Models and Methods of University Technology Transfer." *Foundations and Trends in Entrepreneurship* 9 (6). http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2380317.

Bush, Vannevar. 1945. "Science: The Endless Frontier." *Transactions of the Kansas Academy of Science (1903)*, 231–64.

Campbell, David, Michelle Picard-Aitken, Grégoire Côté, Julie Caruso, Rodolfo Valentim, Stuart Edmonds, Gregory Thomas Williams, et al. 2010. "Bibliometrics as a Performance Measurement Tool for Research Evaluation: The Case of Research Funded by the National Cancer Institute of Canada." *American Journal of Evaluation* 31 (1): 66–83.

Chen, Zhen, Huan Meng, Gengmei Xing, Chunying Chen, Yuliang Zhao, Guang Jia, Tiancheng Wang, et al. 2006. "Acute Toxicological Effects of Copper Nanoparticles in Vivo." *Toxicology Letters* 163 (2): 109–20.

Chesbrough, Henry. 2003. “The Logic of Open Innovation: Managing Intellectual Property.” *California Management Review* 45 (3): 33–58.

Cierpicki, Steven, Malcolm Wright, Byron Sharp, and others. 2000. “Managers’ Knowledge of Marketing Principles: The Case of New Product Development.” *Journal of Empirical Generalizations in Marketing Science* 5 (3): 771–90.

Cohen, Wesley M., Richard R. Nelson, and John P. Walsh. 2000. “Protecting Their Intellectual Assets: Appropriability Conditions and Why US Manufacturing Firms Patent (or Not).” National Bureau of Economic Research. http://www.nber.org/papers/w7552.

Colvin, Vicki L. 2003. “The Potential Environmental Impact of Engineered Nanomaterials.” *Nature Biotechnology* 21 (10): 1166–70.

Commission, EC-European, and others. 2005. “Nanosciences and Nanotechnologies: An Action Plan for Europe 2005-2009.” *Brussels: EC*.

Co-operation, Organisation for Economic, and Development. 1994. *The Measurement of Scientific and Technological Activities: Proposed Standard Practice for Surveys of Research and Experimental Development: Frascati Manual 1993*. OECD. http://www.sidalc.net/cgi-bin/wxis.exe/?IsisScript=COLPOS.xis&method=post&formato=2&cantidad=1&expresion=mfn=035177.

Crawley, Tom. 2007. “Commercialization of Nanotechnology–Key Challenges.” In *Workshop Organised by Nanoforum, Helsinki*. http://www.nanowerk.com/nanotechnology/reports/reportpdf/report73.pdf.

Crichton, Michael, and Patrick Berthon. 2004. *La Proie*. Pocket. http://cops.bozs.org/download/2667/La+proie+-+Michael+Crichton.pdf.

Darby, Michael R., and Lynne G. Zucker. 2003. “Grilichesian Breakthroughs: Inventions of Methods of Inventing and Firm Entry in Nanotechnology.” National Bureau of Economic Research. http://www.nber.org/papers/w9825.

Development, Organisation for Economic Co-operation and. 1993. *Small and Medium-Sized Enterprises: Technology and Competitiveness*. Organisation for Economic Co-operation and Development.

Domeniconi, Carlotta, Daniel Barbará, Harsh Chaudhary, Ali Al-Timimi, and Curtis Jamison. 2005. “DATA-MINING TECHNIQUES FOR MICROARRAY DATA ANALYSIS.” *Next Generation of Data-Mining Applications*, 295.

Dong, Guozhu, Jinyan Li, and Limsoon Wong. 2004. “The Use of Emerging Patterns in the Analysis of Gene Expression Profiles for the Diagnosis and Understanding of Diseases.” *New Generation of Data Mining Applications, IEEE Press/Wiley*. https://personal.comp.nus.edu.sg/~wongls/psZ/guozhu-bk-7.pdf.

Dosi, Giovanni, Christopher Freeman, Richard Nelson, Gerald Silverberg, Luc Soete, and others. 1988. *Technical Change and Economic Theory*. Vol. 988. Pinter London. http://tocs.ulb.tu-darmstadt.de/9738215.pdf.

Dreher, Kevin L. 2004. “Health and Environmental Impact of Nanotechnology: Toxicological Assessment of Manufactured Nanoparticles.” *Toxicological Sciences* 77 (1): 3–5.

Drexler, K. Eric. 1992. *Nanosystems: Molecular Machinery, Manufacturing, and Computation*. John Wiley & Sons, Inc. http://dl.acm.org/citation.cfm?id=135735.

Etzkowitz, Henry, and Loet Leydesdorff. 2000. "The Dynamics of Innovation: From National Systems and 'Mode 2' to a Triple Helix of University–industry–government Relations." *Research Policy* 29 (2): 109–23. doi:10.1016/S0048-7333(99)00055-4.

Fall, C. J., A. Törcsvári, K. Benzineb, and G. Karetka. 2003. "Automated Categorization in the International Patent Classification." *SIGIR Forum* 37 (1): 10–25. doi:10.1145/945546.945547.

Feynman, Richard P. 1960. "There's Plenty of Room at the Bottom." *Engineering and Science* 23 (5): 22–36.

Fitzgibbons, Kevin, and Chuck McNiven. 2006. "Towards a Nanotechnology Statistical Framework." In *Blue Sky Indicators Conference II*. http://www.oecd.org/sti/inno/37450255.pdf.

