

# Sommaire

## Inhoudstafel

62<sup>ème</sup> année - 2008 - n° 1

62ste jaargang - 2008 - nr 1

- Éditorial – Woord vooraf 3
- Outil d'extraction automatisée et en temps réel d'une bibliographie 4  
Frédérique de Ruiter, Arie de Ruiter, Charles-Victor Boutet et Luc Quoniam
- Van folksonomiën naar ontologieën 12  
Céline Van Damme
- Donner du sens aux brevets 18  
Typologie des utilisations de l'information brevets  
Fabienne Monfort-Windels
- Het Nederlandse programma DARE 2003>06 23  
Leo Waaijers
- Notes de lecture – Boekbesprekingen 34
- Nouvelles parutions – Nieuwe publicaties 36
- Regards sur la presse – Een blik op de pers 40
- Index 2007 54



par  
door

Simone JÉRÔME

traduit par  
vertaald door

Dominique VANPÉE

"Je suis venu te dire que je m'en vais"... C'est fini. Je ne serai plus rédacteur-en-chef des *Cahiers de la Documentation*. Désormais, mon collègue Guy Delsaut sera seul responsable de la ligne éditoriale du périodique de notre association. Contrairement à la chanson, mon départ ne se fait pas dans les larmes mais au contraire avec le sentiment du devoir accompli. Avec Guy, dont je souligne le remarquable travail depuis le jour où il a intégré l'équipe, nous avons pu donner à la publication un contenu de qualité et une nouvelle présentation. C'est donc sans regret et avec confiance que je lui cède définitivement les rênes d'un attelage que nous conduisions à deux depuis quelques numéros.

Entrée à l'ABD en 1973, j'ai été active pendant cette période des années 70, qui a vu l'éclosion des technologies numériques dans le champ de la documentation. J'ai eu la chance de travailler à cette époque avec de fortes personnalités dont plusieurs, hélas, nous ont déjà quittés. Après une longue période consacrée avant tout au travail dans une unité de documentation académique qui prenait d'année en année plus d'importance, je suis entrée au conseil d'administration de l'ABD au moment où Internet venait bouleverser les habitudes de travail des professionnels de la documentation. Ce fut à nouveau une période de découvertes, de diversification des orientations et des techniques ainsi que de contacts humains extrêmement enrichissants. Merci à mes collègues du conseil d'administration et à tous ceux qui ont compté pour moi pendant ces années qui furent aussi les dernières de ma vie professionnelle.

Mais ce n'est pas encore un adieu définitif, je reste membre de l'ABD et membre de l'équipe de rédaction des *Cahiers*. Nos futurs auteurs risquent encore fort de me rencontrer au tournant d'un texte parsemé de remarques de style car, pour moi, le langage reste la pierre angulaire de tout travail de documentation. Que les nouvelles technologies qui ouvrent l'espace et dévorent le temps, ne nous fassent pas oublier dans leur tourbillon que l'essentiel reste le message et que celui-ci passe d'abord par le langage.

"Je suis venu te dire que je m'en vais" [Serge Gainsbourg - "Ik kwam je zeggen dat ik je verlaat"] Het liedje is uit(gezongen). Ik zal niet langer meer hoofdredacteur zijn van de *Bladen voor documentatie*. Voortaan zal mijn collega Guy Delsaut alleen verantwoordelijk zijn voor de editoriale lijn van het tijdschrift van onze vereniging. In tegenstelling tot het liedje zal mijn vertrek niet met tranen gepaard gaan maar integendeel met het gevoel van de vervulde plicht. Guy en ikzelf, waarvan ik het voortreffelijke werk wens te onderstrepen vanaf de eerste dag dat hij geïntegreerd werd in de ploeg, hebben de publicatie een meer kwaliteitsvolle inhoud en een nieuwe look gegeven. Het is dan ook zonder spijt en met vertrouwen dat ik hem definitief de teugels van het span - dat wij samen sinds enige nummers bestuurden - overdraag.

Ik sloot mij aan bij de BVD in 1973 en was gedurende deze periode van de jaren zeventig actief waarin we het ontluiken meemaakten van de digitale technologieën in het documentatielandchap. Ik had het geluk om in die tijd te werken met sterke persoonlijkheden, waarvan meerderen ons helaas al verlaten hebben. Na een lange periode vóór alles gewijd aan het werk in een academische documentatie-eenheid die met de jaren steeds aan belang won, ben ik in de raad van bestuur van de BVD gekomen op het moment dat het Internet de werkgewoonten der professionelen uit de documentatie kwam overhoopgooien. Dit was opnieuw een periode van ontdekkingen, van diversificatie in oriëntaties en technieken en van menselijke contacten die extreem verrijkend waren. Met dank aan mijn collega's van de raad van bestuur en aan allen die voor mij meetelden gedurende deze jaren die tevens de laatste van mijn professioneel leven waren.

Maar het is nog geen definitief vaarwel. Ik blijf lid van de BVD en lid van de redactie van de *Bladen*. Onze toekomstige auteurs lopen nog sterk het gevaar me tegen te komen rond de wending van een tekst doorspekt met opmerkingen rond de schrijfstijl. Want voor mij blijft de taal de hoeksteen van het documentatiewerk. Dat de nieuwe technologieën - die ruimte openen en tijd verslinden - ons niet laten vergeten in hun maalstroom dat het essentiële de boodschap blijft en dat deze vooreerst via de taal komt.

