

CHINESE SCIENTIFIC INFORMATION; AN ESSENTIAL INFORMATION SOURCE TO ANY BUSINESS INTELLIGENCE COMPETITIVE WATCH.

Nadège GUENEC, nadegeguenec@gmail.com

Université Paris-Est, Sciences et Ingénierie de l'Information Stratégique (S3IS), 5 bd Descartes, 77454 Marne la Vallée, France

Mots clefs :

Intelligence économique, veille compétitive, information scientifique et technique, data-mining, découverte de connaissance, Chine, Bases de données.

Keywords :

Business Intelligence, competitive watch, scientific and technical information, data-mining, knowledge discovery, China, Databases.

Palabras claves :

Inteligencia de Negocios, vigilancia competitiva, información científica y técnica, data-mining, descubrimiento de conocimientos, China, bases de datos.

Résumé

Toute démarche d'Intelligence Economique accompagnant l'implantation d'une activité à l'étranger démarre par la capacité à déchiffrer l'environnement d'affaire. Or une grande part de l'information à portée stratégique est issue de l'analyse du jeu des acteurs opérants dans le secteur d'activité visé. L'incapacité à « lire » la Chine et à comprendre les enjeux qui s'y déroulent, les écueils dans l'accès à l'information concernant le marché chinois sont un facteur important de ralentissement du développement des entreprises françaises sur l'un des marchés les plus dynamiques du monde. La Chine ayant mis en place une politique de développement scientifique extrêmement précise, elle vise à s'accaparer des pans entiers de l'économie et de la connaissance de demain via la prise de brevets.

Cet article propose une méthode automatique de création de connaissances de la Chine en tant que marché. Ce type d'informations à haute valeur ajoutée est créé par l'analyse de corpus bibliographiques extraits d'une base documentaire chinoise et fournit des connaissances stratégiques nécessaires à toute démarche d'Intelligence Economique sur le marché chinois. L'intérêt est qu'elle est accessible à toute personne, même non sinisante.

Abstract

While implementing its activity abroad, any organisation, has to deal with a new business environment. The ability to decipher this business environment is the condition for implanting a relevant business intelligence strategy. Such strategic information will emerge from the analysis of the different players in the relevant sector and their relation. The inability to "read" China and to understand the issues that take place there, the obstacles in the access to the right information about the Chinese market are important factors that slow down the development of French companies on one of the most dynamic markets in the world. The Chinese government has set up an extremely precise policy of scientific development which aims to capture large sections of the tomorrow world economy and knowledge via the patent and the intellectual property.

This article proposes an automatic method for creating knowledge about the Chinese market. This type of information with high added value is created by the analysis of large corpus extracted from a Chinese bibliographic database and provides strategic knowledge for a business intelligence approach of the Chinese market. This method is accessible to anyone, even for non-Chinese speakers.

1 Introduction

The opening of national economies and the acceleration of global trade have, in barely a decade, transformed the competitive environment of enterprises. The area of activity has expanded by the opening of new markets with very attractive potential. So are the BRIC (Brazil, Russia, India and China). Among these four countries, all impressive by their size, population and economic potential they represent, China is the least accessible. Brazil and India are using language close to ours. We have a very deep historical relationship with Russia and its manners we are well known. China is more distant and closed to our understanding because of a linguistic system radically different from the Indo-European languages on the one hand and of the fact of a culture and a thought system at odds with those of Western countries. Yet for a company of international size, which wants to extend its influence or simply to maintain its market position, including its own market, it is now absolutely essential to be present on the Chinese market.

How does a western company operate on a market that appears at first as inherently complex and enigmatic because of its otherness? During six years of observation in China, we have found out the pitfalls in access to information about the Chinese market. As on many markets, our companies are subject to some unimaginable destabilization. This unsatisfactory and powerlessness situation is shared by almost all public and private actors. [1] The inability to "read" China and understand the issues that take place in spite of sustained efforts, the tactical errors which arise from a misjudgement of the market or a biased understanding of the game players led us to consider a methodology that could provide French companies an approach to China as a market.

If a cultural approach is made of human interactions and subtleties, a market approach is now possible by the automatic processing of information and its modelling. In any process of economic intelligence accompanying the establishment of a foreign operation, a large part of the strategic information comes from analysis of the game players operating in the same sector of activity. Such automation of knowledge creation is, in addition to the human approach to the field, a real added value for understanding the interactions between the players because it provides a set of knowledge, taking into account more entities large, are more comprehensive, elusive anywhere else. Because China has highly developed technologies linked to the knowledge economy, it is now possible to explore the scientific and technological sources of information science in China. We are also convinced that Chinese sources of information will take a more and more crucial importance in any global watch. It is therefore an urgent need for organizations to get solutions that not only allow the access to this information but also are able to handle the masses of information from these sources.

Getting strategic information necessary to the survival of the company just in time, in a context of increasing internationalization and of multiplication of the sources of scientific, technical, economic ... made the important concept of business intelligence competitive watch. [2] We develop in this paper a method that will appeal any decision-maker or watcher wishing detects Chinese players in a given industry. We will take as a field of application for this example that of agricultural biotechnology and more specifically the theme of hybrid wheat in order to illustrate this information analysis approach of China. The analysis of this sector, which is an area of fundamental research, experimental and applied current gives rise to the acquisition of patents and to the marketing of commercial products. It is a very current topic.

¹ English is an official language of India and is the one used in the economic, political and scientific.

The analysis by the data-mining of a corpus of scientific articles extracted from a Chinese database will permit the understanding of how this sector is organized in the interior of China, what are the avenues of research, specific actors, how they evolve over time, etc.. ...

2 Method for analyzing Chinese scientific information

2.1. Sources of information - CNKI

The portal CNKI, China National Knowledge Infrastructure, is a project bringing together the various Chinese academic databases. Its development has been and remains strongly supported by the Chinese government whom relied on this portal to support the growth of the information society and the development in China of business intelligence through strategic information. [18] In addition to the Chinese databases, CNKI begins to open up to foreign bases. One such agreement has been signed with Springer in 2008 who find there an entry on the Chinese market. The consultation of bibliographic records is free, only downloading the full text is to be paid and requires a subscription. In late 2007, over twenty five million articles were referenced among the various base of CNKI.

The functionalities of the database are constantly transformed. Thus, between spring and autumn 2008, the configuration of CNKI was totally changed from the viewpoint of the user.

In addition with the very slow extraction of the corpus (it can not be done automatically), two new problems have emerged:

- On the one hand, a line of Chinese text contains up to three different fonts alternating one character to another. This cannot be remarked when reading the text but generates noise in the encoding.
- Secondly, the key words have disappeared while downloading the corpus; they no longer appear in the fields of a descriptive article.

These recent changes make it more complex bibliometric analysis. This shows the responsiveness of the Chinese on issues of information systems. However the automated processes have been adapted and the results have been obtained. It remains to be analyzed.

2.2. The corpus

2.2.1. Extraction

In part this work, we will strive to demonstrate what can be achieved with regard to automatic processing of scientific information from China.

In order not to overload the demonstration, we will work voluntarily on a little corpus. To do this we will launch a request focused precisely on male sterility, the principal condition of the hybridization process. Our research period runs from 2000 to today. We obtained 302 responses. We can thus through the analysis of this corpus see evolve over the last 8 years research on hybridization of wheat.

