CYBERMetrics
International Journal of Scientometrics, Informetrics and Bibliometrics
ISSN 1137-5019

VOLUME 4 (2000). ISSUE 1. PAPER 2

Mapping University-Industry-Government Relations on the Internet: The Construction of Indicators for a Knowledge-Based Economy

Loet Leydesdorff and Michael Curran
Science & Technology Dynamics, Faculty of Social and Behavioural Sciences,
OZ Achterburgwal 237, 1012 DL Amsterdam, The Netherlands
loet@leydesdorff.net; **mickeyfenn@hotmail.com**
http://www.leydesdorff.net

Abstract

University-industry-government relations can be indicated by using advanced search engines on the Internet. This methodology provides us with opportunities to construct time series, compare among nations, distinguish between the use of national languages as against English, map relational patterns, etc. On the occasion of the Third Triple Helix Conference in Rio de Janeiro (April 2000), we focused on the comparison between Brazil and The Netherlands, with reference to an international baseline. The data can easily be reproduced for other countries. In addition to studying the different domains and their university-industry-government relations, we analyze, among other things, the role of the respective (national) languages. The major findings are: (1) patterns of development are similar in shape, but different in size when comparing Brazil with The Netherlands, and with the international environment; (2) "industry-government" relations are enhanced in the national dimension, while "university-industry" relations are profiled in the (international) "any language" domain; and (3) labels which can be controlled by actors (like title words of web pages) reflect upon (and thus lag behind) structural features of the database

Keywords: Triple Helix, national innovation systems, university-industry relations, S&T indicators

Introduction

The process of the production of scientific knowledge is embedded in contexts of application. Gibbons *et al.* (1994) proposed to distinguish between the traditional format of disciplinarily oriented production of scientific knowledge ("Mode 1") versus

Information Centres
V FP (EU)
S&T in the Web
WORLD SITATION REPORT
WSR (siting)
WSR (sited)
IWR&D-Library

Editor

Updated 31/05/00

an emerging “Mode 2” of scientific knowledge production. In the latter mode the contexts of applications can increasingly be considered as co-constitutive for the development of the research front. The RTD networks of the European Union may provide an example of “Mode 2”: collaboration across institutional and sectorial boundaries has often become conditional for the funding of research. Thus, the reward structure of science and thereby the intellectual organization may also be affected (Whitley, 1984).

While cognitive dynamics can be expected to drive the knowledge production processes mainly in a recursive mode and within the institutions (“Mode 1”), the continuous interaction at the network level would induce “Mode 2”-type of research. “Mode 2” research is organized in temporary projects, it is said to be transdisciplinary, problem-oriented, and the results are evaluated using standards other than those of traditional research. Scientometric indicators were generated for measuring “Mode 1”-type of research outcomes. Can webometric indicators be used for measuring also “Mode 2”?

In this study, we raise the question of whether the reflections of the communications between universities and third parties on the Internet can be used to make this interactive mode of knowledge production accessible to the measurement. Here we focus on the mapping of the structural and institutional relations between universities, industries, and governments as they can be traced using the tags of a search engine. In a follow-up study, we intend to compare this operationalization with the results of traditional scientometric research using, among other things, science and patent citation indices by measuring the same events (e.g., innovations) across different contexts.

In general, one is able to distinguish between citations as indicators of codification within scientific knowledge and other indicators of the relationships between scientific knowledge and its applications (e.g., words and co-words; Callon *et al.*, 1985; Leydesdorff, 1997). Le Pair (1988), for example, found that citations do not provide us with an accurate representation of technological achievements, because knowledge can be built into technological artifacts without necessarily leaving the formal trace of a citation in the scientific literature (Els *et al.*, 1989). Using the Internet, however, one may expect that the transfer of applicational knowledge can also be made visible. Is one able to trace this “Mode 2” of knowledge production and distribution by following the hyperlinks?

We focus on university-industry-government relations because these relations can be considered as a knowledge infrastructure that we have elsewhere called a “Triple Helix of University-Industry-Government Relations” (Etzkowitz & Leydesdorff, 1997; Leydesdorff, 2000). Both university-industry collaborations and triple helix configurations have been the subject of extensive bibliometric research efforts (e.g., Godin & Gingras, 2000; cf. Hicks & Katz, 1997). Leydesdorff & Wouters (1999) suggested that Triple Helix configurations on the Internet can be searched by using hyperlinks between industrial (www.*.com), academic (www.*.edu), and governmental (www.*.gov) texts (cf. Aguillo, 1999).

Boudourides *et al.* (1999) used the advanced search technology of AltaVista for the measurement, because it allows for searching on domain names and using Boolean operators. Following this lead, we were able to analyze the differences and similarities between national systems of innovation as indicated by national domain names (e.g., “.br” or “.nl”) as against the so-called “generic Top Level Domains” (like “.edu” and “.com”). Subsequent questions to be addressed concern the use of national languages versus English, the baseline for international comparisons, the differences in using the various search terms, etc.

Material and methods

Citation analysis has been used to 'map' the manner in which information is shared and passed through printed media. In webometrics *link* analysis could yield similar information about the manner in which information is diffused on the Internet (Clever Project, 1999; Butler, 2000). Rousseau (1997) has called this "sitation" analysis. Similarly, AltaVista's keyword function *text* can be considered as an analogue to scientometric co-word analysis (Callon *et al.*, 1986). By using this command, pages within a specified domain that contain the same textual unit are returned by the search engine. If one wishes to search on only title words one can further specify the keyword function *title*.

The search engines at the Internet offer a rapidly increasing number of search possibilities and there is a growing number of papers about the coverage of the Internet by search engines (e.g., Lawrence & Giles, 1999; Butler, 2000; see searchenginewatch.com for an overview of search engine logics). These search engines cover the Internet only to a limited and variable degree.

In our opinion, the Internet should be considered as an emerging phenomenon that is an algorithmic result of the geometrical representations. From this perspective, the Internet itself remains an hypothetical domain to be indicated by using one or more search engines. Reflexively, we can study the quality of the search engine, for example, by making comparisons among them.