---

# OUTIL D'EXTRACTION AUTOMATISÉE ET EN TEMPS RÉEL D'UNE BIBLIOGRAPHIE

---

**Frédérique de RUITER**

Master 2 Intelligence économique et territoriale

**Arie de RUITER**

Architecte de systèmes d'information

**Charles-Victor BOUTET**

Master 2 Dispositifs sociotechniques de l'information et de la communication, Université de Toulon – Institut Ingemedia

**Luc QUONIAM**

Professeur, Université de Toulon – Institut Ingemedia

▪ Nous présentons dans les pages qui suivent un outil d'extraction automatisée de toutes les références bibliographiques basée sur une sélection de mots-clés et qui fonctionne à partir de la plupart des sites nationaux d'Amazon à travers le monde. Cet outil qui permet de paramétrer en amont et en temps réel l'extraction de l'information, le tri ainsi que son classement sur une base de données exploitable suivant les besoins de l'utilisateur se caractérise par l'absence de bruit dans les résultats obtenus du fait du choix des sites source.

▪ Wij beschrijven in de hierop volgende bladzijden een geautomatiseerd extractie-instrument voor alle bibliografische referenties gebaseerd op een trefwoordselectie en dat werkt vanaf het grootste deel van de nationale sites van Amazon in de wereld. Dit gereedschap dat toelaat om het voorafgaande stadium en in real time de informatie-extractie, de selectie alsook zijn klassemment binnen een bruikbare gegevensbank volgens de gebruikersbehoeften te parametriseren karakteriseert zich door de afwezigheid van ruis binnen de bekomen resultaten dank zij de keuze van de bronsites.

Le XXI<sup>ème</sup> siècle consacre l'âge de l'information marqué par la turbulence, une accélération du temps dans tous les domaines, grâce à des technologies toujours plus performantes, devenues accessibles à tous et qui se traduisent par des délais de réaction de plus en plus courts. Les entreprises qui se maintiennent et accroissent leurs performances sont celles qui savent être à l'écoute systématique de leur environnement. Mais, l'information étant devenue tellement accessible qu'elle en devient surabondante, l'important est lié à l'usage qui en est fait, grâce au traitement qui lui donne du sens et permet son actualisation.

Comme l'a souligné Herbert Simon<sup>1</sup> "Les systèmes de traitement de l'information de notre monde contemporain baignent dans une abondance excessive d'informations et de symboles. Dans un tel monde, la ressource rare n'est pas l'information, mais la capacité de traitement pour s'occuper de cette information". On mesure ainsi l'importance de la capacité de traitement, particulièrement pour une structure de veille dont l'objectif principal est l'aide à la décision.

Dans le cadre de ce travail, nous considérerons la veille comme "un processus de collecte de données pertinentes dont l'objectif est d'aboutir à une prise de décision" [Antonio Da Silva, 2002]. Pour ce faire, il faut disposer d'une bonne méthode de sélection des informations, sans

laquelle "il n'est pas de veille stratégique viable" [LESCA, 1998].

Il convient ainsi de disposer de la méthode et de l'outil pertinents. Comme toute collecte d'informations, elle requiert pour être efficace, une stratégie de recherche. En effet pour éviter les interminables itérations de la recherche traditionnelle de bibliographie, il importe de réfléchir à une méthode de recherche qui soit rapide et exhaustive.

L'objectif de notre travail consistera ainsi à définir la bonne stratégie et l'outil le plus pertinent pour l'acquisition d'une base bibliographique actualisée qui soit susceptible d'être intégrée dans une base de données Excel aux fins de son exploitation adéquate.

Ainsi, l'outil "Macro Recherche" permet de lancer une extraction automatique de références d'ouvrages pertinents sélectionnés à partir de mots clés correspondant à tout sujet possible de recherche et qui donne en résultat une conséquente base bibliographique disponible sur le marché. Cet outil offre également la capacité de filtrer, par exemple par l'année de publication. Il est naturellement possible de filtrer les doublons émanant des publications nationales de chaque pays dans lequel Amazon possède un website, ainsi que par N° ISBN, par titre, mots dans le titre, auteur, éditeur etc. Ainsi il devient aisé pour un chercheur de trouver rapidement plus

d'ouvrages qui ont été publiés sur un sujet donné que dans une librairie. L'outil "Macro Recherche" lui donne en temps réel des réponses lorsqu'il souhaite approfondir un sujet d'études ou en appréhender les généralités. Grâce aux requêtes il a été possible d'extraire à la fois les documents primaires (les références sur l'ouvrage) et les documents secondaires notamment les résumés des ouvrages.

Nous exposerons dans les pages qui suivent la restitution de l'expérience et une application de cet outil à la recherche bibliographique sur le thème de l'Intelligence Économique.

## Cadrage théorique

### Analyse des besoins

Diverses études sur l'analyse des comportements des internautes lorsqu'ils effectuent une recherche sur les moteurs ont montré une tendance à s'arrêter sur les premiers résultats ou sur les premières pages<sup>2</sup>.

Cela pose un réel problème vu le nombre relativement élevé des pages renvoyées par les moteurs de recherche. A fortiori quand certaines études révèlent l'absence de pertinence fréquente des premiers résultats du fait de la présence de liens sponsorisés<sup>3</sup>.