A more relevant analysis should collect all the articles that mention the word “wheat”. This request would permit to have a first hand view concerning the research on hybridization processes versus all the research areas related to wheat. But even if we limit the request on only one year and if we select the items

containing the word "wheat" only in the title, we obtain a corpus of 2597 responses. (Fig. 1) The analysis of such a corpus or a larger corpus if one fails the temporal restriction, would go beyond the scope of this article. That's why we have voluntarily restricted the corpus collected in a specific area.

快速检索

标准检索

高级检索

专业检索

引文检索

作者发文检索

科研基金检索

句子检索

工具书及知识元搜索

文献来源

发表时间：

具体日期

从 2007-10-22

到 2008-10-22

题名

小麦

并含

模糊

并且

题名

并含

精确

并且

题名

并含

精确

并且

题名

并含

精确

在结果中检索

检索文献

中英文扩展检索

检索结果不错，

生成检索报告

定制或收藏本次检索式

分组分析方法：

学科类别

中文关键词

研究层次

文献作者

作者单位

文献出版来源

研究获得资助

来源数据库

发表年度

不分组

排序分析方法：

相关性

发表时间

被引频次

下载频次

显示方式：

列表

摘要

显示记录数：100 50

共有记录2597条

REQUEST AREA

NUMBER OF CORRESPONDING ARTICLES

CORPUS

序号	题名	作者	文献来源	发表时间↓	来源库	被引频次	下载频次
1	2BMF-5固定茎小麦免耕播种机的设计	李太伟;李洪文;何进	农机化研究	2008-10-01	期刊	0	0
2	小麦施用锌肥增产效果明显	曲善功 郝建成	山东科技报	2008-09-22	报纸	0	0
3	加强小麦种子管理 确保明年小麦丰收	冯金莲	石家庄日报	2008-09-21	报纸	0	0
4	近红外检测技术在小麦品质及面制品研究中的应用	王玮;张泽俊;薛文通;张惠	食品科技	2008-09-20	期刊	0	0
5	切实做好小麦播种期病虫害防治工作	临汾市病虫区域测报站	临汾日报	2008-09-19	报纸	0	0
6	小麦全蚀病 拌种可防控	定州市农牧局 王虎	河北科技报	2008-09-18	报纸	0	0
7	如何搞好小麦备播	隆尧县农业局 张静辉	河北科技报	2008-09-18	报纸	0	0
8	国审小麦新品种石麦15号生产应用价值	武金焱;李夕军;陈丽;杨英茹	农业科技通讯	2008-09-17	期刊	0	0
9	栽培环境条件对专用小麦籽粒品质的影响	杨利玲;刘慧;陆志芳	农业科技通讯	2008-09-17	期刊	0	0
10	优质强筋小麦郑农16育种家种子生产技术方案	邱冬云;李宾;廖先静;张赞育	农业科技通讯	2008-09-17	期刊	0	0
11	河南省优质强筋小麦育种进展	马香花;王保林;周	农业科技通讯	2008-09-17	期刊	0	0

FIG. 1: Screenshot of the query "wheat" for the period October 2007 - October 2008 CNKI

The extracted corpus will be picked as follows (Fig.2):
[Wheat AND male sterility] during the period from October 2000 to October 2008

The screenshot displays the CNKI search interface. At the top, the '发表时间' (Publication Time) is set to '具体日期' (Specific Date) from '2000-10-22' to '2008-10-22', which is circled in red. Below this, the search criteria are defined using a combination of '题名' (Title) and '关键词' (Keywords) with '并含' (AND) logic. The first criterion is '小麦' (Wheat) AND '雄性不育' (Male sterility), also circled in red. Subsequent criteria are empty. The interface includes buttons for '在结果中检索' (Search in results), '检索文献' (Retrieve literature), and '中英文扩展检索' (Expand search in Chinese and English). Below the search criteria, there are options for '检索结果分组筛选' (Filter search results by group) and '生成检索报告' (Generate search report). The '分组分析方法' (Group analysis method) is set to '学科类别' (Discipline category). The '排序分析方法' (Sorting analysis method) is set to '相关性' (Relevance). The '显示方式' (Display mode) is set to '列表' (List). The '显示记录数' (Number of records displayed) is set to '10 20 50', with '50' circled in red. The total number of records is '共有记录302条' (Total records: 302).

FIG. 2: Screenshot of the query (wheat AND male sterility) from October 2000 to October 2008 on CNKI.

Then, the results of this request can be downloaded as a corpus shown as follow: (Fig. 3):

将你选中的以下文献

 到个人/机构馆中,或按照选择的输出格式

<input checked="" type="radio"/> CNKI桌面版个人数字图书馆 下载软件 <input type="radio"/> CAJ-CD格式引文 <input type="radio"/> RefWork <input type="radio"/> EndNote <input type="radio"/> NoteExpress <input type="radio"/> 查新 <input type="radio"/> 自定义	<p>DataType: 1 题名: 2BMF-5固定茎小麦免耕播种机的设计 作者: 李太伟,李洪文,何进; 单位: 中国农业大学工学院; 摘要: 针对我国西北地区固定茎保护性耕作条件下,玉米茬地茎作免耕播种小麦秸秆堵塞严重和播种质量差等问题,设计了一种驱动圆盘式固定茎小麦免耕播种机。整机主要由驱动圆盘式破茬装置、开沟施肥装置、单体仿形播种装置及镇压传动装置等组成。田间试验结果表明,驱动圆盘刀的平均入土深度为10cm,平均播种和施肥深度分别为5cm和10cm。该机一次作业可完成破茬、开沟施肥、播种和镇压等工序,作业时土壤扰动小,播种作业质量可满足西北地区农艺要求。 年: 2008 期: 10</p> <p>DataType: 4 题名: 小麦施用锌肥增产效果明显 作者: 曲善功,都建成 来源: 山东科技报 年: 2008</p> <p>DataType: 4 题名: 加强小麦种子管理 确保明年小麦丰收 作者: 冯金莲 来源: 石家庄日报 年: 2008</p> <p>DataType: 1 题名: 近红外检测技术在小麦品质及面制品研究中的应用 作者: 王玮,张泽俊,薛文通,张惠; 单位: 中国农业大学食品科学与营养工程学院; 摘要: 近红外技术(NIRS)是近年来迅猛发展起来的新技术,其原理是由于不同基团产生的光谱在吸收峰位置和强度上有所不同,根据朗白-比尔吸收定律,随着样品成分组成或者结构的变化其光谱特征也将发生变化,从而可以实现复杂物质的定性鉴别和定量分析。简要介绍了近红外分析技术的原理和特点,并对其在小麦品质、面制品加工中的应用做了综述,并详细介绍在面团调制、面包老化方面的研究进展。 年: 2008 期: 09</p> <p>DataType: 4 题名: 切实做好小麦播种期病虫害防治工作 作者: 临汾市病虫区域测报站 来源: 临汾日报 年: 2008</p>
--	---

FIG. 3: Screenshot of a page of bibliographic notices proposed as a result for the query.

We note that the base is very well structured. We will thus be able to operate a number of analysis based on the metadata contained in these bibliographic notices of this database to extract a value-added information not otherwise discernible.

2.2.2. Specificity of a Chinese corpus

Our analysis focuses on the metadata contained in the bibliographic records of this corpus. This information is not visible to the simple reading and it is necessary to have a specific tool to perform the analysis by crossing metadata fields and by counting the frequency of datas. It is not simply to model information based on their content but in terms of uses that will be made.[19]

The software Tetralogie, developed by the “Institut de Recherche en Informatique de Toulouse” (IRIT) is specifically dedicated to data-mining [20] and allows the emergence of networks of actors and their dynamics from a given corpus, to highlight the evolution of concepts and topics and to detect weak signals. [21] Its use has been lauded on many occasions and our goal is to use it in a new linguistic environment. It has been necessary to think the evolution of the software in order to adapt to the environment of the Chinese language on the one hand and to the structure of the database CNKI on the other hand. Each database has its own structure and we had to adapt to the new format of CNKI. It was therefore necessary to create a descriptor of the database structure. This descriptor defines the different fields of the database, identifies their banner, their separators, their utility and the various types of information they contain. It also allows locating the beginning of each notice as well as the physical structure of the record (format and number of occurrences of banners).