Given these severe limitations in presenting the web data, the quality of the organization of the data within each representation provides an additional criterion for using one search engine or another. Most search engines are user-oriented, but some of them provide analytical tools in so-called "advanced" versions. Among the major search engines, the *AltaVista Advanced Search Engine* provides hitherto superior capacities for combining both date delimiters and a full set of Boolean operators.[1]

By using this search engine one is enable to picture a complex representation at each moment in time and to compare results over the time axis dynamically. However, the dynamic representation is from a hindsight perspective, while the web crawler is continuously rewriting the representation. For this dynamic reason and for the already mentioned static problem of sampling from the population, we cannot claim validity for our inferences beyond the AltaVista domain.

The AltaVista search engine

We searched using four functions available in AltaVista: domain, link, text, and title. "Domain" refers to the number of web pages contained within a single domain (such as a country). Thus, one is able to compare the relative size of the representations of the three helices among national systems. Likewise, the tag "link" enables us to quantify the number of web pages containing a link to another helix; and "text" or "title" corresponds to the number of web pages that share the same text (or title words). Calendar years were used for the time delineation.

Because of our employment of AltaVista, our methodology is limited to its search technology and our results inform us about the structure of this domain. Further information about the advanced search technology can be found in the help link at AltaVista.com. The list in Table 1 shows the available functions and how they work.

Keyword	Function
anchor:text	Finds pages that contain the specified word or phrase in the text of a hyperlink. **Anchor:"Click here to visit garden.com"** would find pages with "Click here to visit garden.com" as a link
applet:class	Finds pages that contain a specified Java applet. Use **applet:morph** to find pages using applets called morph
domain:domainname	Finds pages within the specified domain. Use **domain:uk** to find pages from the United Kingdom, or use **domain:com** to find pages from commercial sites
host:name	Finds pages on a specific computer. The search **host:www.shopping.com** would find pages on the Shopping.com computer, and **host:dilbert.unitedmedia.com** would find pages on the computer called dilbert at unitedmedia.com
image:filename	Finds pages with images having a specific filename. Use **image:beaches** to find pages with images called beaches
link:urltext	Finds pages with a link to a page with the specified URL text. Use **link:www.zip2.com** to find all pages linking to Zip2.com
text:text	Finds pages that contain the specified text in any part of the page other than an image tag, link, or URL. The search **text:graduation** would find all pages with the term graduation in them
title:text	Finds pages that contain the specified word or phrase in the page title (which appears in the title bar of most browsers). The search **title:sunset** would find pages with sunset in the title
url:text	Finds pages with a specific word or phrase in the URL. Use **url:zip2** to find all pages on all servers that have the word zip2 in the host name, path, or filename—the complete URL, in other words

Table 1: Tags in the Advanced Search Engine of AltaVista (from http://doc.AltaVista.com/help/search/search_help.shtml#fancy)

The various keyword functions have a different status from a sociological perspective. While the links and the title words are intentionally chosen by the author of a text, the domain field is generated institutionally. It is attributed by the host machine. The (indexed) text field can furthermore be compared with the free text search. Although words in texts (and titles) can be considered as categories used by authors, one should not expect the meaning of words or co-words to be codified when comparing among texts or over time. Words vary in meaning across texts and the meaning may change upon interaction (Leydesdorff, 1997). Thus, the structure of the network can be expected to carry a meta-narrative, which is beyond the control of individual actors (Burt, 1982).

National domains

In preparation for the Third Triple Helix Conference in Rio de Janeiro (April 2000), we have selected three domains for our query: Brazil, The Netherlands, and a combination with an OR-statement of the generic extensions for commercial sites (.com), governmental sites (.gov), educational sites (.edu), organizations (.org), military (.mil), and internet organizations (.net). These so-called "generic Top Level Domains" or gTLDs (OECD, 1999; EU, 1999) were combined into an international system of reference by using Boolean OR-operators.

The Netherlands was chosen as a country of interest for the comparison because of our affiliation with it. The inclusion of Brazil serves two ends: first, it provides a counterbalance to a small European country by being a large South American economy; and second, Brazil served as the host of the third international Triple-Helix

conference for which this paper was being prepared. In terms of Triple-Helix relations, Brazil can be considered as a so-called "big emerging market," while The Netherlands can be considered as an example of a knowledge-based economy.

The paper elaborates a methodology that can easily be applied to other nation states for comparative reasons. The Netherlands was also chosen for reasons of comparison with Brazil. The combination of generic Top Level Domains provided us with a system of reference for assessing the relative contributions of these nations globally.

The American contribution is embedded in and dominant in these gTLDS, but to a variable extent. The '.gov'-domain, for example, refers almost exclusively to the U.S. government, while the '.com'-domain is increasingly internationalized. The gTLDs provide us with a set, which is neither "global" nor purely "American." The U.K., for example, is relatively underrepresented, since it maintains a similar codification *within* its own (".uk"-)domain. Additionally, there exists a domain ".us". This domain is of the size of the French national contribution. Thus, the contribution of the United States is mainly organized into web pages of various types at the generic Top Level.

Languages

At the global level, English can be expected to prevail as the dominant language. However, national languages play an important role on the Internet. The AltaVista search engine allows for a choice of specific languages or of "any language". We will compare the national language contributions to the "any language" category for each national domain, and compare across national domains using the "any language" definition.

Neither The Netherlands nor Brazil coincides with a single linguistic domain. Portuguese is also spoken in Portugal and in other parts of the world (like the former Portuguese colonies), and Dutch is also a native language in Belgium (Flanders). As with these comparisons between national domains and natural languages, we can distinguish in the case of the gTLDs between using specifically "English" or the option of "any language."

When using Dutch as the search language, we used as search terms: "universiteit," "industrie," and "overheid." The literal translation of "government" into Dutch is "regering," but relations with a "regering" would mean that the relations are maintained with a specific (that is, politically elected) administration. The institution of government is more precisely expressed by the Dutch word "overheid."