Concevoir un outil permettant de lever cet écueil afin d'exploiter au mieux les requêtes faites à travers les moteurs de recherche demeure important.

Étant entendu que l'objectif est de fournir la bonne information, celle qui répond aux besoins du décideur, une surveillance systématique des publications sur un domaine déterminé semble d'un intérêt stratégique pour une structure de veille.

En effet, une revue de bibliographie est nécessaire pour à la fois l'information spécialisée fournie par les monographies mais aussi pour l'approfondissement de certaines questions qu'elle rend aisé. Elle permet également de repérer les auteurs importants, les éditeurs, les titres d'ouvrages et éventuellement les sites web à connaître afin de suivre un domaine.

La notion de bibliographie peut recouvrir diverses acceptions mais nous retiendrons la suivante "une liste de références ou de notices bibliographiques classées selon certains critères pour permettre le repérage des documents référencés" [Beaudiquez, 1989].

La méthode traditionnelle que pratiquent souvent étudiants et chercheurs rompus à la recherche d'auteurs de référence est l'interrogation de bases de données de bibliothèques, la consultation éventuelle de renvois bibliographiques en fin d'article ou d'ouvrage.

Cette recherche traditionnelle requiert de connaître les caractéristiques des langages utilisés par les spécialistes de la documentation pour décrire le contenu des documents scientifiques, qu'il s'agisse de livres, de chapitres d'ouvrages, d'articles de périodiques, etc. [Piolat, Annie, 2002].

Le travail a consisté à définir la manière de lever l'écueil de la recherche manuelle de documents dans un contexte de surcharge de l'information et de l'absence de normalisation de l'indexation par les principaux moteurs de recherche? La connaissance de la description des produits au niveau d'Amazon, a permis d'établir les requêtes de la macro. L'organisation des résultats obtenus au sein d'une feuille permettra leur intégration dans une base de données Wini-sis<sup>4</sup>.

### Le choix de la source d'information

Deux raisons principales justifient le choix porté sur Amazon :

Amazon offre la possibilité de lancer simultanément la recherche sur tous ses sites (en France, aux États-Unis, au Canada, en Grande Bretagne, en Belgique, en Allemagne, en Chine etc. et de pouvoir commander immédiatement un choix d'ouvrages reçus trois à cinq jours plus tard.

En outre, La présentation des pages est similaire sur chacun des différents sites nationaux. Parfois, en fonction des sites, la structure en ligne (Internet) peut être très variable, ce qui complique d'autant les méthodes de recherche automatiques et simultanées sur plusieurs sources. La similarité de présentation des pages sur Amazon permet de pallier ce problème.

## Résultats

Pour les raisons évoquées ci-dessus nous avons choisi d'établir une extraction bibliographique à partir des différents sites web nationaux d'Amazon. En effet, Amazon fournit des "Services Web" ("Amazon Web Services")<sup>5</sup> qui permettent d'effectuer des recherches automatisées directement dans leurs bases de données. Par ailleurs, ces services nécessitent une connaissance étendue des outils informatiques spécifiques. C'est pourquoi nous avons choisi d'utiliser un

fichier Excel, comportant une macro en Visual Basic.

### Le fichier Excel

Afin de garantir un bon fonctionnement de la macro, notre fichier Excel comporte un nombre de feuilles. Vous trouverez dans le tableau ci-dessous les noms de ces différentes feuilles et la description de leurs objectifs.

#### Définition des mots clés

Afin d'alimenter la feuille "Searchstring", nous avons choisi un ensemble de vingt sept mots-clés, principalement à l'aide du référentiel de formation en intelligence économique en France

Nom	Utilisation
<b>SearchString</b>	Contient les mots clés pour lesquels l'extraction est à effectuer.
<b>Exclude</b>	critère d'exclusion : les URL contenant la/les chaîne/s présente/s dans le champ "Exclude" ne seront pas prises en compte.
<b>Sites</b>	Contient les différents sites d'Amazon à interroger ainsi que des informations complémentaires en différentes langues, afin d'extraire les détails d'un livre (voir paragraphe 2.5)
<b>Result</b>	Stockage de tous les résultats de l'extraction
<b>Web</b>	Stockage temporaire du résultat d'une requête Google
<b>Content</b>	Stockage temporaire d'une page détaillée d'un livre sur Amazon
<b>Duplicate</b>	Stockage temporaire des N° ISBN déjà trouvés pendant l'extraction

<[http://ocsima.neuf.fr/Referentiel\\_formation\\_commission\\_Juillet.pdf](http://ocsima.neuf.fr/Referentiel_formation_commission_Juillet.pdf)> (consulté le 29/01/2008).

L'ensemble de mots-clés est un mélange de mots français et anglais sur le thème de l'Intelligence Économique.

Toutefois, notons qu'un chercheur sélectionnera ses propres mots-clés au fur et à mesure et ne lancera que rarement une requête simultanée de 27 mots-clés, surtout s'il souhaite obtenir des résultats quasi-immédiats.

### La stratégie de recherche : la requête sur Google

Afin de trouver les pages sur les sites web d'Amazon correspondant à un mot clé choisi (ou combinaison de mots clés), la macro effectue d'abord une requête sur Google, pour les raisons techniques suivantes :

- Pour lancer une requête de recherche des livres directement sur

Amazon, la composition de l'URL n'est pas "standardisée"

- Les résultats d'une recherche sur Amazon ne contiennent pas l'URL de la page détaillée d'un livre spécifique (bien sûr il existe un lien avec la page détaillée, mais comme le résultat est importé en texte brut dans la feuille Excel, il n'est pas possible de le récupérer).