An excerpt from our corpus as it appears during the download allows seeing that the articles are presented in the same way (Fig. 4):

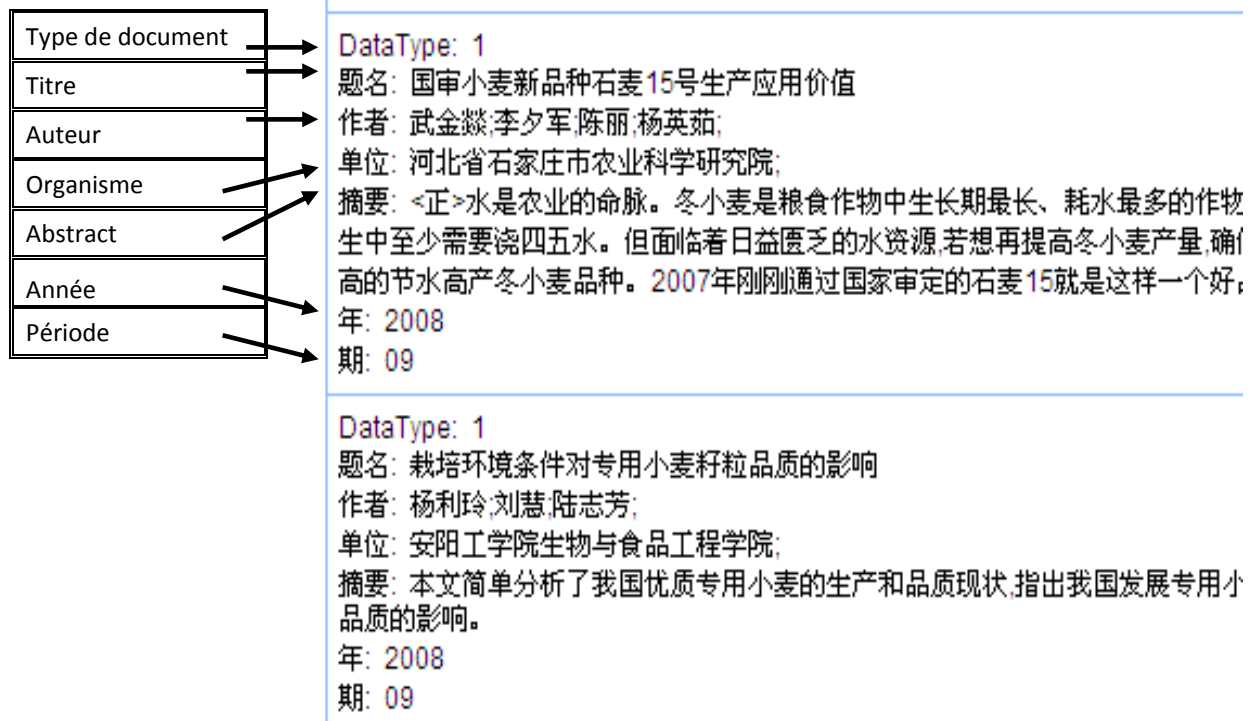


FIG. 4: Description of the metadata fields of a bibliographic notice.

This will allow automatic processing of all the bibliographic records of this corpus and will propose, ultimately, strategic analysis perfectly accessible to the non-Chinese reader even though the processed information is given in Chinese language. Concerning the processing of Chinese characters, we will then go by the Unicode codification in order to translate the metadata fields and to mark them. (Table 1)

TAB.1: Fields correspondence chinese character / Unicodewith English translation

Chinese banner	English translation	Unicode
题名	Title	题名
作者	Author	作者
单位	Organism	单位
摘要	Abstract	摘要
年	Year	年
期	Period	期

An additional challenge for treatment is that the police of the script constantly change. A sentence may contain up to three different fonts, which increases the noise in the coding and make the treatment more complex. For example, the following notice: the title field (题名) has two characters. The first is written with the police Simsun and the second is written with the police MS UI Gothic.

Visualisation of the code just for the field « title » :

```
??>DataType:&nbsp;&nbsp;&nbsp;1<br>
</span><span lang=ZH-CN style='font-size:9.0pt;font-family:SimSun;mso-bidi-font-family:
SimSun'>&#39064;</span><span lang=ZH-CN style='font-size:9.0pt;font-family:
"MS UI Gothic";mso-bidi-font-family:"MS UI Gothic">&#21517;</span><span
style='font-size:9.0pt;font-family:??>:&nbsp;&nbsp;&nbsp;</span>
```

Thus, data processing is done directly on the computer code of the Chinese language. Here is the coding of a single bibliographic notice. It will then be necessary to establish a reformer to clean up this record to keep only the Chinese characters and bibliographic fields.

Thus, the title of a notice stated:

题名:

不同时期茉莉酸甲>酯处理对光温敏雄性不育小麦颖花开放的诱导效应

is coded as follows:

```
??>DataType:&nbsp;&nbsp;&nbsp;<br>
</span><span lang=ZH-CN style='font-size:9.0pt;font-family:SimSun;mso-bidi-font-family:
SimSun'>#39064;</span><span lang=ZH-CN style='font-size:9.0pt;font-family:
"MS UI Gothic";mso-bidi-font-family:"MS UI Gothic">#21517;</span><span
style='font-size:9.0pt;font-family:??':&nbsp;&nbsp;&nbsp;</span><span lang=ZH-CN
style='font-size:9.0pt;font-family:"MS UI Gothic";mso-ascii-font-family:??;
mso-hansi-font-family:??'>#19981;#21516;</span><span lang=ZH-CN
style='font-size:9.0pt;font-family:SimSun;mso-bidi-font-family:SimSun'>#26102;</span><span
lang=ZH-CN style='font-size:9.0pt;font-family:"MS UI Gothic";mso-bidi-font-family:
"MS UI Gothic">#26399;#33545;#33673;#37240;#30002;</span><span
lang=ZH-CN style='font-size:9.0pt;font-family:SimSun;mso-bidi-font-family:
SimSun'>#37231;#22788;</span><span lang=ZH-CN style='font-size:9.0pt;
font-family:"MS UI Gothic";mso-bidi-font-family:"MS UI Gothic">#29702;</span><span
lang=ZH-CN style='font-size:9.0pt;font-family:SimSun;mso-bidi-font-family:
SimSun'>#23545;</span><span lang=ZH-CN style='font-size:9.0pt;font-family:
"MS UI Gothic";mso-bidi-font-family:"MS UI
Gothic">#20809;#28201;#25935;#38596;#24615;#19981;#32946;#23567;#40614;</span><span
style='font-size:9.0pt;font-family:??'>BS366</span><span lang=ZH-CN
style='font-size:9.0pt;font-family:SimSun;mso-bidi-font-family:SimSun'>#39062;</span><span
lang=ZH-CN style='font-size:9.0pt;font-family:"MS UI Gothic";mso-bidi-font-family:
"MS UI Gothic">#33457;</span><span lang=ZH-CN style='font-size:9.0pt;
font-family:SimSun;mso-bidi-font-family:SimSun'>#24320;</span><span
lang=ZH-CN style='font-size:9.0pt;font-family:"MS UI Gothic";mso-bidi-font-family:
"MS UI Gothic">#25918;#30340;</span><span lang=ZH-CN style='font-size:
9.0pt;font-family:SimSun;mso-bidi-font-family:SimSun'>#35825;#23548;</span><span
lang=ZH-CN style='font-size:9.0pt;font-family:"MS UI Gothic";mso-bidi-font-family:
"MS UI Gothic">#25928;</span><span lang=ZH-CN style='font-size:9.0pt;
```

font-family:SimSun;mso-bidi-font-family:SimSun'>#24212;