In Portuguese, we used "universidade," "industria," and "governo" as search terms. We did not specifically search for "indústria" with an accent since both search terms occur in the case of Brazil. AltaVista to a large extend subsumes "indústria" under "industria." [2]

The longitudinal dimension

Although the Internet has a longer history, browsers (like Mosaic and then Netscape) have only been available since 1992. Before 1993 most files contained only plain text, and one is not able to retrieve hypertext structures from these texts with hindsight. Therefore, the years specified for our query were the last six complete years: 1993-1998. These six points in time create a dynamic data set, which should yield sufficient findings for exploring the options for a dynamic analysis. We did not wish to break the years down into months because of the possibility of seasonal effects.

Note that there are limitations to this methodology due to the search engine logic of the AltaVista technology. AltaVista returns pages that were last *modified* within the

specified search dates. Thus, the system changes dynamically, and insofar as the system is changed, its history is overwritten. The original date stamps are then lost for retrieval.

The possibility of structural change with hindsight accords with our evolutionary perspective: the state of the (constructed) system in the present has analytical priority over its historical construction in the past. We are here interested in the expected information content of the database as a memory trace from the past, and we are not testing the quality of the representation as a historical record (Leydesdorff *et al.*, 1994). The Internet being represented remains an external reference.

In addition to the ongoing reconstruction, the search system itself extends at a high speed by using "spiders." Actually, AltaVista's robots were not scanning the web until 1995. The spiders are continuously crawling in the web and thereby they hit on pages, which have not been found before. These discoveries also affect the relative number of pages in past years. This development is fast and not under control from the user's end. Searches are sometimes not reproducible within a period of ten days or so. Thus, all relevant searches for the sake of a comparison have preferably to be done in a relatively short time span (of a few days only) and one has to control for differences among search results at different moments in time (Rousseau, 1999). All data collected for this study were retrieved between 15 November and 15 December 1999, and we always checked a few days later whether the results were stable.

Results

The overall distributions

Before we turn to university-industry-government relations in detail, let us first provide an overview. How are these three geographical and linguistic domains developing over time on the Internet as windowed by the AltaVista domain? Are there significant differences between Brazil and The Netherlands with reference to the overall system? Is the role of Portuguese and Dutch different in relation to the international environment of these national languages?

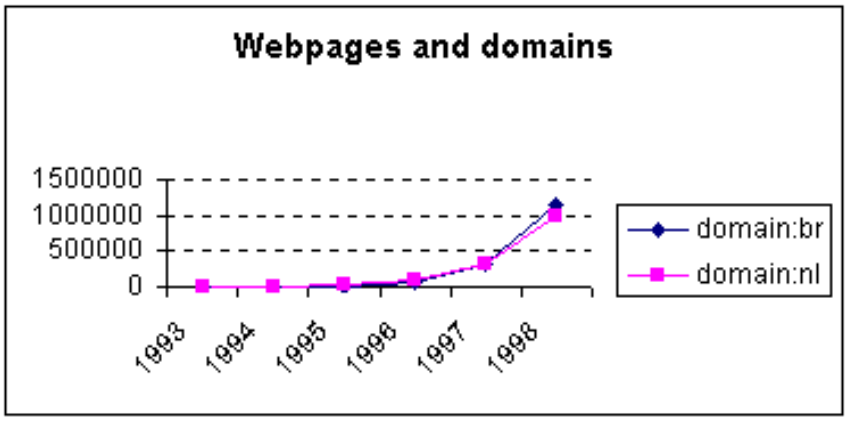

Figure 1: Relative contributions of Brazil and The Netherlands to the Web

First, one observes (in Figure 1) that Brazil and The Netherlands are approximately of the same size in terms of the numbers of webpages carrying the respective domain names of these countries. While Holland had a somewhat speedier take-off on the net, recent years seem to indicate that Brazil is now growing a bit faster. This is illustrated

in Figure 2: here we normalized with reference to the global system of gTLDs (as specified above). Both countries contribute on the order of 3% of the reference system, but the share of Dutch publications is no longer growing in relative terms. (Yet, the Internet itself is still multiplying in size each year).

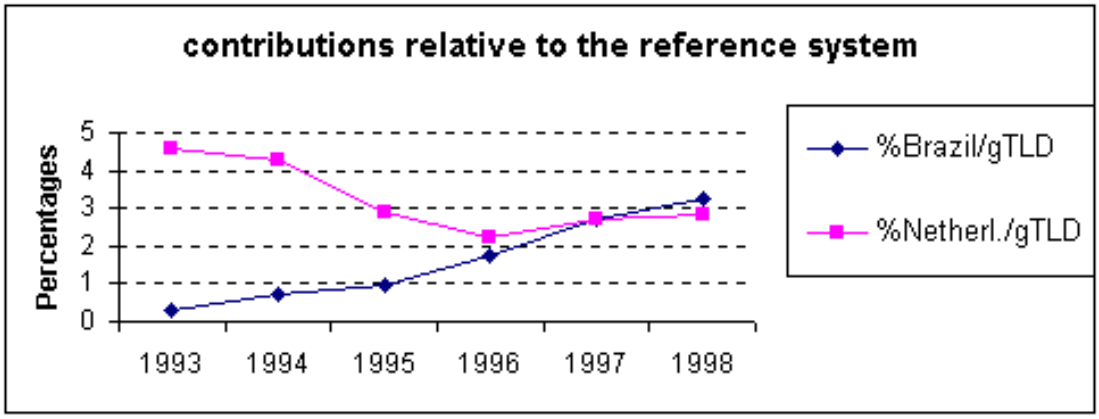

Figure 2: Relative growth of the two national systems with reference to the (OR) combination of generic Top Level Domains

Another difference between The Netherlands and Brazil becomes visible if we plot (in Figure 3) the relative contribution of these countries' web pages to the English-language domain (in the restricted sense of only "English"). Obviously, web pages from both countries have increasingly begun to use English, but The Netherlands' contribution is more advanced in this respect than Brazil's.