Afin de retrouver toutes les pages référencées par Google sur le site de **Amazon.com** pour obtenir, par exemple, la combinaison "**competitive intelligence**" dans la barre des titres, on lance la requête suivante :

- **site:www.amazon.com intitle: "competitive intelligence"**

Nous avons choisi de ne retrouver que les pages contenant les mots clés sélectionnés dans leurs titres. Si l'on souhaite obtenir chacune des pages

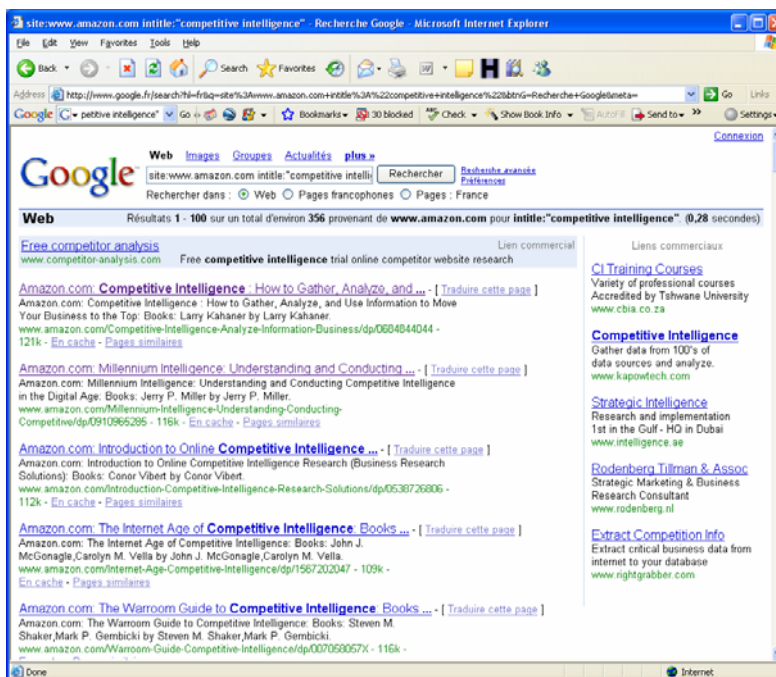


Fig. 1 : Translittération de la requête "compétitive intelligence" Macro/Google.

contenant n'importe où ces mots clés, il suffit de supprimer le mot "intitle:" de la requête. La figure 1 montre le résultat de notre requête.

- Premièrement, ce qui est intéressant dans ce résultat est la traduction par Google de notre requête en l'URL suivant :

<http://www.google.fr/search?hl=fr&q=site%3Awww.amazon.com+intitle%3A%22competitive+intelligence%22&btnG=Recherche+Google&meta=>

(voir la barre "Address" dans figure 1), où on retrouve bien notre requête lancée sur Google (avec un codage de certains caractères : → %3A et " → %22) :

- Deuxièmement, les résultats d'une requête Google consistent toujours en trois rubriques distinctes :

- Le titre de la page référencée
- Le contexte (la phrase) dans lequel les mots clés se retrouvent sur cette page
- L'URL de cette page

Le fonctionnement de la macro en Visual Basic dans le fichier Excel

La base de l'extraction automatisée est composée des cellules de la première colonne de la feuille "SearchString" (à partir de la deuxième ligne) contenant notre sélection de mot-clés per-

tinents pour la recherche, comme le montre la figure 2.

Pour chaque cellule remplie de la feuille "SearchString", la macro compose un URL de la façon précédemment indiquée dans le paragraphe 2.3. puis établit une connexion en utilisant cet URL sur Internet, et enregistre les résultats obtenus dans la feuille "Web" de notre fichier Excel. La figure 3 montre le résultat dans la feuille "Web" pour la première combinaison de mots clés "competitive intelligence" de notre exemple<sup>6</sup>.

Dans cette feuille on peut voir que chaque résultat de la recherche Google consiste en 3 lignes et

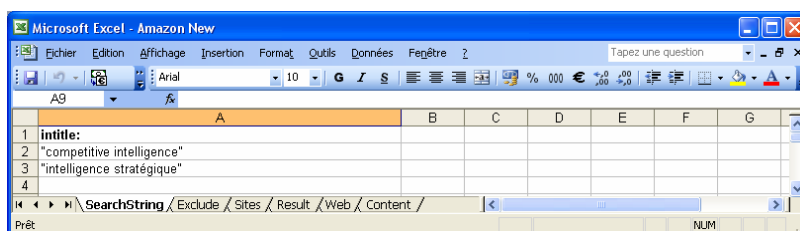


Fig 2 : Structure des feuilles Excel de la Macro - les mots-clés.

que la troisième ligne contient l'URL de la page détaillée du produit<sup>7</sup> sur Amazon.com.

De plus, lorsqu'une ligne commençant avec le texte "Page de résultats :" contient le mot "Suivant" dans une des lignes de la même colonne, cela signifie qu'une prochaine page de résultats de Google est à traiter.