Gans, Joshua S., David H. Hsu, and Scott Stern. 2000. "When Does Start-up Innovation Spur the Gale of Creative Destruction?" National Bureau of Economic Research. http://www.nber.org/papers/w7851.

Genet, Corine, Khalid Errabi, and Caroline Gauthier. 2012. "Which Model of Technology Transfer for Nanotechnology? A Comparison with Biotech and Microelectronics." *Technovation*, Nanotechnology: Introducing the Future, 32 (3–4): 205–15. doi:10.1016/j.technovation.2011.10.007.

Gibbons, Michael, Camille Limoges, Helga Nowotny, Simon Schwartzman, Peter Scott, and Martin Trow. 1994. *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies*. SAGE.

Godin, Benoit. 2004. *Measurement and Statistics on Science and Technology: 1920 to the Present*. Routledge. https://books.google.ca/books?hl=en&lr=&id=vFjVxQIw6aEC&oi=fnd&pg=PP1&dq=Godin,+2005+Measurement+and+Statistics+on+Science+and&ots=dpsFwmqYm8&sig=yTyIiBFtdl0X2Wl5Hn1LKyr0jRc.

Godin, Benoît. 2006. "The Linear Model of Innovation the Historical Construction of an Analytical Framework." *Science, Technology & Human Values* 31 (6): 639–67.

Godin, Benoit, and Yves Gingras. 2000. "The Place of Universities in the System of Knowledge Production." *Research Policy* 29 (2): 273–78. doi:10.1016/S0048-7333(99)00065-7.

Gök, Abdullah, Alec Waterworth, and Philip Shapira. 2014. "Use of Web Mining in Studying Innovation." *Scientometrics*, September, 1–19. doi:10.1007/s11192-014-1434-0.

Grinbaum, Alexei. 2006. "Cognitive Barriers in Perception of Nanotechnology." *The Journal of Law, Medicine & Ethics* 34 (4): 689–94.

Hullmann, Angela. 2007. "Measuring and Assessing the Development of Nanotechnology." *Scientometrics* 70 (3): 739–58. doi:10.1007/s11192-007-0310-6.

Islam, Nazrul, and Kumiko Miyazaki. 2010. "An Empirical Analysis of Nanotechnology Research Domains." *Technovation* 30 (4): 229–37. doi:10.1016/j.technovation.2009.10.002.

IV, John Elder, and Thomas Hill. 2012. *Practical Text Mining and Statistical Analysis for Non-Structured Text Data Applications*. Academic Press.

Kamei, S. 2002. "Promoting Japanese Style Nanotechnology Enterprises." *Mitsubishi Research Institute*.

Kim, Jinyoung, Sangjoon John Lee, and Gerald Marschke. 2008. "Impact of University Scientists on Innovations in Nanotechnology." https://books.google.ca/books?hl=en&lr=&id=_g3oAwAAQBAJ&oi=fnd&pg=PA141&dq=Impact+of+University+Scientists+on+Innovations+in+Nanotechnology&ots=G1MGKu9MxE&sig=IgVBG7StLuwEHqYfsKYFqXuXL0c.

Kleinknecht, Alfred, Kees Van Montfort, and Erik Brouwer. 2002. "The Non-Trivial Choice between Innovation Indicators." *Economics of Innovation and New Technology* 11 (2): 109–21. doi:10.1080/10438590210899.

Kostoff, Ronald, Raymond Koytcheff, and Clifford Lau. 2007. "Global Nanotechnology Research Metrics." *Scientometrics* 70 (3): 565–601. doi:10.1007/s11192-007-0303-5.

Kotsiantis, Sotiris B., I. Zaharakis, and P. Pintelas. 2007. *Supervised Machine Learning: A Review of Classification Techniques*. https://books.google.ca/books?hl=en&lr=&id=vLiTXDHr_sYC&oi=fnd&pg=PA3&dq=Kotsiantis,+2007&ots=CXnssB2Bmn&sig=5sSxzM6dTMFy7VzvlFlha-I2z70.

Krugman, Paul R. 1991. *Geography and Trade*. MIT press. https://books.google.ca/books?hl=en&lr=&id=AQDodCHOgJYC&oi=fnd&pg=PP7&dq=Krugman+(1991)&ots=Pzh-2iHJaq&sig=UUoD4rJTlz6gzpNSXMQKhN5tOiE.

Lamy, Jérôme. 2007. "Penser les rapports entre sciences et politique : enjeux historiographiques récents." *Cahiers d'histoire. Revue d'histoire critique*, no. 102 (October): 9–32.

Landis, J. Richard, and Gary G. Koch. 1977. "The Measurement of Observer Agreement for Categorical Data." *Biometrics*, 159–74.

Lee, Cheol-Ju, SuKap Lee, Myung S. Jhon, and Juneseuk Shin. 2013. "Factors Influencing Nanotechnology Commercialization: An Empirical Analysis of Nanotechnology Firms in South Korea." *Journal of Nanoparticle Research* 15 (2): 1–17.

Lee, Raymond M. 2000. *Unobtrusive Methods in Social Research*. Open Univ Pr.