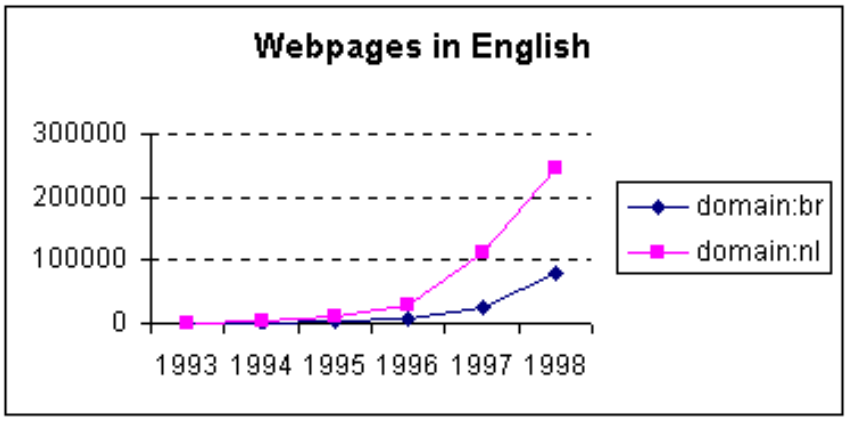

Figure 3: Contributions of Dutch and Brazilian webpages to the English language domain

Figure 4 compares the papers in these respective domains using the national languages as a percentage of the contributions in the "any language" domain. For reasons of comparison, English is now considered as the "natural" language for the generic Top Level Domains (as defined in the methods section above).

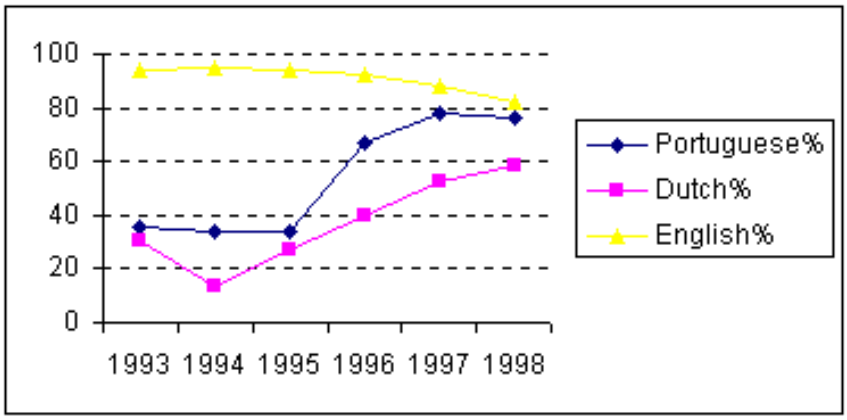

Figure 4: Percentage of Brazilian, Dutch and gTLD web pages using native languages, respectively.

One should keep in mind that the overall use of the Internet is nowadays predominantly commercial. It is somewhat interesting to speculate about the importance of the home markets given the trends visible in Figure 4.

The Triple Helix of university-industry-government relations

Although we have collected only a limited amount of data, the number of possible comparisons and analyses is already large. One can compare among the countries, over time, in terms of using different languages, and in terms of bilateral and trilateral relations, using the various options of the search engine.

Let us first explore the differences between Brazil, The Netherlands, and the combined gTLDs in terms of the number of hits for the Triple-Helix categories. Figures 5a, 5b, and 5c provide the comparison using the free text keywords “university,” “industry,” “government,” and the combinations of these three sets with Boolean AND-operators.