Pour chaque résultat de la recherche Google la macro établit encore une connexion Internet en utilisant l'URL trouvée dans la troisième ligne du résultat et enregistre les résultats dans la feuille

"Content" de notre fichier Excel. La figure 4 montre le résultat de la page détaillée d'Amazon dans la feuille "Content" pour le deuxième résultat Google (Figure 3).

Chaque page de présentation d'un ouvrage sur Amazon, contient différentes rubriques dont il a fallu tenir compte pour mieux extraire les informations intéressantes dans le cadre de notre travail.

Pour ce faire, l'extraction suit la logique suivante<sup>8</sup> :

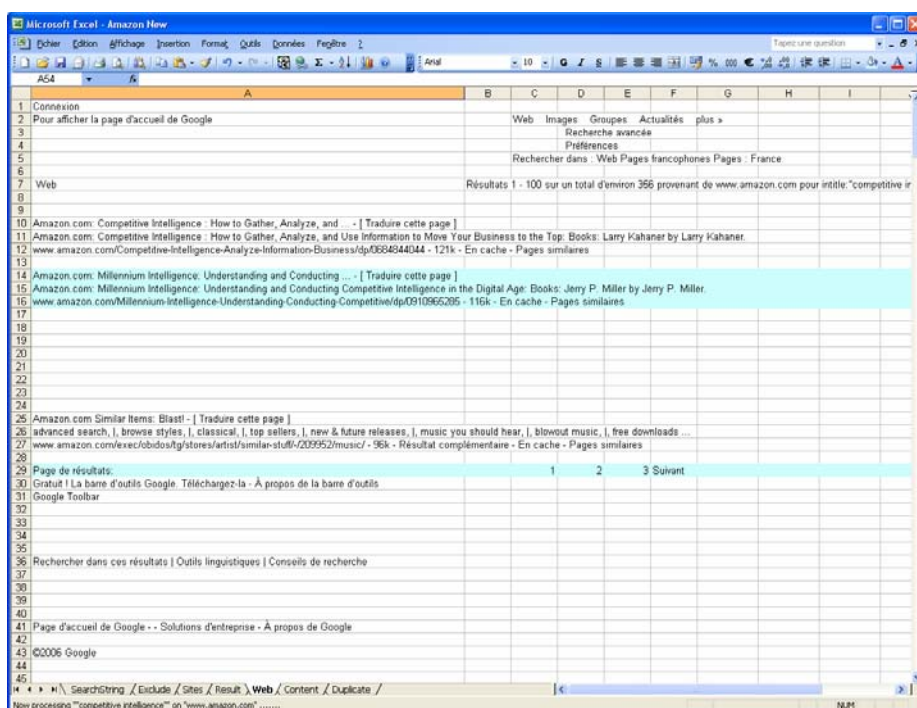


Fig 3 : Résultats de recherche bruts.