Levin, Richard C., Alvin K. Klevorick, Richard R. Nelson, Sidney G. Winter, Richard Gilbert, and Zvi Griliches. 1987. "Appropriating the Returns from Industrial Research and Development." *Brookings Papers on Economic Activity*, 783–831.

Leydesdorff. 2011. "The Triple Helix, Quadruple Helix, …, and an N -Tuple of Helices: Explanatory Models for Analyzing the Knowledge-Based Economy? - Springer." doi:10.1007/s13132-011-0049-4.

Libaers, Dirk, Martin Meyer, and Aldo Geuna. 2006. "The Role of University Spinout Companies in an Emerging Technology: The Case of Nanotechnology." *The Journal of Technology Transfer* 31 (4): 443–50.

Linton, Jonathan D., and Steven T. Walsh. 2004. "Integrating Innovation and Learning Curve Theory: An Enabler for Moving Nanotechnologies and Other Emerging Process Technologies into Production." *R&D Management* 34 (5): 517–26.

Lipsey, Richard G., Kenneth I. Carlaw, and Clifford T. Bekar. 2005. *Economic Transformations : General Purpose Technologies and Long-Term Economic Growth: General Purpose Technologies and Long-Term Economic Growth*. OUP Oxford.

Li, Xin, Hsinchun Chen, Zan Huang, and Mihail C. Roco. 2007. "Patent Citation Network in Nanotechnology (1976–2004)." *Journal of Nanoparticle Research* 9 (3): 337–52.

Lorenzoni, Gianni, Damiano Russo, and Simone Ferriani. 2010. "Unfolding Local Knowledge and Field Emergence: The Rise of Bologna Nanotech." Working Paper, Department of Management. University of Bologna.

Lundvall, Bengt-\AAke, Björn Johnson, Esben Sloth Andersen, and Bent Dalum. 2002. "National Systems of Production, Innovation and Competence Building." *Research Policy* 31 (2): 213–31.

Lux Research. 2007. "The Nanotech Report 5th Edition -- Volume 1." https://portal.luxresearchinc.com/research/document_excerpt/2861.

Manach, Jean-Marc. 2009. "Nanotechnologies : le syndrome OGM." *Le Monde.fr*, April 10. http://www.lemonde.fr/technologies/article/2009/04/10/nanotechnologies-le-syndrome-ogm_1179126_651865.html.

Mangematin, Vincent, Khalid Errabi, and Caroline Gauthier. 2011. "Large Players in the Nanogame: Dedicated Nanotech Subsidiaries or Distributed Nanotech Capabilities?" *The Journal of Technology Transfer* 36 (6): 640–64.

Mangematin, Vincent, and Steve Walsh. 2012. "The Future of Nanotechnologies." *Technovation*, Nanotechnology: Introducing the Future, 32 (3–4): 157–60. doi:10.1016/j.technovation.2012.01.003.

Maniet, Françoise. 2010. "L'encadrement Juridique Des Nanotechnologies Au Canada et Dans l'Union Européenne." http://www.archipel.uqam.ca/id/eprint/3615.

Mann, Henry B., and Donald R. Whitney. 1947. "On a Test of Whether One of Two Random Variables Is Stochastically Larger than the Other." *The Annals of Mathematical Statistics*, 50–60.

Maynard, Andrew D., and others. 2006. "A Research Strategy for Addressing Risk." *Nanotechnology, Woodrow Wilson International Center for Scholars*. http://tinhoahoc.com/Nanotechnology/RiskRelatedResearch_Maynard_7-06-Final.pdf.

McNeil, R. D., J. Lowe, T. Mastroianni, D. Ferk, and J. Cronin. 2007. "Barriers to Nanotechnology Commercialization." *Report Prepared for U.S. Department of Commerce, The University of Illinois at Springfield, College of Business and Management.* http://www.ntis.gov/assets/pdf/Report-BarriersNanotechnologyCommercialization.pdf.

McNiven, Chuck. 2007. *Overview and Discussion of the Results of the Pilot Survey on Nanotechnology in Canada*. Statistics Canada, Science and Technology Surveys Section, Science, Innovation and Electronic Information Division. http://www.statcan.gc.ca/pub/88f0006x/88f0006x2007005-eng.pdf.

Mendonça, Sandro, Tiago Santos Pereira, and Manuel Mira Godinho. 2004. "Trademarks as an Indicator of Innovation and Industrial Change." *Research Policy*, What do we know

Innovation? Selected papers from an International Conference in honour of Keith Pavitt, 33 (9): 1385–1404. doi:10.1016/j.respol.2004.09.005.

Miyazaki, Kumiko, and Nazrul Islam. 2007. “Nanotechnology Systems of innovation—An Analysis of Industry and Academia Research Activities.” *Technovation* 27 (11): 661–75. doi:10.1016/j.technovation.2007.05.009.

Moore, Gordon E. 2005. “Cramming More Components onto Integrated Circuits, Electronics, 38: 8 (1965).” *URL: Ftp://download. Intel. Com/research/silicon/moorespaper. Pdf* 16.

Nevens, T. Michael. 1990. “Commercializing Technology: What the Best Companies Do.” *Planning Review* 18 (6): 20–24.

OCDE, and Eurostat. 2005. *Manuel d’Oslo*. Paris: Organisation for Economic Co-operation and Development. http://www.oecd-ilibrary.org/content/book/9789264013124-fr.