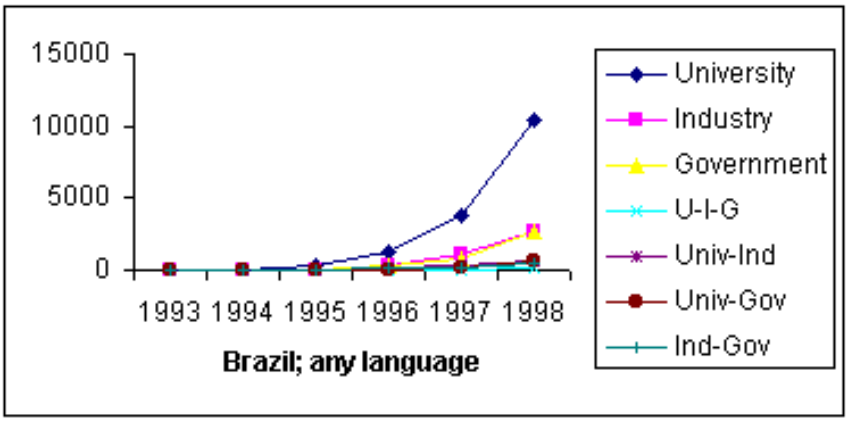

Figure 5a: Triple Helix components and combinations in the Brazilian domain

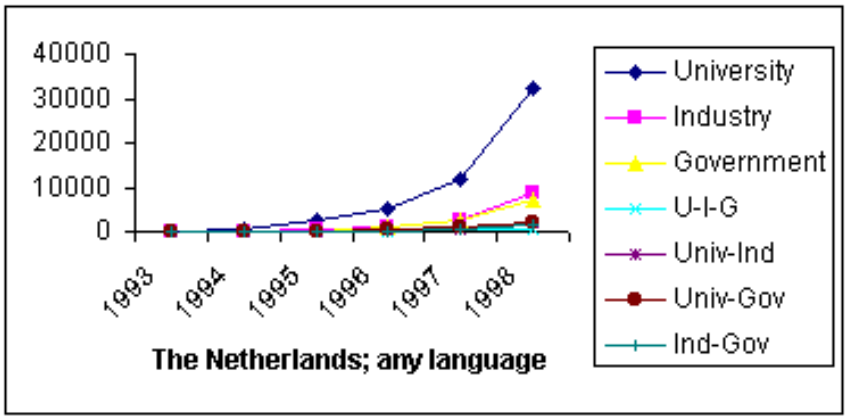

Figure 5b: Triple Helix components and combinations in the Dutch domain

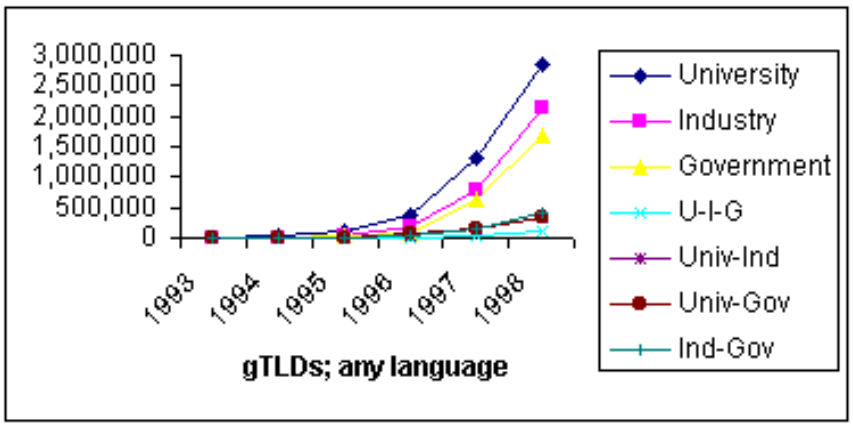

Figure 5c: Triple Helix components and combinations in the reference domain of gTLDs

Note that the pattern is almost identical for the cases of Brazil and The Netherlands, but the respective sizes are considerably different. All contributions are more than twice as large in the case of The Netherlands compared with the Brazilian contributions. In the case of the reference system of the combined gTLDs (Figure 5c), the words "government" and "industry" are more dominant than in the Brazilian or Dutch set. In these latter two cases "industry" and "government" seem still in their take-off phase, while "university" is the leading keyword.

One reason for these differences in the recall using the various search terms seems to be the international orientation of the term "university." The words "government" and "industry" are not so strongly internationalized. When using the national languages, however, "university" remains the most important term, followed by "government." We return to this issue below. Note that the various regression lines all indicate exponential growth patterns (with $r > .95$).

The Triple Helix relations

Let us now focus on the trends based on searches combining two and three of the search terms "university," "industry," and "government." Thus, we zoom in on the lower lines in Figures 5a, b, and c. Figures 6a, b, and c explore the fine structure of these Boolean searches.

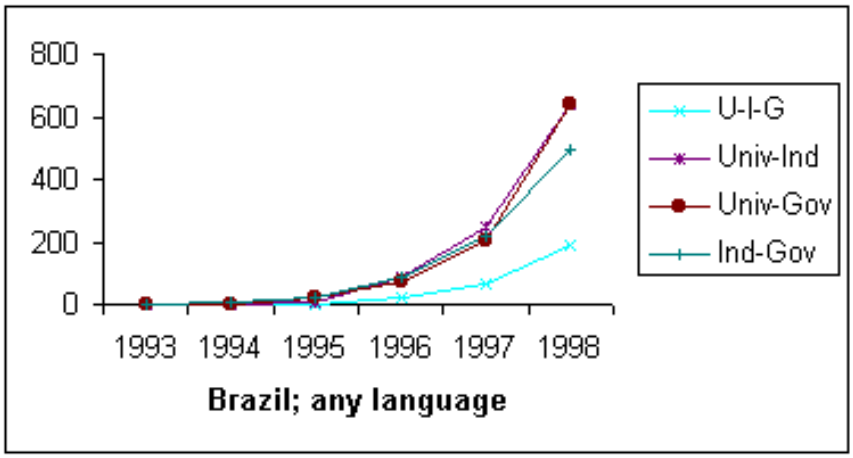

Figure 6a: Bilateral and Triple Helix relations in the case of Brazil

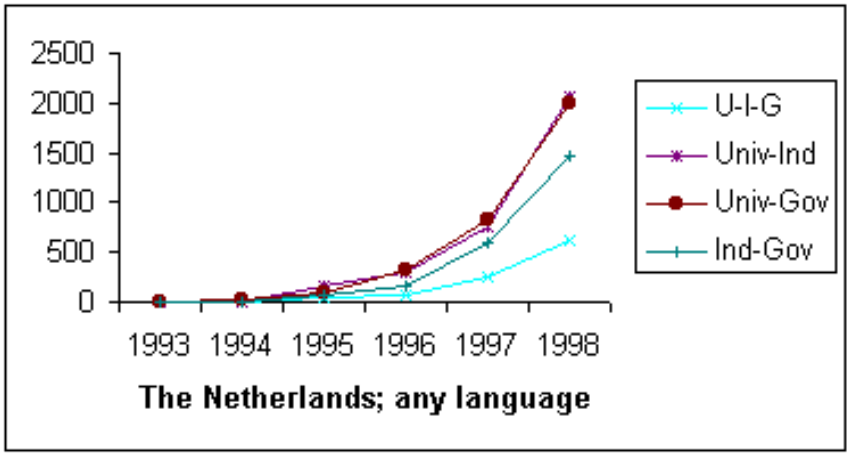

Figure 6b: Bilateral and Triple Helix relations in the case of The Netherlands

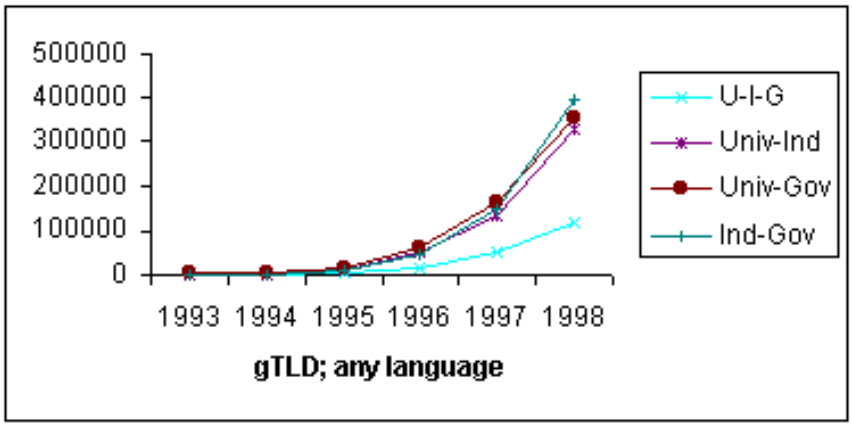

Figure 6c: Bilateral and Triple Helix relations in the case of the reference set of combined gTLDs

The correspondence between the picture for Brazil and for The Netherlands is again striking, and the effects of scale are even more pronounced. It is now visible that "industry-government" relations lag behind when compared to the global data set. As we shall see in Figures 7 —which depict these same relations but using the national languages— this phenomenon can be explained as a language effect.