The reformer that we have incorporated in Tetralogie will clean out the above notice. The following result will be reached, its format then being able to be manipulated for a bibliometric analysis:

DataType:1

#39064;#21517;#19981;#21516;#26102;#26399;#33545;#33673;#37240;#30002;#37231;#22788;#29702;#23545;#20809;#28201;#25935;#38596;#24615;#19981;#32946;#23567;#40614;#39062;#33457;#24320;#25918;#30340;#35825;#23548;#25928;#24212;

After having cleaned the corpus, treatment will be made. We give here some examples of results, but we invite the reader to refer himself to previous publications about Tetralogie in order to have a clearer idea of the wide range of analysis and treatments of information that can produce this software. We present here some results that can be obtained after treatment with the software Tetralogie on the corpus male sterility.

2.3. Results and analysis

Here is a first representation of the networks of players in the corpus "male sterility". We can quickly note out that the research teams are very distinct and disconnected from each other, some of them even containing only three authors. We will not talk about these little groups and will target first the biggest one. (Fig 5) Indeed, not only the collaborations between authors are numerous , which induces the existence of a real research team within which exchanges are of good profit, but more is the representation of some authors is bigger which indicates that these authors are also the most prolific.

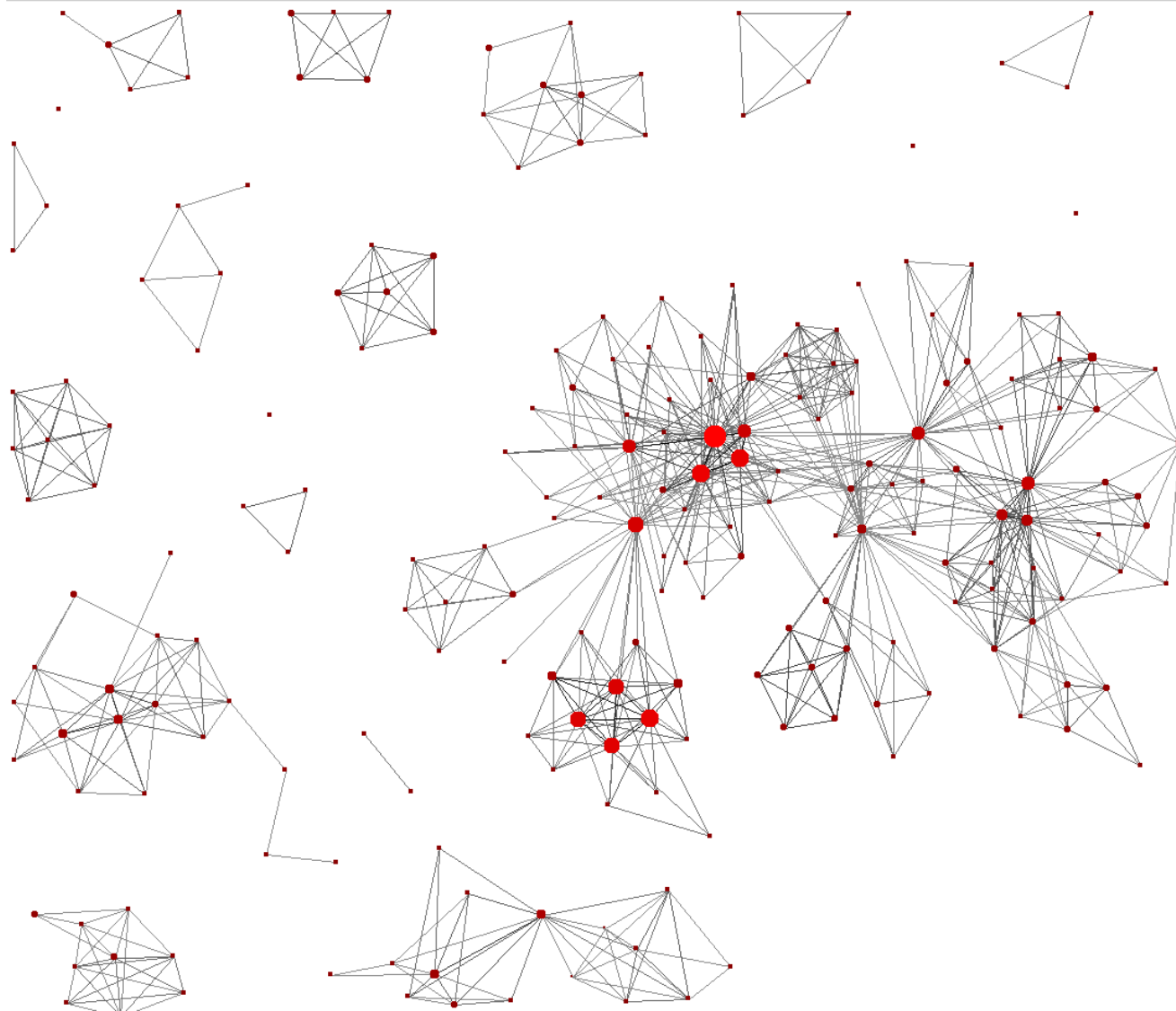


FIG. 5: Visualisation of author's collaborations.

Having thus identified the main network, we can now refer to the matrix that will allow to identify the players and also to monitor their evolution

Tétralogie V7.0 Tableau 2D Fichier : AU-DP											
		2000-1	2002-3	2004-5	2006-8	CLASSE	POIDS				
1	203095		7	6	15		1				
2	2435210	8	13	17	13		1				
3	395329		7	6	12		1				
4	329933		5	6	19		1				
5	232751	1	2	7	11						
6	234352		6	6	10						
7	2646924	1	8	5	8						
8	396401	5	3	2	7						
9	2644614		1	2	7						
10	362139			2	7						
11	370411				7						
12	2435236			2	5						
13	312062			1	5						
14	234353				5						
15	2646911				6						
16	339911		2	2	4						
17	2646919		1	2	4						
18	263613				4						
19	200521				4						
20	243927	1	2		3						
21	20313	2	4		3						
22	2644630	2	1	2	3						
23	214942			1	3						
24	2646923			1	3						
25	274941			1	3						
26	20167			1	3						
27	36911			1	3						
28	23391			1	3						
29	264172				3						
30	28966				3						
31	295799				3						
32	37085				3						
33	2101622				3						
34	2644612				3						
35	362139				3						
36	2101619				3						
37	28519				3						
38	2646925				3						
39	2644640			1	2						
40	29279			1	2						
41	2435219	7	9	1	2						
42	203096	1		1	2						
43	2957921				2						
44	35449				2						
45	40898		1	2	2						
46	23447	1			2						
47	2435225			1	2						
48	21333			1	2						
49	295791			1	2						
50	203091			1	2						
51	31206			1	2						

FIG. 6: Unicode visualisation of the name of the related author group.