Palmberg, Christopher. 2007. “The Transfer and Commercialisation of Nanotechnology: A Comparative Analysis of University and Company Researchers.” *The Journal of Technology Transfer* 33 (6): 631–52. doi:10.1007/s10961-007-9059-6.

Palmberg, Christopher, Hélène Dernis, and Claire Miguet. 2009. *Nanotechnology: An Overview Based on Indicators and Statistics*. OECD. http://recherche.chu-sainte-justine.org/documents/General/CdeR/Chaires/R%C3%A9seau%20NE3LS/OCDE-Nanotechnology%20An%20Overview...%20juin%202009.pdf.

Pavitt, K. 1985. “Patent Statistics as Indicators of Innovative Activities: Possibilities and Problems.” *Scientometrics* 7 (1-2): 77–99. doi:10.1007/BF02020142.

Porter, Alan L., Jan Youtie, Philip Shapira, and David J. Schoeneck. 2008. “Refining Search Terms for Nanotechnology.” *Journal of Nanoparticle Research* 10 (5): 715–28. doi:10.1007/s11051-007-9266-y.

RAMDANI, ANAS, VUE DE, DU DIPLÔME DE MAÎTRISE L’OBTENTION, and others. 2014. “REVUE SYSTÉMATIQUE DE LA LITTÉRATURE SUR LES MESURES DE LA COLLABORATION INTER-ORGANISATIONNELLE DANS UN CONTEXTE D’INNOVATION.” http://publications.polymtl.ca/1624/1/2014_AnasRamdani.pdf.

Ramos, Juan. 2003. “Using Tf-Idf to Determine Word Relevance in Document Queries.” In *Proceedings of the First Instructional Conference on Machine Learning*. https://www.cs.rutgers.edu/~mlittman/courses/ml03/iCML03/papers/ramos.pdf.

Renn, Ortwin, and Mihail C. Roco. 2006. “Nanotechnology and the Need for Risk Governance.” *Journal of Nanoparticle Research* 8 (2): 153–91.

Rieu, Alain-Marc. 2011. “Fourth Helix: Sustaining Research Diversity The Social Turn of Science and Technology Policy. The Case of Japan.” https://hal.archives-ouvertes.fr/hal-00701243/.

Roco, Mihail C. 2001. “International Strategy for Nanotechnology Research.” *Journal of Nanoparticle Research* 3 (5-6): 353–60.

———. 2011. “The Long View of Nanotechnology Development: The National Nanotechnology Initiative at 10 Years.” In *Nanotechnology Research Directions for Societal Needs in 2020*, 1–28. Science Policy Reports 1. Springer Netherlands. http://link.springer.com/chapter/10.1007/978-94-007-1168-6_1.

Rodgers, Joseph Lee, and W. Alan Nicewander. 1988. “Thirteen Ways to Look at the Correlation Coefficient.” *The American Statistician* 42 (1): 59–66. doi:10.1080/00031305.1988.10475524.

Rothaermel, Frank T., and Marie Thursby. 2007. “The Nanotech versus the Biotech Revolution: Sources of Productivity in Incumbent Firm Research.” *Research Policy* 36 (6): 832–49.

Rothen-Rutishauser, Barbara M., Samuel Schürch, Beat Haenni, Nadine Kapp, and Peter Gehr. 2006. “Interaction of Fine Particles and Nanoparticles with Red Blood Cells Visualized with Advanced Microscopic Techniques.” *Environmental Science & Technology* 40 (14): 4353–59.

Saner, Marc, and Jennifer Pelley. 2009. “International Approaches to the Regulatory Governance of Nanotechnology.” *Regulation Papers, April*. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1532390.

Sauermann, Henry, and Michael Roach. 2013. “Increasing Web Survey Response Rates in Innovation Research: An Experimental Study of Static and Dynamic Contact Design Features.” *Research Policy* 42 (1): 273–86. doi:10.1016/j.respol.2012.05.003.

Schiffauerova, Andrea, and Catherine Beaudry. 2009. “Canadian nanotechnology innovation networks: intra-cluster, inter-cluster and foreign collaboration.” *Journal of Innovation Economics* 4 (2): 119. doi:10.3917/jie.004.0119.

Schumpeter, Joseph Alois. 1934. *The Theory of Economic Development: An Inquiry Into Profits, Capital, Credit, Interest, and the Business Cycle*. Transaction Publishers.

Shapira, Philip, and Jan Youtie. 2006. “Measures for Knowledge-Based Economic Development: Introducing Data Mining Techniques to Economic Developers in the State of Georgia and the US South.” *Technological Forecasting and Social Change* 73 (8): 950–65.

Siegel, Donald S., David A. Waldman, Leanne E. Atwater, and Albert N. Link. 2004. “Toward a Model of the Effective Transfer of Scientific Knowledge from Academicians to Practitioners: Qualitative Evidence from the Commercialization of University Technologies.” *Journal of Engineering and Technology Management* 21 (1): 115–42.

Singhal, Amit. 2001. “Modern Information Retrieval: A Brief Overview.” *IEEE Data Eng. Bull.* 24 (4): 35–43.

Svensson, Roger. 2007. “Commercialization of Patents and External Financing during the R&D Phase.” *Research Policy* 36 (7): 1052–69. doi:10.1016/j.respol.2007.04.004.