"Industry-government" relations are more nationally embedded, while "university-industry" relations are relatively more international. In the global set of gTLDs the differences between the two pictures (6c and 7c) are marginal.

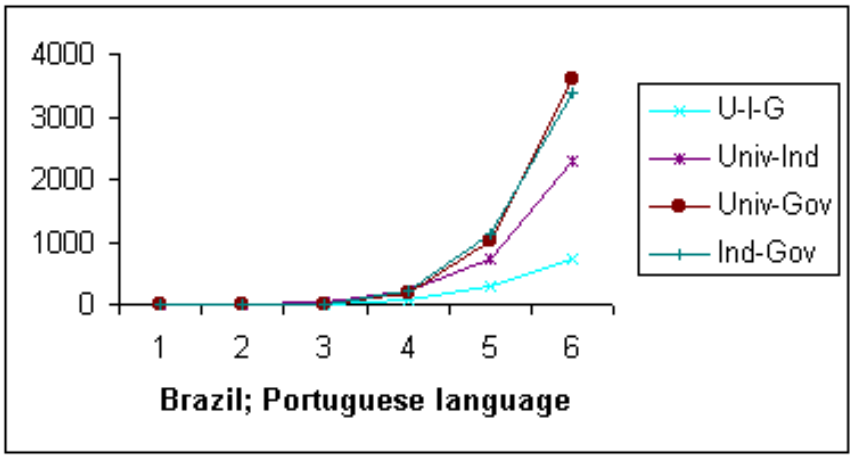

Figure 7a: Bilateral and Triple Helix relations in the case of Brazil using only Portuguese

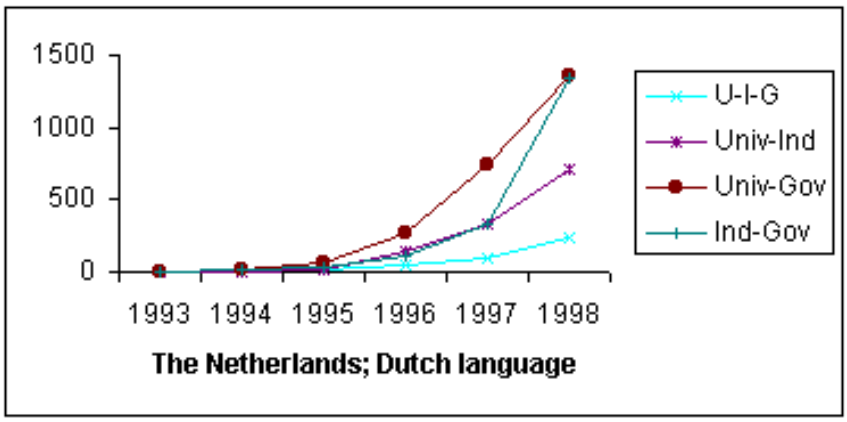

Figure 7b: Bilateral and Triple Helix relations in the case of The Netherlands using only Dutch

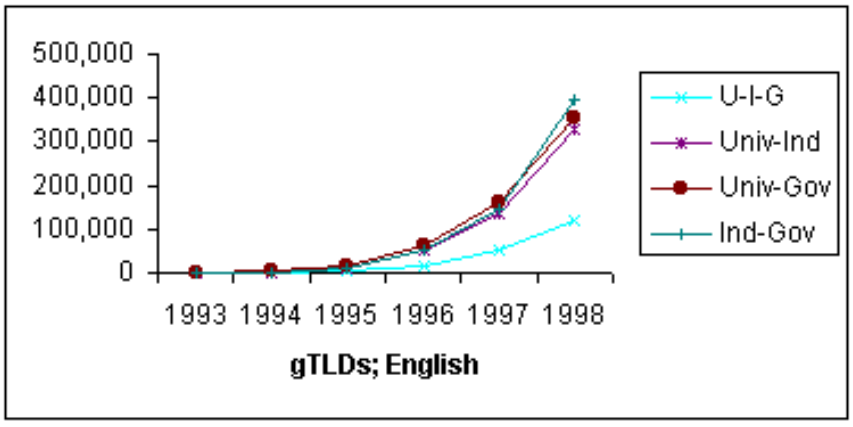

Figure 7c: Bilateral and Triple Helix relations in the case of gTLDs using only English

Trilateral ("university-industry-government") relations behave similarly in the three

datasets. (The lower level is an effect of using two Boolean ANDs.) Note that the Portuguese dataset for Brazil is more than twice as large as the Dutch dataset for The Netherlands, while the reverse was the case in the international datasets discussed above.

In sum, the data indicate that national languages are less prevalent in university-industry relations than in webpages relating industry with government. Portuguese functions as a more important language in the knowledge infrastructure of Brazil than does English, while the situation is reversed in The Netherlands. Note that the figures cannot be compared directly because the search terms are different using the various languages.

Links and hypertext

The Internet enables the user to move from one domain to another since the documents are provided with links by the authors. The links can be considered as the carriers of the hypertext at the Internet, and therefore this relational operator provides the added value of the net. Indeed, AltaVista specifies “link:” as a specific search term. We now turn to the results of using the same searches as above in terms of domains (that is, Brazil, The Netherlands, and the combined gTLDs), but using the terms “university,” “industry,” and “government” as links (instead of free text). In other words, these searches were constrained by a Boolean operator of the following format: (domain:nl AND link:university AND link:industry).

With the exception of the much larger set for the gTLDs, links, which contain both “university” and “industry” or any other bilateral combination of the two, are very scarce. Both in the Dutch and in the Brazilian domain, there is some activity in the combination “link:industry AND link:government” when using the national languages in 1997 and 1998. Brazil provides six such sites in 1997 and ten in 1998; The Netherlands, one in 1997 and eight in 1998. In the gTLDs the number of hits for the combination “industry-government” is an order of magnitude larger than in the other two combinations (Figure 8).