By submitting the correspondence from Unicode to the Chinese characters we get the following names (Table 2):

TAB. 2: Correspondence Unicode / Chinese characters with pinyin translation

Unicode	Chinese	Pinyin translation
张改生	张改生	ZHANG Gaisheng
何蓓如	何蓓如	HE Beiru
马翎健	马翎健	MA Jianling
宋喜悦	宋喜悦	SONG Xiyu
牛娜	牛娜	NIU Na
高庆荣	高庆荣	GAO Qingrong
李宏斌	李宏斌	LI Hongbin
赵昌平	赵昌平	ZHAO Chanping
张风廷	张风廷	ZHANG Fengting
来源小麦遗传育种国际学术讨论会	来源小麦遗传育种 国际学术讨论会	International Symposium on wheat genetic and breeding.
秦志列	秦志列	QIN Zhilie
邱新民	邱新民	BING Xinmin
宋国琦	宋国琦	SONG Guoqi
来源内蒙古农业大学	来源内蒙古农业大学	Inner Mongolia agricultural university.

With this method, we could quickly identify the most Chinese researchers working in the field of hybrid wheat research. It will be easier for French decision-makers to conduct research on their behalf to be able to contact them then, to monitor their work accurately and possibly to enter into negotiations with these people.

We call for very special vigilance on the names of Chinese authors. Indeed, the Chinese surnames are few. The popular Chinese saying speaks about only one hundred Chinese family’s names. If in fact the number of Chinese names is more than a hundred, it is still small, especially given the immensity of the Chinese population. One will therefore have to be very careful with regard to verification of the names so as not to confuse people. The family name is not enough, it is imperative to search the full name of the person. One will take extra precautions during a pinyin transcription of a Chinese name because it was then that the ambiguity will arise. It could eventually be useful to combine the search by name with a keyword. For example, the engine CNKI proposes keywords related to a given search. When launching a query with an author's name, other people names are given. Here are a few. (Table 3)

TAB. 3: Part of the list proposed by CNKI, for authors with similar names to the one originally sought.

张改	张改娥	张改红	张改慧
张改惠	张改兰	张改利	张改莲
张改连	张改玲	张改梅	张改娜
张改琴	张改芹	张改清	张改荣
张改生	张改侠	张改香	张改英

All these names start with “ZHANG Gai” and only the last character of each first name makes the distinction between the different people. Moreover, if it is clear, reading the Chinese language, that 张改莲 and 张改连 is not the same person, the last character of each name being different, this is far less clear in pinyin, the two names being transcribed “ZHANG Gai Lian”. Similarly for 张改慧 and 张改惠, both are bearing the Chinese name "ZHANG Gai Hui" with a different character for the last phoneme “hui”. Here appears an additional interest of the software Tetralogie that realise the analysis directly on the Chinese language without any translation, limiting thus any confusion.

2.4. Validation and comparison

To validate our results, we can verify the identity of the authors proposed above by the system. This article does not allow us to deepen each author; we propose to verify three authors taken at random from which we will dwell a little in order to ensure they match with the criteria of our research.

HE Beiru² :

Professor and director of research, involved since the 1970s in research on male sterility of wheat with three lines; Creator of several varieties of wheat and other triticales; Key member in the people's committee of Shaanxi Province and in the Chinese Academy of Sciences; Numerous awards for scientific excellence; Highly politically involved in the planning of the Chinese agricultural field crops; Testing of wheat at high altitude performance (in Tibet).

ZHAO Chanpin :

Director of the Research and engineering hybrid wheat centre, Beijing Academy of Agriculture and Forestry; Direction of genetic research on wheat and molecular biology; Research on thermo-photo-sensitive factors of male sterility of wheat and on the improvement of wheat yield by two lines of reproduction; Study the DNA of wheat, database and germplasm creation of innovative new varieties, cloning; Numerous awards for scientific excellence and president of various government projects, including some from the Ministry of Agriculture and some from the Ministry of Science and Technology; part of the commissions of awards research grants; Involved in political decisions on food safety.

ZHANG Gaisheng

Member of the People's Consultative Conference of the Shaanxi Province; PhD in agricultural sciences, major in genetics and plant breeding in 1993 after studies in France, England and Switzerland; Post-doc at the North-west Agricultural University on the breeding of wheat; He currently teaches there as a professor and laboratory director at the centre selection and Biotechnology Yangling and at the national centre for the improvement of wheat yield; is also Executive Director of the Institute of Chinese culture and genetic resources, the editor of several scientific journals and received several prizes for his research on male sterility of wheat by means of molecular genetics and cellular. It is engaged in political programs concerning the development of science and food safety.

A short presentation of these authors has enabled us to validate their importance in research in Chinese wheat hybrid. The authors being proposed by Tetralogie are really key-characters at the national level, not only in terms of research but also at political level. Tetralogie has identified the actors in a simple way, while in the mass of articles; all authors do not have the weight as proposed.

One can go further in the validation of our approach, taking the example of the author ZHANG Gaisheng. If we run a query with his name on CNKI, we get 236 articles, including 19 on the year 2008. (Fig.7) (See the publication dates besides).

2008-07-15	2008-12-16	2008-11-10	2008-10-01	2008-09-15
2008-08-20	2008-08-15	2008-08-15	2008-07-25	2008-07-15
2008-06-15	2008-06-12	2008-06-10	2008-05-12	2008-04-20
2008-04-10	2008-03-10	2008-02-15	2008-01-15	

² We got these informations from Baidu Baike, the Chinese “wikipedia”. <http://baike.baidu.com/>

快速检索 | 标准检索 | **高级检索** | 专业检索 | 引文检索 | 作者发文检索 | 科研基金检索 | 句子检索 | 工具书及知识元搜索 | 文献出版来源

发表时间： 具体日期 从 到

作者 张改生 并含 精确

并且 题名 并含 精确

并且 题名 并含 精确

并且 题名 并含 精确

在结果中检索 检索文献 中英文扩展检索

检索结果分组筛选：

生成检索报告 定制或收藏本次检索式

分组分析方法： 学科类别 中文关键词 研究层次 文献作者 作者单位 文献出版来源 研究获得资助 来源数据库 发表年度 不分组

排序分析方法： 相关度 发表时间 被引频次 下载频次

显示方式： 列表 摘要 显示记录数：10 20 50

全选 清除 定制 存盘 上页 下页 共有记录236条

序号	题名	作者	文献来源	发表时间	来源库	被引频次	下载频次
1	小麦多子房性状外显率影响因子的初步研究	张国慧; 张改生; 葛峰辉; 牛娜; 马守才; 潘栋梁; 汪奎	植物遗传资源学报	2008-12-16	期刊	0	1
2	油菜数量性状QTL定位研究进展	王俊生; 董育红; 张改生; 李殿荣; 王淑; 田建华	西北农林科技大学学报(自然科学版)	2008-11-10	期刊	0	48
3	三个小麦抗白粉病新种质创制及其抗性基因的染色体定位和分子标记	王长有; 吉万全; 张改生; 王秋英; 刘素兰	中国遗传学会第八次代表大会暨学术讨论会	2008-10-01	会议	0	0
4	山羊草属不同细胞质对小麦籽粒戊聚糖含量的影响	潘建刚; 张改生; 牛娜; 马利兵	麦类作物学报	2008-09-15	期刊	0	6
5	小麦花药蛋白组双向电泳技术体系的优化	陈蕊红; 张改生; 刘卫; 叶景秀; 牛娜	核农学报	2008-08-20	期刊	0	81
6	49份小麦种质资源中Glu-1, Glu-3, Gli-1位点基因组成分析	王瑞; 张改生; 王宏; F J ZELLER; S L K HSAM	河南农业大学学报	2008-08-15	期刊	0	13
7	小麦遗传型与生理型雄性不育花药蛋白组双向电泳分析	刘卫; 陈蕊红; 张改生; 牛娜	遗传	2008-08-15	期刊	0	192
8	小麦籽粒戊聚糖研究进展	潘建刚; 马利兵; 张	陕西农业科学	2008-07-25	期刊	0	10

FIG. 7: Screenshot of the query about ZHANG Gaisheng on CNKI and the dates of its publications in 2008.