Tahmooresnejad, Leila, Catherine Beaudry, and Andrea Schiffauerova. 2015. “The Role of Public Funding in Nanotechnology Scientific Production: Where Canada Stands in Comparison to the United States.” *Scientometrics* 102 (1): 753–87.

Taniguchi, Norio, and others. 1974. “On the Basic Concept of Nanotechnology.” In *Proc. Intl. Conf. Prod. Eng. Tokyo, Part II, Japan Society of Precision Engineering*, 18–23.

Tether, Bruce S. 2002. “Who Co-Operates for Innovation, and Why: An Empirical Analysis.” *Research Policy* 31 (6): 947–67.

Tsuzuki, Takuya. 2009. “Commercial Scale Production of Inorganic Nanoparticles.” *International Journal of Nanotechnology* 6 (5/6): 567. doi:10.1504/IJNT.2009.024647.

———. 2013. *Nanotechnology Commercialization*. CRC Press. https://books.google.ca/books?hl=en&lr=&id=-U_GAAAAQBAJ&oi=fnd&pg=PP1&dq=commercialization+nanotechnology&ots=tPPyhIFXT0&sig=M_cd0nxrz3DegVvkkW2K9VAJOpM.

Van der Lei, T. E., and S. W. Cunningham. 2006. "Use of the Internet for Future-Oriented Technology Analysis." In *Second International Seville Seminar on Future-Oriented Technology Analysis, September, Seville*.

Youtie, Jan, Diana Hicks, Philip Shapira, and Travis Horsley. 2012. "Pathways from Discovery to Commercialisation: Using Web Sources to Track Small and Medium-Sized Enterprise Strategies in Emerging Nanotechnologies." *Technology Analysis & Strategic Management* 24 (10): 981–95. doi:10.1080/09537325.2012.724163.

ANNEXE A – INSTALLATION ET CONFIGURATION DE NUTCH AVEC MYSQL

Installation de Mysql :

Ligne de commande d'installation de Mysql : ***sudo apt-get install mysql-server mysql-client***

L'ordinateur doit être redémarré pour prendre en compte l'installation.

Pour verifier installation, taper ***sudo netstat -tap | grep mysql***.

Si l'installation c'est correctement déroulée le message suivant apparait:

tcp 0 0 localhost:mysql *:* LISTEN

Création de la table webpage:

Pour se connecter au serveur, taper ***mysql -u xxxxx –p*** avec ***xxxx*** le nom d'utilisateur.

Dans l'éditeur Mysql, nous créons premièrement une nouvelle base de données :

CREATE DATABASE nutch DEFAULT CHARACTER SET utf8mb4 DEFAULT COLLATE utf8mb4_unicode_ci;

On utilise la base de données avec l'instruction suivante :

use nutch;

On créé la table "webpage" :

```
CREATE TABLE `webpage` (
`id` varchar(767) NOT NULL,
`headers` blob,
`text` longtext DEFAULT NULL,
`status` int(11) DEFAULT NULL,
`markers` blob,
`parseStatus` blob,
`modifiedTime` bigint(20) DEFAULT NULL,
`prevModifiedTime` bigint(20) DEFAULT NULL,
`score` float DEFAULT NULL,
`typ` varchar(32) CHARACTER SET latin1 DEFAULT NULL,
`batchId` varchar(32) CHARACTER SET latin1 DEFAULT NULL,
`baseUrl` varchar(767) DEFAULT NULL,
`content` longblob,
`title` varchar(2048) DEFAULT NULL,
`reprUrl` varchar(767) DEFAULT NULL,
`fetchInterval` int(11) DEFAULT NULL,
```

```
`prevFetchTime` bigint(20) DEFAULT NULL,
`inlinks` mediumblob,
`prevSignature` blob,
`outlinks` mediumblob,
`fetchTime` bigint(20) DEFAULT NULL,
`retriesSinceFetch` int(11) DEFAULT NULL,
`protocolStatus` blob,
`signature` blob,
`metadata` blob,
PRIMARY KEY (`id`)
)

ENGINE=InnoDB
ROW_FORMAT=COMPRESSED
DEFAULT CHARSET=utf8mb4;
```

Installation de Nutch 2.2.1

Télécharger la dernière version à l'adresse suivante : http://www.apache.org/dyn/closer.cgi/nutch/

Dans le fichier nutch, éditer le document ${APACHE_NUTCH_HOME}/ivy/ivy.xml pour s'assurer que la dépendance entre MySQL et Nutch est disponible

Changer:
<dependency org="org.apache.gora" name="gora-core" rev="0.3" conf="*->default"/>
en:
<dependency org="org.apache.gora" name="gora-core" rev="0.2.1" conf="*->default"/>

Dé commenter gora-sql
<dependency org="org.apache.gora" name="gora-sql" rev="0.1.1-incubating" conf="*->default" />

Dé commenter le connecteru mysql
<!– Uncomment this to use MySQL as database with SQL as Gora store. –>
<dependency org="mysql" name="mysql-connector-java" rev="5.1.18" conf="*->default"/>