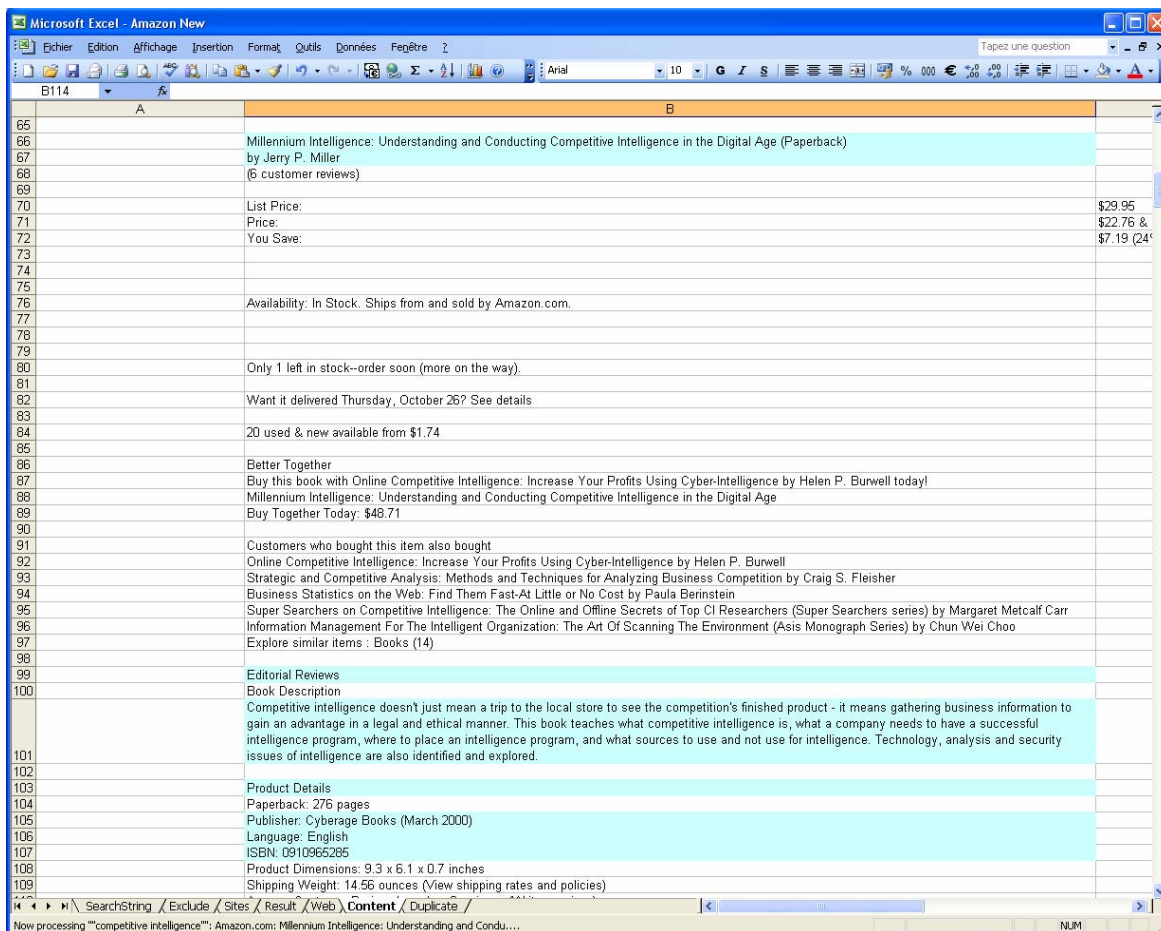


Fig 4 : Page produit Amazon sous Excel.

Afin de trouver la colonne contenant les détails sur le produit, la macro cherche le texte "Product Details".

Elle récupère ainsi l'éditeur (cellule commençant avec "Publisher: "), l'année de publication (les 4 chiffres avant la fin de l'éditeur), la langue ("Language:"), l'ISBN, ou l'ASIN, (respectivement "ISBN:" ou "ASIN:"), l'auteur ("by") et le titre (la ligne précédente celle de l'auteur).

Pour trouver une description du produit, la macro recherche le texte "Editorial Reviews" dans les cellules de la même colonne que les autres données. Dans le but de retrouver une description "utile" la macro interroge les cellules des 6 lignes suivantes, établit la longueur du contenu de chacune, et choisit le contenu le plus long comme description du produit.

### La recherche Multi-sites Amazon

Afin de chercher des produits sur les différents sites nationaux d'Amazon, la macro utilise des

informations stockées dans la feuille "Sites" du fichier Excel (figure 5) au sein de laquelle on retrouve les différents sites disponibles dans le monde et ses correspondances ou traductions de textes dans le langage national du site, utilisés pour l'extraction des informations, comme décrit dans le paragraphe précédent.

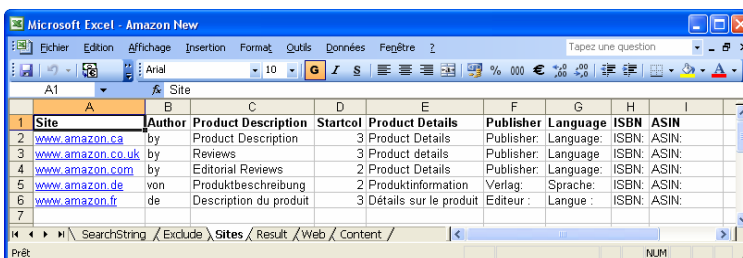


Fig 5 : Structure des feuilles Excel de la Macro – la base des sites explorés.

L'exclusion de certains résultats de la recherche Google

L'expérimentation sur les résultats obtenus de Google a montré qu'il existe des résultats dans lesquels l'URL d'Amazon renvoie à une page du website Amazon s'il contient des informations complémentaires aux détails donnés sur le pro-



duit spécifique. Afin d'exclure ces URLs de l'extraction, la macro vérifie si l'URL du résultat Google contient un des textes contenus dans la feuille "Exclude" du fichier Excel. La Figure 6 montre par exemple les différents textes établis pendant notre expérimentation.

### La signalisation de doublons

Le doublon est la double occurrence d'un ouvrage identifiable par son ISBN ou ASIN. Afin d'indiquer si un produit a déjà été trouvé auparavant, la macro stocke les ISBN/ASIN's dans une feuille temporaire "Duplicate" du fichier Excel. Cette feuille est créée dynamiquement pendant l'extraction et est supprimée à la fin.

Les résultats de la feuille "Result" contiennent une colonne "Duplicate" qui aura une valeur "No" pour la première occurrence de l'ISBN/ASIN et "Yes" pour chaque récurrence. En effet, il arrive souvent qu'un même ouvrage soit commercialisé simultanément sur Amazon.co.uk, Amazon.com et Amazon.ca (...)

### Résultats

L'utilisation de la macro à partir d'une base de recherche sur 27 mots clés spécifiques à l'intelligence économique a généré 2291 résultats (1090 résultats uniques) avec un temps d'extraction de 3 heures et 40 minutes. La figure 7 montre un extrait de ces résultats, dans lesquels nous retrouvons 3 occurrences du livre qui nous a servi d'exemple à travers cette présentation.