By searching with the name "ZHANG Gaisheng" by author on PubMed, we got no result (Fig. 8):

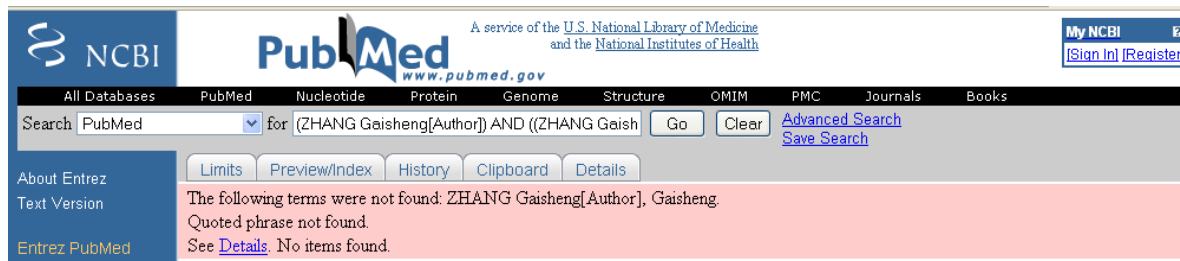


FIG. 8: Screenshot of the query about ZHANG Gaisheng on the PubMed database.

4 articles are available on the French database INIST but only the 2 first are really relevant concerning the author (fig. 9):

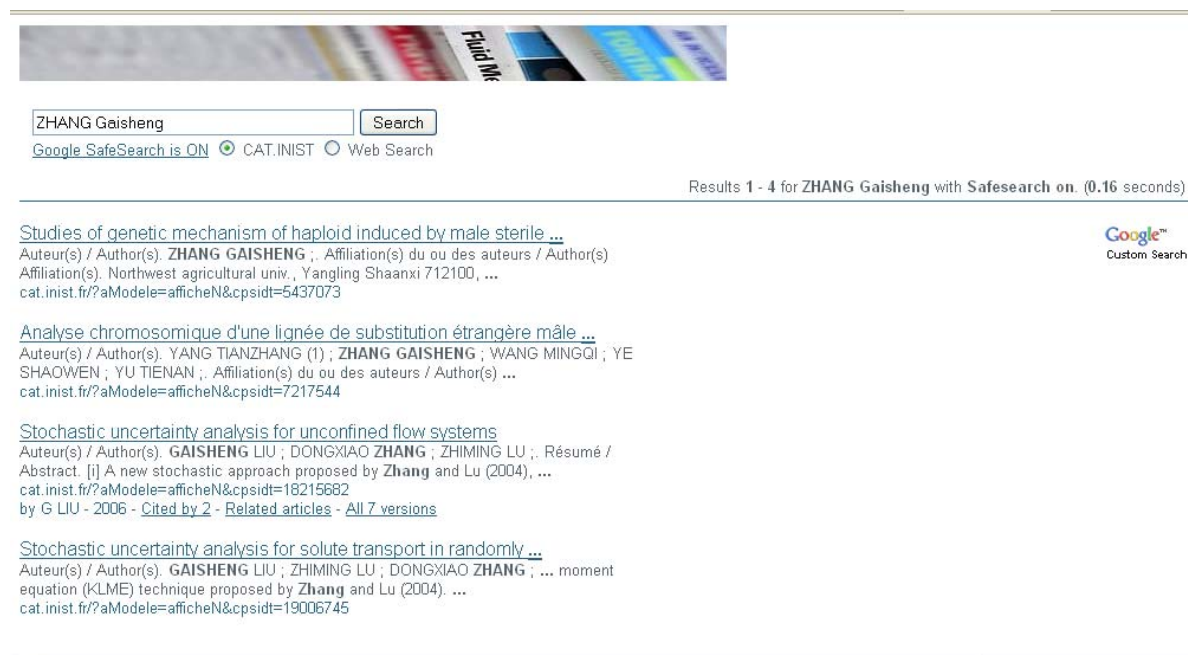


FIG. 9: Screenshot of the query about ZHANG Gaisheng on the INIST database.

Research on the year 2008 with Google Scholar engine has proposed two articles, the first one being referenced by Elsevier and the second by Wanfang data³ (fig. 10).

Google Scholar BETA

ZHANG Gaisheng depuis 2008 Rechercher

Rechercher sur le Web Rechercher les pages en français

Scholar [Tous les articles](#) - Articles récents Résultats 1 - 2 sur 2 pour ZHANG Gaisheng. (0.07 secondes)

Essayez avec cette orthographe : [ZHANG Baosheng](#)

[Investigation of flow and transport processes at the MADE site using ensemble Kalman filter](#)
 G Liu, Y Chen, D **Zhang** - Advances in Water Resources, 2008 - Elsevier
 ... **Gaisheng** Liu a , Corresponding Author Contact Information , E-mail The
 Corresponding Author , Yan Chen b and Dongxiao **Zhang** c , d. ...
[Autres articles](#) - [Recherche sur le Web](#) - [Les 2 versions](#)

[普通小麦-Elymus rectisetus衍生系的细胞学鉴定及其特征特性研究](#)
 王长有, 王秋英, 吉万全, 薛秀庄, 刘雪利 - ACTA AGRICULTURAE BOREALI-SINICA, 2008 - 万方数据资源系统
 ... 吉万全, 张改生, 范宏军, 王秋英, 任志龙, 张宏, 蔡东明, 薛秀庄, Wang
 Changyou , Ji Wanguan , **Zhang Gaisheng** , Fan Hongjun , Wang ...
[Autres articles](#) - [Recherche sur le Web](#)

Essayez avec l'orthographe suivante : [ZHANG Baosheng](#)

ZHANG Gaisheng Rechercher

FIG. 10: Screenshot of the query about ZHANG Gaisheng using Google scholar engine.

Finally, we add that the results provided by this analysis have effectively permitted to identify the key elements of the Chinese research in wheat hybrid. The content of this article does not allow detailing the entirety of the study that was conducted but we can say that this analysis has led a French company to approach major Chinese players in biotechnology research. We have seen that the areas of research, economics and politics in China were much more closely linked than in Western countries. A research partnership with a Chinese, even institutional, opens the door to market itself. This analysis provides policy makers a quick and an appropriate targeting of key players. Then it is the sensitivity of the negotiators who will be on the task.

³ Wanfang data is a Chinese database for scientific articles. It is the main competitor for CNKI. It lists fewer stories than the latter but gives the advantage to the Chinese public to access to major international portals and databases, enabling it to retain a certain percentage of market shares on the Chinese market provider's information.

A more thorough study inevitably follows this analysis. Having identified some key authors, a selection of their articles can be carried out for translation. This phase of translation is situated only at the very final steps of the analysis and is specifically targeted. It is done on safe items.