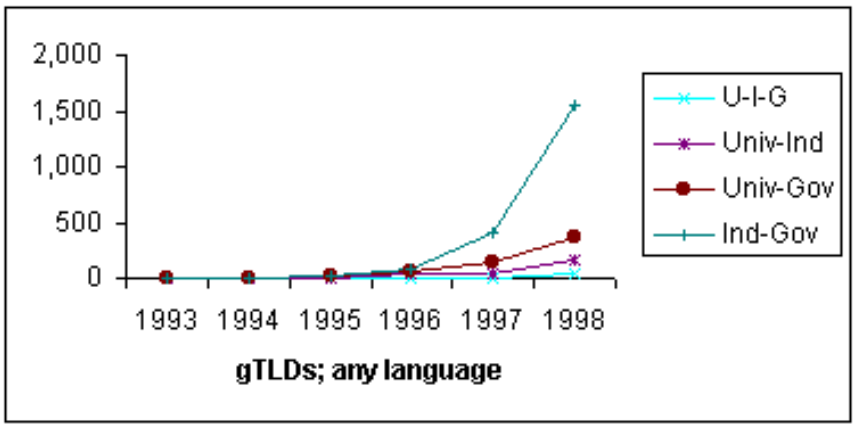

Figure 8: shared links between “university,” “industry,” and “government” at the level of the reference set of combined gTLDs

Inspection of the retrieved sites in the case of The Netherlands shows that these sites are maintained by organizations that service both sectors, such as professional organizations, regional planning offices, and sometimes large industries or banks which incorporate these services. In summary, this infrastructure of the national economy focuses on “industry” and “government,” while “universities” have hitherto been a less

interesting clientèle for this particular layer. Let us note that this service structure seems to be locally oriented and not yet internationalized as far as Holland and Brazil are concerned.

From the perspective of our research question, the use of hypertextual links as an indicator of university-industry-government relations was less fruitful than the approach of using the domains in combinations with (free) texts as search operators. Why? First, one observes that the knowledge infrastructure is more internationalized than the infrastructures of the national economies. Thus, the two layers do not necessarily couple in terms of hypertextual links. We have seen this indicated above by the relative decrease of university-industry relations shown if we move to the national languages (Figures 7a, 7b, and 7c). Second, keywords may not be good indicators for links, for example, if the latter are use abbreviations (like "univ"). Finally, if a specific domain that has an active interest in serving two or three of the sectors involved is now manifesting itself increasingly at the Internet, it remains a carrier different from the processes of mutual interest formation and the emerging overlay of university-industry-government relations. This observable domain of an emerging service sector can perhaps be considered as a retention mechanism for the increased relationships among the networked agencies.

A theoretical reflection can be added to the above conclusion. In general, the study of the links points us to the analysis of domains, while the study of domains raises questions with respect to the relations among domains. These two are related as the warp and the woof of a fabric. Thus, if one is interested in the study of the relations, one may wish to focus on the domains as the units of analysis and consider the relations as variables. How is the communication between domains enhanced by links, and how does this process enable the participants to become embedded in an emerging hypertextual dimension?

Using the "text:" anchor

The differences between using the free text search or the "text:" delimiter available among the Advanced Search options of AltaVista are marginal in terms of the number of hits. The two resulting sets are correlated above the .99 level in all six of the sets of domains and languages we have studied. However, the two underlying sets are not necessarily the same. Some sites may be indexed under the "text:" delimiter, but not retrievable using the free text option, and vice versa. For example, free text searches include title words, which do not always or necessarily occur in the body of the text.

Note that the free text search option in the Advanced Search is not identical to the Main Search option of AltaVista, that is, the search without Boolean operators. For example, if one searches for "university" using this main engine, one generates 27,321,230 hits (on 12 December 1999). At this same date, a free text search using the advanced search technique provided only 12,615,296 hits, while a search with "text:university" provided 12,768,737 hits. The "OR" statement between the latter two generated the same number of hits, suggesting that the free text search provides us with a subset of the controlled text search. However the search "university AND NOT text:university" provided another 12,168 hits indicating an inconsistency of one per thousand. We found inconsistencies of this order of magnitude throughout our searches.

We are not able to explain these differences. Extending our search by using Advanced Search with the wildcat "universit*", that is, including the plural "universities" within the search domain, provided 14,547,561 hits, still falling short by approximately 13 million hits if compared to the Main Search result. However, the quality of the search engine was not the research question of this study. For the research presented here (using the Advanced Search engine), the differences between searching with free text

or with controlled text terms seemed negligible.

Using the "title:" anchor

A final field which merits our attention is the possibility of searching with the tag "title:". In HTML authoring title words can be added deliberately in order to facilitate searching the webpage. With standard browsers, title words do not appear when reading the page. Thus, they are intentional "actor categories" to a greater extent than any of the other searchable tags.

Using the tag "title:", we found a later development of our search categories than in the case of the full text and free text searches. This lag is even more pronounced than in the case of the other deliberately constructed text-element, that is, "link:". The titles with these headings do not really appear before 1996 and the combinations remain almost empty for both the Netherlands and Brazil, irrespective of the use of language. Title words are more pronounced in the national languages of both these countries than in English. In the case of the set of combined gTLDs there is some activity in the bilateral relations (using a single AND-operator), but hardly any when the three search terms are combined (using two ANDs).

In other words, authors are inclined to indicate their own institutional domains in terms of the sector. Relations with other institutional agencies are visible within the text itself, and to a much smaller extent in the textual codifications (like title words and hyperlinks). The development of linkages within and among texts seems to precede the deliberate insertion of these linkages by reflexive agencies. While the meta-narrative is socially constructed at a macro-level, the network tends to drive the micro-agencies in choosing their preferred terminologies.

Conclusions

We have demonstrated how it is possible to study Triple Helix relations between universities, government, and industries using the Advanced Search options of the AltaVista search engine. The main findings are:

· Brazil and The Netherlands are of comparable size as represented in terms of the number of sites on the Internet. While The Netherlands seems to have stabilized its percentage share, Brazil is still growing as a relative (percentage) rate.