Éditer le fichier ***${APACHE_NUTCH_HOME}/conf/gora.properties*** et remplacer le Default SqlStore Properties. Par le MySQL properties suivant en remplacant xxxxx par le nom d'utilisateur et le mot de passe défini lors de l'installation.

```
################################
# MySQL properties #
################################
gora.sqlstore.jdbc.driver=com.mysql.jdbc.Driver
```

gora.sqlstore.jdbc.url=jdbc:mysql://localhost:3306/nutch?createDatabaseIfNotExist=true
gora.sqlstore.jdbc.user=xxxxx
gora.sqlstore.jdbc.password=xxxxx

Éditer le fichier ${APACHE_NUTCH_HOME}/conf/gora-sql-mapping.xml pour changer la longueur de la clée primaire au deux endroits.
<primarykey column="id" length="767"/>

Configurer ${APACHE_NUTCH_HOME}/conf/nutch-site.xml en changeant les valeurs si nécessaire. C'est dans ce fichier que les plugins sont à intégrer ainsi que les règles de parcourt de lien

```
<property>
<name>http.agent.name</name>
<value>YourNutchSpider</value>
</property>

<property>
<name>http.accept.language</name>
<value> en-us,en-gb,en;q=0.7,*;q=0.3</value>
<description>Value of the "Accept-Language" request header field.
This allows selecting non-English language as default one to retrieve.
It is a useful setting for search engines build for certain national group.
</description>
</property>

<property>
<name>parser.character.encoding.default</name>
<value>utf-8</value>
<description>The character encoding to fall back to when no other information
is available</description>
</property>

<property>
<name>storage.data.store.class</name>
<value>org.apache.gora.sql.store.SqlStore</value>
<description>The Gora DataStore class for storing and retrieving data.
Currently the following stores are available: ….
</description>
</property>
```

Installation du compilateur ant: ***sudo apt-get install ant***

Aller dans le repertoire ${APACHE_NUTCH_HOME} et taper ***ant runtime*** pour compiler le logiciel. Cela peut prendre plusieurs minutes.

Creer un nouveau fichier seed.txt dans ***${APACHE_NUTCH_HOME}/runtime/local/URL*** contenant les sites Internet à parcourir.

:

bin/nutch inject urls

bin/nutch generate -topN 20
bin/nutch fetch -all
bin/nutch parse –all
bin/nutch updatedb

ANNEXE B – SCRIPT DE DÉTECTION DE LANGAGE

```
# coding=utf-8
# Author        : Ellias Zainea
#
# This script goes through every page recovered by Nutch
# (where ‘status = 2 and text is not null’) to detect and
# set the language (either ‘english’, ‘french’ or ‘other’).
#

import sys
import argparse
import MySQLdb
import cld2full

# Create an argument parser and add an argument to it.
parser = argparse.ArgumentParser(description=‘Detect the language of parsed webpages with a status = 2 and where the text is no null.’)
parser.add_argument(‘inputTableName’, help=‘The name of the Nutch table where the parsed webpages are saved.’)
# Parse the command line argument.
args = parser.parse_args()

# Generator function that grabs data by chunks of ‘arraysize’.
def ResultIterator(cursor, arraysize=1000):
  ‘An iterator that uses fetchmany to keep memory usage down’
  while True:
    results = cursor.fetchmany(arraysize)
    if not results:
      break
    for result in results:
      yield result
```

```
# Open database connection to the ‘Nutch’ table.
try:
	db = MySQLdb.connect(host="localhost",user="nutchproject",passwd="elliasconstant",db="nutch")
except StandardError as e:
	sys.exit("DB connection failed: %s" % e)

# Prepare two cursor objects (one for getting the data and another to update the data).
cursorSelect = db.cursor()
cursorUpdate = db.cursor()

# Execute SQL query to get the required data to analyse.
sql = "SELECT id, text FROM %s WHERE status = 2 AND text IS NOT NULL AND language IS NULL" % db.escape_string(args.inputTableName)
try:
	cursorSelect.execute(sql)
except StandardError as e:
	db.close()
	sys.exit("An error occurred while querying for the pages to analyse: %s" % e)
# Iterate as long as there is another row to fetch.
for result in ResultIterator(cursorSelect):
	# Encode into UTF8 bytes the text to analyse.
	textBytes = unicode(result[1], "utf-8", "ignore")

	# Detect language of text.
	try:
		isReliable, textBytesFound, details = cld2full.detect(textBytes, isPlainText=True)
	except StandardError as e:
		print "An exception occurred when trying to detect the language. Data id = ‘%s’: %s" % (result[0], e)
```

```
language1, language2 = details[0][0].lower(), details[1][0].lower()
confidence1, confidence2 = int(details[0][2]), int(details[1][2])
score1, score2 = int(details[0][3]), int(details[1][3])

if not (language1 == "english" or language1 == "french"):
lang = "other"
else:
lang = language1
# Sometimes, a page can have both languages or be wrongfully labeled as English.
# In those cases, we rather have the page labeled as French.
if language1 == "english" and language2 == "french":
if confidence1 - confidence2 <= 10 and score1 < score2: # These values were chosen based
on observations.
lang = "french"

# Prepare SQL query to UPDATE the language.
sql = "UPDATE %s SET language = ‘%s’ WHERE id = ‘%s’" %
(db.escape_string(args.inputTableName), lang, db.escape_string(result[0]))

try:
# Execute the SQL command.
cursorUpdate.execute(sql)
# Commit changes in the database.
db.commit()
except StandardError as e:
# Rollback in case there is any error.
db.rollback()
print "An exception occurred when trying to update the language of id = ‘%s’: %s" %
(result[0], e)



# disconnect from server
db.close()
```