3 Conclusion

In a competitive environment increasingly complex, organizations have the imperative need to acquire strategic information in record time. In a first stage of competitive watch, China is to be considered as a global entity, in order to determine its global involvement in a given sector of research. Then, it is important in a process of business intelligence to have a clear vision of dynamic networks of actors in terms of relational analysis to identify swiftly and surely the players to approach. Grasp the reality of the potential of the relationnal of an organization is a way to adjust its strategy and tactics and to deploy adequate tactics especially in the commercial field. [22]

A so large country has many national entities operating on the same market sector. We gave an example for identify the key players in agricultural biotechnologies and in particular wheat hybrid. The same approach could equally well have been applied to other market segments.

This presentation will permit the reader to become aware of the need to conduct active monitoring and analyzing relevant information directly extracted from the Chinese databases, bypassing the translation in a first step. The Western Web, including the specialized databases providing professional scientific information is not sufficient resources to collect information about China. It is reckless to make an analysis of the Chinese market without going through the sources of information in China. Deprive ourselves of the Chinese sources, is to deprive ourselves from a significant part of the scientific world.

The relevant information obtained in real time becomes a key factor of success in every strategic step or action and within the decision making process. This is true especially in a rapid growth market, where relations between individuals are more difficult to highlight. The software Tetralogie finds out in the analysis of the Chinese IST a new opening. It will assess the operational capabilities of China in general and a Chinese entity in particular, to suggest a strategy for action in the Chinese market.

The major benefit of this approach is that it is quickly applicable by any individual, even those who don't speak Chinese. In addition, the software Tetralogie coming out from the French public research, its approach arose in response to the statement by Mr. JUILLET said that "we have a terrible lack of tools from French or European origins." [23] The futures searches will focus on analyzing multi-sources corpus in Chinese language.

4 References:

- [1] **CARAYON Bernard**, *Intelligence Economique, compétitivité et cohésion sociale*, Editions La Documentation Française, Paris, 2003.
- [2] **DOU Henri**, *Veille Technologique et compétitivité : l'Intelligence Economique au service du développement industriel*. Ed. Dunod, Paris, 1995.
- [3] **Gouvernement chinois**, site Web officiel, *China presents top awards to outstanding scientists*, 09-01-09. http://english.gov.cn/2009-01/09/content_1200919.htm Consulté le 26-01-09.
- [4] **OST**, *Edition 2008 du rapport de l'Observatoire des Sciences et Techniques*, OST, Paris, 2008.
[http://www.obs-ost.fr/dossiers/article/publication-de-ledition-2008-du-rapport-de-lost.html?tx_ttnews\[backPid\]=5&cHash=7008cb6c80](http://www.obs-ost.fr/dossiers/article/publication-de-ledition-2008-du-rapport-de-lost.html?tx_ttnews[backPid]=5&cHash=7008cb6c80) Consulté le 26-01-09.

- [5] **OCDE**, *Perspectives de l'OCDE de la science, de la technologie et de l'industrie*, éd. 2006, OCDE, Paris, 2006.
http://www.oecd.org/document/61/0,3343,fr_2649_33703_37743997_1_1_1_1,00.html Consulté le 26-01-09.
- [6] **BUTLER D.**, *China: The great contender*. Nature 454, <http://www.nature.com/news/2008/080723/full/454382a.html>, 23 July 2008.
- [7] **ISTIC**, *2008 年中国和世界十大科技进展新闻评出 (2008 China and the world's top ten scientific and technological progress in selected news)* 2008-11-27.
<http://www.chaxin.org/EducationDetail.aspx?ArticleID=86505> Consulté le 26-01-09.
- [8] **PEREZ Alain**, *R&D : Pékin miné par la fuite des cerveaux*, Les Echos du 06-07-2007. http://www.chine-informations.com/actualite/chine-rd-pekine-mine-par-la-fuite-des-cerveaux_7468.html, Consulté le 26-01-09.
- [9] **Gouvernement chinois**, Ministère de la Science et de la Technologie, *More scientific paper from China*, Newsletter n°531, 10-12-08.
http://www.most.gov.cn/eng/newsletters/2008/200812/t20081210_66158.htm Consulté le 26-01-09.
- [10] **VILLALONGA André**, *La Chine se hisse dans le peloton de tête du classement des publications scientifiques*, ADIT, BE Chine 56, 8/01/2009.
<http://www.bulletins-electroniques.com/actualites/57207.htm>, Consulté le 26-01-09.
- [11] **ZHOU Ping, LEVDESORFF Loet**, *The emergence of China as a leading nation in science*. Research Policy, Volume 35, Issue 1, February 2006, pages 83-104.
<http://users.fmg.uva.nl/lleydesdorff/ChinaScience/ChinaScience.pdf>, Consulté le 26-01-09.
- [12] **OCDE**, *Science, technologie et industrie : Perspectives de l'OCDE, édition 2008*, Paris, 2008. http://www.oecd.org/document/41/0,3343,fr_2649_34273_41564137_1_1_1_1,00.html, Consulté le 26-01-09.
- [13] **QUOTIDIEN DU PEUPLE**, *La Chine dresse les grandes lignes d'un projet de développement de la bio-économie*. Quotidien du peuple. Le 28.06.2007
<http://french.peopledaily.com.cn/Economie/6200702.html> Consulté le 26-01-09.
- [14] **DAVID Amos, SIDHOM Sahbi**, *Intégration de la démarche d'intelligence économique dans l'architecture fonctionnelle d'un système d'information*. Le système National d'information économique ; état et perspectives 2005. CERIST. Alger, Algérie.
- [15] **CNNIC**, *Statistical Survey Report on the Internet Development in China*, « China Internet Network Information Center », Janvier 2008, Pékin, Chine.
<http://www.cnnic.net.cn/uploadfiles/pdf/2008/2/29/104126.pdf>, Consulté le 26-01-09.
- [16] **ZHENG Shengli**, *L'incidence économique de la protection des bases de données en Chine*, dans Organisation mondiale de la propriété intellectuelle, Comité permanent du droit d'auteur et des droits connexes, septième session, Genève, 13-17 Mai 2002.
http://www.wipo.int/edocs/mdocs/copyright/fr/sccr_7/sccr_7_6.pdf, Consulté le 26-01-09.
- [17] **MA Jiaji**, « *Bibliothèque et Document Numérique en Chine* », note de synthèse pour le RTP-DOC, CNRS, Juin 2005.
- [18] **Quotidien du peuple**, *La Chine dévoile un moteur de recherche des journaux académiques*, 12-10-2006.
<http://french.people.com.cn/Sci-Edu/4910811.html>, Consulté le 26-01-09.
- [19] **DAVID Amos**, *Organisation des connaissances dans les systèmes d'information orientés utilisation – Contexte de veille et d'intelligence économique*, PU Nancy, Avril 2005.
- [20] **DOUSSET Bernard**. *Intégration de méthodes interactives de découvertes de connaissances pour la veille stratégique*. HDR, Université de Toulouse III, 2003.
- [21] **LOUBIER Eloïse, BAHSOUN Wahiba, DOUSSET Bernard** “*La prise en compte de la dimension temporelle dans la visualisation de données par morphing de graphe*”. Dans : Colloque Veille Stratégique Scientifique et Technologique (VSST 2007), Marrakech, 21/10/2007-25/10/2007, IRIT, (support électronique), octobre 2007b.
- [22] **BRUNEAU Jean-Maurice, MARCON Christian**. *Information stratégique informelle : le potentiel relationnel d'une organisation*. VSST 2006.
- [23] **JUILLET Alain**, *Nous manquons cruellement d'outils informatiques d'origine française ou européenne*. 01 DSI n°10. Interview par N. ARPAGIAN. le 01-03-2005.
http://www.spyworld-actu.com/IMG/_article_PDF/article_232.pdf, Consulté le 26-01-09.