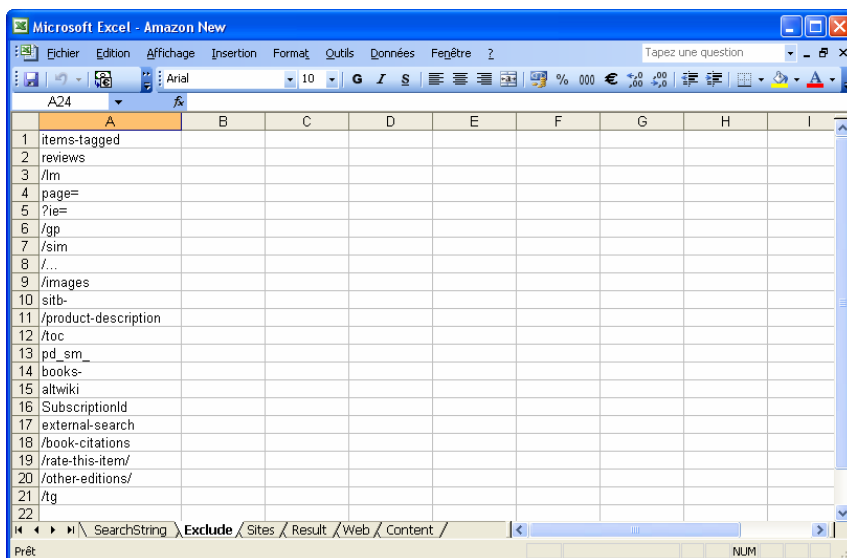


Fig 6 : Structure des feuilles Excel de la Macro – les mots-clés exclus de la recherche.

SearchString	Duplicate	ISBN/ASIN	Title	Author	Publisher	Language	Year	URL
"competitive intelligence"	No	0699303739	A New Archetype for Competitive Intelligence (Hardcover)	John J. McGonagle (Author), Carolyn M. Vella (Author)	Quorum Books (June 1996)	English	1996	www.amazon.ca/gp
"competitive intelligence"	Yes	0699303739	A New Archetype for Competitive Intelligence (Hardcover)	John J. McGonagle, Carolyn M. Vella, et John J. McGonagle	Greenwood Press (30 Jun 1996)	English	1996	www.amazon.co.uk
"competitive intelligence"	Yes	0699303739	A New Archetype for Competitive Intelligence (Hardcover)	John J. McGonagle, Carolyn M. Vella	Quorum Books (June 30, 1996)	English	1996	www.amazon.com
"competitive intelligence"	Yes	0699303739	A New Archetype for Competitive Intelligence (Gebundene Ausgabe)	John J. McGonagle, Carolyn M. Vella	Greenwood Press (30. Juni 1996)	English	1996	www.amazon.de/gp
"competitive intelligence"	Yes	0699303739	A New Archetype for Competitive Intelligence (Relié)	John J. McGonagle, Carolyn M. Vella	Quorum Books (Jui 1996)	English	1996	www.amazon.fr/gp
"competitive intelligence"	No	0910965205	Millennium Intelligence: Understanding and Conducting Competitive Intelligence in the Digital Age (Paperback)	Jerry P. Miller (Author)	CyberAge Books & Information Today, Inc. (Jan 23 2003)	English	2003	www.amazon.ca/gp
"competitive intelligence"	Yes	0910965205	Millennium Intelligence: Understanding and Conducting Competitive Intelligence in the Digital Age (Paperback)	Jerry Miller (Editor)	CyberAge Books (Mar 2000)	English	2000	www.amazon.co.uk/Competitive/dp/0910965205
"competitive intelligence"	Yes	0910965205	Millennium Intelligence: Understanding and Conducting Competitive Intelligence in the Digital Age (Paperback)	Jerry P. Miller	Cyberage Books (March 2000)	English	2000	www.amazon.com/Competitive/dp/0910965205
"competitive intelligence"	No	0910965641	Super Searchers on Competitive Intelligence: The Online and Offline Secrets of Top CI Researchers (Paperback)	foundng director, Motorola's intellige, Jan Hering (Foreword), Margaret Metcalf Carr (Author), Reva Basch (Éditeur)	Information Today, Inc. (Jun 26 2003)	English	2003	www.amazon.ca/gp
"competitive intelligence"	Yes	0910965641	Super Searchers on Competitive Intelligence: The Online and Offline Secrets of Top CI Researchers (Super Searchers series) (Paperback)	Margaret Metcalf Carr, founding director, Motorola's intellige, Jan Hering (Foreword), Reva Basch (Éditeur)	Information Today, Inc. (June 1, 2003)	English	2003	www.amazon.com/Sup Searchers/dp/0910965641
"competitive intelligence"	Yes	0910965641	Super Searchers on Competitive Intelligence: The Online and Offline Secrets of Top CI Researchers (Broché)	Jan P. Hering (Préface), Margaret Metcalf Carr, Reva Basch (Pour la direction de)	Cyberage Books (Jui 2003)	English	2003	www.amazon.fr/gp/Sup Searchers/dp/0910965641

Fig 7 : Structure des feuilles Excel de la Macro – les résultats de recherche raffinés.

### Conclusion

Le travail effectué à l'aide de l'outil "Macro Recherche" développé dans Excel nous a permis d'extraire un nombre considérable d'ouvrages. La macro balayant les différents sites Amazon, pourrait être assimilée à un méta moteur de recherche bibliographique. Cette caractéristique est à notre sens son principal intérêt. En effet, cet outil permet pour une même requête, d'interroger plusieurs moteurs de façon simultanée, de rapatrier les résultats, les synthétiser et même