ANNEXE C - SÉLECTION DE PAGES PERTINENTES

```
# coding=utf-8

# Author        : Ellias Zainea
#
# This script goes through every english or french page and checks if it is relevant for our
#project. English pages are always relevant in our case. French pages will be relevant only
# if a page with an english translation is not available. This script also extracts and saves the domain of each webpage.
#
# Note that for the moment, we only check if a domain has english pages. If so, all french pages #from this domain are discarded (regardless if it is unique). A more thorough method is
#desirable.
import sys
import argparse
import MySQLdb
from urlparse import urlparse


# Create an argument parser and add an argument to it.
parser = argparse.ArgumentParser(description="Check whether a webpage is relevant or not based on it’s language.")
parser.add_argument(‘inputTableName’, help=‘The name of the Nutch table where the parsed webpages are saved.’)
# Parse the command line argument.
args = parser.parse_args()
```

```
# Generator function that grabs data by chunks of ‘arraysize’.
def ResultIterator(cursor, arraysize=1000):
	# An iterator that uses fetchmany to keep memory usage down
	while True:
		results = cursor.fetchmany(arraysize)
		if not results:
		  break
		for result in results:
		  yield result

# Open database connection.
try:
	db = MySQLdb.connect(host="localhost",user="nutchproject",passwd="elliasconstant",db="nutch")
except StandardError as e:
	sys.exit("DB connection failed: %s" % e)



# Prepare three cursor objects.
cursorSelect = db.cursor()
cursorUpdate = db.cursor()
cursorCheck = db.cursor()
```

```
# Execute the SQL command to get all non relevant english pages.
sql = "SELECT id, baseUrl FROM %s WHERE isRelevant = 0 AND language REGEXP ‘english’" % db.escape_string(args.inputTableName)
try:
	cursorSelect.execute(sql)
except StandardError as e:
	db.close()
	sys.exit("An error occurred while querying for the english pages to analyse: %s" % e)

# Iterate over the english results and set the pages to relevant and grab their domain.
for result in ResultIterator(cursorSelect):
	# Grab the domain.
	parseObj = urlparse(result[1])
	domain = parseObj.netloc
# Update the english pages
	try:
		# UPDATE the isRelevant field of every english page.
		sql = "UPDATE %s SET isRelevant = 1, domain = ‘%s’ WHERE id = ‘%s’" % (db.escape_string(args.inputTableName), db.escape_string(domain), db.escape_string(result[0]))
		cursorUpdate.execute(sql)
		# Commit changes in the database.
		db.commit()

except StandardError as e:
		# Rollback in case there is any error.
		db.rollback()
		print "An exception occurred when trying to update the isRelevant and domain fields of the english page id = ‘%s’: %s" % (result[0], e)
```

```
# Execute SQL query to get all non relevant french pages.
sql = "SELECT id, baseUrl FROM %s WHERE isRelevant = 0 AND language REGEXP ‘french’ AND domain IS NULL" % db.escape_string(args.inputTableName)
try:
	cursorSelect.execute(sql)
except StandardError as e:
	db.close()
	sys.exit("An error occurred while querying for the french pages to analyse: %s" % e)

# Iterate as long as there is another row to fetch.
for result in ResultIterator(cursorSelect):
	# Compare baseUrl of french pages to check if their is an equivalent in english.
	# For our needs, if there is a single english page in the DB from the same domain web site,
	# we conclude that the web site has a english translation and so the french page is not needed.
	parseObj = urlparse(result[1])
	domain = parseObj.netloc
	try:
		cursorCheck.execute("SELECT COUNT(*) FROM %s WHERE domain = ‘%s’ AND language REGEXP ‘english’" % (db.escape_string(args.inputTableName), db.escape_string(domain)))
	except StandardError as e:
		db.close()
		sys.exit("An exception occurred when trying get the number of english rows with a domain of ‘%s’: " % (domain, e))
	count = cursorCheck.fetchone()[0]

# If their is not an english equivalent, then the french page is relevant.
	if count == 0:
```

```
# Prepare SQL query to UPDATE the isRelevant field and set their domain.
        sql = "UPDATE %s SET isRelevant = 1, domain = ‘%s’ WHERE id = ‘%s’" % (db.escape_string(args.inputTableName), db.escape_string(domain), db.escape_string(result[0]))
        try:
            # Execute the SQL command.
            cursorUpdate.execute(sql)
            # Commit changes in the database.
            db.commit()
        except StandardError as e:
            # Rollback in case there is any error.
            db.rollback()
            print "An exception occurred when trying to update the isRelevant and domain fields of the french page id = ‘%s’: %s" % (result[0], e)
    # If their is an english equivalent, then just update the page domain.
    else:
        sql = "UPDATE %s SET domain = ‘%s’ WHERE id = ‘%s’" % (db.escape_string(args.inputTableName), db.escape_string(domain), db.escape_string(result[0]))
        try:
            # Execute the SQL command.
            cursorUpdate.execute(sql)
            # Commit changes in the database.
            db.commit()
        except StandardError as e:
            # Rollback in case there is any error.
            db.rollback()
            print "An exception occurred when trying to update the domain field of the french page (non relevant) id = ‘%s’: %s" % (result[0], e)
# disconnect from server
db.close()
```