Figure list:

FIG. 1: Screenshot of the query "wheat" for the period October 2007 - October 2008 CNKI.

FIG. 2: Screenshot of the query (wheat AND male sterility) from October 2000 to October 2008 on CNKI.

FIG. 3: Screenshot of a page of bibliographic notices proposed as a result for the query.

FIG. 4: Description of the metadata fields of a bibliographic notice.

FIG. 5: Visualisation of author's collaborations.

FIG. 6: Unicode visualisation of the name of the related author group.

FIG. 7: Screenshot of the query about ZHANG Gaisheng on CNKI and the dates of its publications in 2008.

FIG. 8: Screenshot of the query about ZHANG Gaisheng on the PubMed database.

FIG. 9: Screenshot of the query about ZHANG Gaisheng on the INIST database.

FIG. 10: Screenshot of the query about ZHANG Gaisheng using Google scholar engine.

Table list:

TAB. 1: Fields correspondence Chinese character / Unicode with English translation.

TAB. 2: Correspondence Unicode / Chinese characters with pinyin translation.

TAB. 3: Part of the list proposed by CNKI, for authors with similar names to the one originally sought.

Annexes :

Query launch with Google scholar engine with the key words « male sterility gene » realised on the 09-01-09.

Résultats 1 - 10 sur un total d'environ 938 pour "Male sterility gene"

[Cité 57 fois](#) - [Autres articles](#) - [Recherche sur le Web](#) - [Importer dans BibTeX](#) - [Les 2 versions](#)

[Cité 18 fois](#) - [Autres articles](#) - [Recherche sur le Web](#) - [Importer dans BibTeX](#) - [Les 4 versions](#)

[Cit  13 fois](#) - [Autres articles](#) - [Recherche sur le Web](#) - [Importer dans BibTeX](#) - [Les 5 versions](#)

Cité 13 fois - Autres articles - Recherche sur le Web - Importer dans BibTeX - Les 3 versions

[Cité 118 fois](#) - [Autres articles](#) - [Recherche sur le Web](#) - [Importer dans BibTeX](#) - [Les 5 versions](#)

[Cité 46 fois](#) - [Autres articles](#) - [Recherche sur le Web](#) - [Importer dans BibTeX](#) - [Les 2 versions](#)


mitochondrial morphology in yeast and plant cells. Yann ...

... Dalian, China, p 37. Kou HS, Song YH, Heu MH, Lee HS, McCouch SR (1999) Molecular mapping of a new genetic **male-sterility gene** causing chalky endosperm in rice

[Cité 34 fois](#) - [Autres articles](#) - [Recherche sur le Web](#) - [Importer dans BibTeX](#) - [Les 5 versions](#)

Auteurs clés: [B Wang](#) - [H Nguyen](#) - [D Page](#) - [Y Tao](#) - [N Dong](#)

Same query as above with the Google Scholar engine with the key words « male sterility gene » in Chinese realised on the 09-01-09.



-

[Recherche avancée Scholar](#)
[Préférences Scholar](#)
[Aide Scholar](#)

☒ Rechercher sur le Web
 ☐ Rechercher les pages en français

Scholar
Tous les articles - [Articles récents](#)

Résultats 1 - 10 sur un total d'environ 1 320 pour **\"性别不育基因\"** (0,12 secondes)

[水稻野败型细胞质雄性不育基因的SSR和SSLP标记分析](#)
 景渭春, 朱英国 · 中国农业科学, 2000 · [cqvip.com](#)
 @10-中国农业科学2000, 33(2): 47-15 Scientia Agricultura Sinica
[水稻野败型细胞质雄性不育恢复基因的SSR和SSLP标记分析彭 ...](#)
[Cite 15 fois · Autres articles · Rechercher sur le Web](#)

[玉米雄性不育基因\(ms30\)的RFLP作图](#)
 梁业红, 周洪生 · 作物学报, 2000 · [cqvip.com](#)
 , 21a 第26卷第3期2000年5月作物学报ACTA AGRONOMIC SINICA Vol. 26, NO. 3 May, 2000 玉米雄性不育基因(ms30)的RFLP作图: . 口斌J...
[Cite 15 fois · Autres articles · Rechercher sur le Web](#)

[一个用于甘蓝显性雄性不育转育辅助选择的SCAR标记](#)
 王晓武, 方智远 · 园艺学报, 2000 · [cqvip.com](#)
 ... 北京100081 方智远 中国科学院蔬菜花卉研究所,北京100081 SCAR
 显性雄性不育基因 甘蓝 分子标记辅助选择 2000 027 002 143 ...
[Cite 15 fois · Autres articles · Rechercher sur le Web](#)

[萝卜胞质大白菜雄性不育系的生化分析](#)
 孙日飞, 方智远 · 园艺学报, 2000 · [cqvip.com](#)
 ... carotenat in anthers ofOgura CM S (ms/ s P) *表示18隐5%差异显著水平。
 T-test s'lieant at 5%level 3 讨论植物雄性不育基因在表达 ...
[Cite 21 fois · Autres articles · Rechercher sur le Web](#)


[大豆\(Alium fistulosumL.\)胞质雄性不育基因的RAPD标记](#)
 盖利鹏, 孟祥栋 · 农业生物技术学报, 2002 · [cqvip.com](#)
 农业生物技术学报,lotmaal 0(a- 研究报告: 大(A/ /~
 fistulosumL.)胞质雄性不育基因的RAPD标记盖利鹏孟祥栋 ...
[Cite 10 fois · Autres articles · Rechercher sur le Web · Les 2 versions](#)

[青麻叶结球白菜雄性不育系的转育](#)
 闻凤英, 宋连秋, 王玉龙, 刘晚辉, 赵冰, 庄玉秀, 期刊 · 园艺学报, 2001 · [scholar.iilib.cn](#)
 ... 不育系3A作为不育源,以青麻叶材料黑227为目标亲本,经过杂交、
 自交、兄妹交、测交等转育手段,得雄性不育基因转育 ...
[Cite 11 fois · Rechercher sur le Web](#)

[属间体细胞杂交创建甘蓝型油菜细胞质雄性不育系及其鉴定](#)
 胡琳, 李芸昌, 梅德宝, 方小平, LN Hansen, SE ... · 中国农业科学, 2004 · [cqvip.com](#)
 ... 4~6代的不育单株中发现了2条本线粒体的重组[21]。因而nsa雄性不育
 的遗传决定因素(即细胞质雄性不育基因)有2种可能来源 ...
[Cite 17 fois · Autres articles · Rechercher sur le Web · Les 2 versions](#)

[一个与甘蓝显性雄性不育基因连锁的RFLP标记](#)
 刘玉梅, 方智远, MO McMullen, 庄木, 杨丽梅, 王晓武, ... · 园艺学报, 2003 · [cqvip.com](#)
 园艺学报2003, 30(5):549-553 Aam Hooiaulan'ae Sime'一个与甘蓝显性雄性
 不育基因连锁的RFLP标记刘玉梅方智远Michael D. McMullen2 ...
[Cite 7 fois · Autres articles · Rechercher sur le Web](#)

Auteurs clés: [方智远](#) · [王晓武](#) · [朱英国](#) · [孙日飞](#) · [凌杏元](#)

Gooooooooogle
 

Page de résultats:
 [1](#)
[2](#)
[3](#)
[4](#)
[5](#)
[6](#)
[7](#)
[8](#)
[9](#)
[10](#)
[Suivant](#)