proposer un récapitulatif des réponses données. En ce sens, c'est un outil qui, vu sa relative facilité de fabrication est d'une utilité essentielle pour l'usage d'une petite structure de recherche. En effet, l'intérêt de cet outil est justement le fait qu'il est aisé de bénéficier des puissantes fonctionnalités du logiciel Excel<sup>9</sup> pour faire du tri, du filtrage, retrouver l'occurrence de certains mots clés et exporter aisément de ces résultats dans d'autres types de bases de données pour divers types de traitements. Toutefois, le premier biais qui apparaît est le nombre considérable de doublons obtenus, en réalité, autant de doublons que d'ouvrages recensés. Ceci s'explique par le fait que la recherche a été réalisée effectivement de manière simultanée sur de multiples sites nationaux. Cependant, la sélection par le N° ISBN a permis de faire le tri nécessaire et d'éliminer les doublons. L'utilisation de cet outil à l'aide de termes généraux risque de générer beaucoup d'informations non pertinentes. Il importe de souligner qu'une recherche plus ciblée sur un nombre limité de mots-clés donne naturellement des résultats plus pertinents. La macro est donc spécialement adaptée à une recherche ponctuelle avec une stratégie bien définie au

préalable, avec un objectif de recherche spécifié pour un gain de temps non négligeable. Il importe de remarquer que l'utilisation de cet outil est efficace dans le cadre de recherches sur des sujets très pointus où l'information est plus rare.

**Frédérique de Ruiter**

**Arie de Ruiter**

55 rue Ravel

83500 La Seyne/Mer

France

frederique.deruiter@free.fr

arie.deruiter@free.fr

**Charles-Victor Boutet**

**Luc Quoniam**

*Institut Ingemedia*

*Université de Toulon*

Avenue de l'Université - BP20132

83957 La Garde Cedex

France

charles-victor-boutet@univ-tln.fr

quoniam@univ-tln.fr

1<sup>er</sup> Octobre 2007

L'outil d'extraction décrit dans cet article est disponible à l'adresse suivante : <http://www.tapalair.net/macro.xls>.

## Bibliographie

Beaudiquez, M. *Guide Bibliographie générale. Méthodologie et Pratique*. Paris : Saur, 1989, nouv. éd., 277p.

Blanco, S. Sélection de l'information à caractère anticipatif : un processus d'intelligence collective. *Actes de la XI ième Conférence de l'AIMS*. 2002, 5-7 juin, Paris, pp. 1-20.

Dou, H. ; Hassanaly, P. ; Quoniam, L. ; Latela, A. Veille technologique et information documentaire : de l'usage de la bibliométrie dans les services documentaires. *Documentaliste-science de l'information*, 1990, vol. 27, n° 3, pp.132-141.

Lefèvre, Ph. *La Recherche d'informations*. Paris : Hermès, 2000, 253 p.

Léon, A. Savoir chercher et interroger : les repères méthodologiques. *Ressources électroniques pour les étudiants, la recherche et l'enseignement*, Formist, 2006, pp 59-61.

Lesca, H. Veille stratégique : comment sélectionner les informations pertinentes ? Concepts, méthodologie, expérimentation, résultats. *Conférence internationale de management stratégique*. Lille, 1996, 13-15 mai, pp. 161-162.

Lesca, H.; Schuler, M. Veille stratégique : comment ne pas être noyé sous les informations. In *Economies et sociétés, sciences de gestion*, série S.G., 1998, n°2, pp159-177.

Lopes Da Silva, A. *L'information et l'entreprise, des savoirs à capitaliser, méthodes, outils et applications à la veille*. Université de Droit, Eco. & Sciences d'Aix-Marseille III 2002. Thèse Sciences de l'information et de la communication.

Pateyron, E.L. Veille stratégique. *Encyclopédie de gestion*. Paris : Economica, 1998, tome 1, pp. 183-194.

Piolat, A. *La Recherche documentaire. Manuel à l'usage des étudiants, doctorants et jeunes chercheurs*. Marseille : Solal, 2003, 150 p.

Rostaing, H. *Veille technologique et Bibliométrie : concepts, outils applications*. Thèse Université de droit et des sciences Aix-Marseille, Faculté des sciences et techniques de Saint-Jérôme, 1993.

## Notes

- <sup>1</sup> Cité par Pateyron, E., Veille stratégique, in Encyclopédie de gestion, 1997, pp. 183-194.
- <sup>2</sup> Première étude co-réalisée par Jupiter Research et Iprospect : analyse des tendances observées sur trois ans (2002, 2004, 2006) ; deuxième étude de Harvest Digital et Métro Research : janvier 2006. <<http://veillepme.blogspot.com/etudes/>> (consulté le 29/01/2008).
- <sup>3</sup> Véronis, Jean. Étude comparative de six moteurs de recherche. <<http://www.up.univ-mrs.fr/veronis/pdf/2006-etude-comparative.pdf>> (consulté le 29/01/2008).
- <sup>4</sup> Logiciel "freeware" développé et distribué par l'Unesco.
- <sup>5</sup> <<http://aws.amazon.com>> (consulté le 29/01/2008).
- <sup>6</sup> La copie d'écran n'est pas une version complète du résultat. Certaines lignes ont été supprimées afin de montrer la façon dont la fin de la page de résultats de Google apparaît dans la feuille Excel.
- <sup>7</sup> Les ouvrages trouvés par le moteur contiennent les mots clés dans leur titre. Si l'on clique sur le lien correspondant, on obtient immédiatement tous les détails les concernant (résumé, avis de lecteurs etc.).
- <sup>8</sup> Dans cette logique, les textes en **gras** sont les textes utilisés pour le site <<http://www.amazon.com>> Les textes équivalents utilisés pour les autres sites de Amazon sont stockés dans la feuille "**Sites**" (voir paragraphe 2.5).
- <sup>9</sup> Microsoft.