ANNEXE D –STRATÉGIE DE RECHERCHE DE MOTS CLÉS DE LA NANOTECHNOLOGIE

(nano-?)(biotechnology|replication|cantelever|encapsulation|container|medicine|complex|prototyping|fabrication|lithography|machining|manufacturing|imprint lithography|mechanic|electronic|device|crystal memory|robotic|switch|computing|transisator|material|composite|film|particle|structure|powder|coating|biocomposite|tube)\w*|(bio-?nano)\w*|(nano-?bio)\w*|(biological anosensor)|(nanobiomagnetic)|(targeted nano-therapeutic)|(biosensor)|(molecular motor)|(biomolecular fabric)|(engineered enzyme)|(engineered protein)|(scanning tunneling icroscopy)|(atomic force microscopy)|(scanning probe microscopy)|(electron-beam lithography)|(self-assembly)|(directed assembly)|(afm)|(focu ion beam)|(fib)|(molecular electronic)|(nems)|(nanoscale magnetic)|(fullerenes?)|(quantum tubes?)|(sol gel)|(graphite)|(growth methode)|(graphene)|(quantum dots?)

ANNEXE E – STRATÉGIE DE RECHERCHE PAR INDICATEUR

Recherche et développement

\b (research and development)|(r&d)|(laborator(y|ie)|(researcher)|(scientist)|(product development)|(technology development)|(development phase)|(technical development)|(development program)|(development process)|(development project)|(development center)|(development facilit(y|ies))|(technological development)|(development effort)|(development cycle)|(development research)|(research & development)|(development activit(y|ies))|((basic|fundamental) research) \b

Collaboration simple

\b((collaborat\w*)|(cooperati\w*)|(alliance\w*)|(partnership\w*))\b

Collaboration universitaire

\b((universit(és?|ies|y)|(academi\w*)|(student))\W+(?:\w+\W+){1,6}?((collaborati\w*)|(cooperati\w*)|(alliance\w*)|(partnership\w*)))|(((collaborati\w*)|(cooperati\w*)|(alliance\w*)|(partnership\w*))\W+(?:\w+\W+){1,6}?(universit(és?|ies|y)|(academi\w*)|(student)))\b

Financement

\b (atlantic canada opportunities)|(business development bank of canada)|(sustainaible development technology)|(venture capital)|(atlantic innovation fund)|(nrc-irap)|(fednor)|(industrial research assistance program)|(public fund\w*)|(joint venture)|(grant)\b

Produit

\b((([A-Z-a-z-]{1,}))(\w{1,})\ ?[™®])\b

Propriété intellectuelle

(intellectual propert(ie|y))|(patent\w*)|(trade secret)

ANNEXE F - SITE D'ENTREPRISES CANADIENNES CLASSÉES PAR DOMAINE

- Métrologie et nanoprocessus :

fibics.com
www.apcfilters.com
www.appliednt.com
www.cmc.ca
www.designw.com
www.edgescientific.com
www.fibics.com
www.hitachi.ca
www.icspicorp.com
www.meaglow.com
www.microbonds.com
www.nanosecurity.ca
www.quartzimaging.com

- Nanodispositifs et nanoélectronique :

skynanotech.com
www.acamp.ca
www.aurorand.com
www.celluforce.com
www.dwavesys.com
www.elevenengineering.com
www.exfo.com
www.ocivm.com
www.skynanotech.com

- Nanomédecine et nanobiotechnologie :

greencorenfc.com
www.axela.com
www.biomedcentral.com
www.biophagepharma.net
www.chemcomp.com

www.chemroutes.com
www.epocal.com
www.flogen.com
www.micralyne.com
www.nanostruck.com
www.nicoyalife.com
www.scimedtechnologies.com
www.supratek.com

- Nanostructures et nanomatériaux :

arciglobal.com
eastgatepharmaceuticals.com
fimatcon.com
indeedlabs.com
intelligentnano.com
kematrix.com
ufluidix.com
www.acrolab.com
www.adsdyes.com
www.am-m.com
www.arkema.com
www.barrday.com
www.bcri.ca
www.cancarb.com
www.celgenecanada.net
www.cg2nanocoatings.com
www.cifmetal.com
www.cvmr.ca
www.cyriumtechnologies.com
www.cytodiagnostics.com
www.design1st.com
www.eaglegraphite.com
www.gensnano.com
www.grafoid.com
www.hy-powernano.com
www.ibioceramix.com
www.ignisinnovation.com
www.indigo.com
www.innovotech.ca
www.integran.com
www.intertek.com
www.itlinc.com

www.lightsource.ca
www.lv-em.com
www.mac.ca
www.mchnanosolutions.com
www.mdscoating.com
www.mdsprad.ca
www.mettech.com
www.mkplastics.com
www.mpb-technologies.ca
www.nanalysis.com
www.nanomed.ca
www.nanoxplore.ca
www.nucryst.com
www.ondinebio.com
www.open-nd.ca
www.patt-technologies.com
www.perkinelmer.com
www.petrosep.com
www.photonetc.com
www.plasmionique.com
www.polymersciencecorp.com
www.polymos.com
www.puroxi.com
www.quantiam.com
www.sealweld.com
www.sentec.ca
www.sjpc.com
www.targray.com
www.vivecrop.com
www.xrail.com