· The Netherlands is more advanced than Brazil in terms of websites in English. In the case of Brazil, Portuguese is a far more important medium than English on the Internet. In general, the number of websites in national languages seems to have become relatively more important in recent years.

· Of the categories "university", "industry," and "government," "university" is the leading term in all comparisons (that is, across countries and languages). Within the national domains of both Brazil and The Netherlands, "government" is more important then "industry." Of course, this reflects the national character of the respective governments.

· In terms of university-industry-government relations, one should first note the size effects of an order of magnitude. The AND-operator introduces a strong restriction. This restriction is further reinforced when two AND operators are used to indicate trilateral relations. In the case of Brazil, these selections are particularly dramatic in the international domain. Using Portuguese search terms, the relations are five times as large as when using the "any language" domain with English search terms. Thus, here again the differences between the countries become pronounced.

· The patterns, however, are rather similar when the size effects are normalized: industry-government relations' lag behind when viewed from the international perspective, and industry-university relations' lag behind when viewed from the national perspective. This accords with our intuition that industry-government relations can be considered as indicators of national economies, while university-industry relations are increasingly knowledge intensive.

· All trend lines exhibit growth patterns, which are exponential. However, this may also be an artifact of the underlying growth of the Internet as a medium.

The present research has taught us that it is possible to measure Triple-Helix relations at the Internet and to compare among institutional agencies. From a theoretical point of view it is important to note that these relations among textual units at the Internet can be considered as less codified counterparts of scientometric distributions retrieved in terms of co-word, co-authorship relations, etc. While the scientometric distributions refer to codified communications, Internet relations are not controlled and are mainly based on free text.

By comparing webometric and scientometric results, one seems able to operationalize the differences between "Mode 1" and "Mode 2" types of scientific knowledge production (Gibbons *et al.*, 1994). However, one should proceed carefully, since the relations were here defined mainly at the institutional level, while the scientometric indicators can be codified in terms of their meaning for developments at the level of cognitive fields and specialties (Collins, 1985; Leydesdorff, 1998). The distinction between a "Mode 1" and "Mode 2" in the production of scientific knowledge, however, was at the epistemological level, that is, as a consequence of differences in the substance of communication. The lack of codification may make it difficult to induce cognitively defined domains that are equivalent to disciplines and specialties, and yet based on a reflexive analysis of the empirical data, that is, without forcing the data into *a priori* classifications.

** The authors are grateful to Moses Boudourides (mboudour@upatras.gr) and two anonymous referees for valuable suggestions. A previous version of this paper was presented at the Third International Conference of "The Triple Helix of University-Industry-Government Relations" in Rio de Janeiro, April 2000.*

References

Aguillo, Isidro (1999). "Statistical Indicators on the Internet: The European Science-Technology-Industry System in the World-Wide Web," at <http://diotima.math.upatras.gr/weborg/aguillo2> (9 November 1999).

Boudourides, Moses A.; Sigrist, Beatrice & Alevizos, Philippos D. (1999). "Webometrics and the Self-Organization of the European Information Society" at <http://hyperion.math.upatras.gr/webometrics> (26 October 1999).

Burt, Ronald S. (1982). *Toward a Structural Theory of Action,* Academic Press, New York, 1982.

Butler, Declan (2000). "Souped-up search engines". *Nature*, Vol. 405, 11 May 2000:112-115.

Callon, Michel; Law, John & Rip, Arie (eds.). *Mapping the Dynamics of Science and Technology*. Macmillan, London (1986).

Clever Project [Soumen Chakrabarti, Byron Dom, S. Ravi Kumar,

Prabhakar Raghavan, Sridhar Rajagopalan, Andrew Tomkins, Jon M. Kleinberg, David Gibson]. (1999). "Hypersearching the Web". *Scientific American*, June 1999, 44-52.
<www.sciam.com/1999/0699issue/0699raghavan.html>
<http://simon.cs.cornell.edu/home/kleinber>

Collins, Harry M. (1985). "The Possibilities of Science Policy", *Social Studies of Science*, 15 (1985), 554-558.

Etzkowitz, Henry, and Loet Leydesdorff (eds.), *Universities in the Global Knowledge Economy: A Triple Helix of University-Industry-Government Relations*, Cassell Academic, London, 1997.

European Commission / ETAN, *Transforming European Science through Information and Communication Technologies: Challenges and Opportunities of the Digital Age*. European Communities, EUR 18916, at <http://www.cordis.lu/etan/src/topic-8.htm>, Luxembourg, 1999.

Gibbons, Michael; Limoges, Camille; Nowotny, Helga; Schwartzman, Simon; Scott, Peter and & Trow, Martin (1994). *The new production of knowledge: the dynamics of science and research in contemporary societies*, Sage, London, 1994.

Godin, Benoît & Gingras, Yves (2000). "The place of universities in the system of knowledge production," *Research Policy* 28 (2000), 273-278.

Hicks, Diana M. & Katz, J. Sylvan (1997). "The changing shape of Bristish industrial research," *STEEP special report no. 6*, SPRU, 1997.

Lawrence, Steve, and C. Lee Giles (1999). "Accessibility of information on the web", *Nature*,400 (8 July 1999):107-109.

Le Pair, Cees (1998). "The citation gap of applicable science", in: A.F.J. van Raan (ed.), *Handbook of Quantitative Studies of Science and Technology*, Elsevier Science/North-Holland, Amsterdam, 1988, 537-553.

Leydesdorff, Loet (2000). "The Triple Helix: An evolutionary model of innovations". *Research Policy*, 29(2) (2000), 243-256.

Leydesdorff, Loet (1997). "Why Words and Co-Words Cannot Map the Development of the Sciences". *Journal of the American Society for Information Science*, 48 (1997), 418-27.

Leydesdorff, Loet (1998). "Theories of Citation?". *Scientometrics*, 43 (1998), 5-25.

Leydesdorff, Loet & Wouters, Paul (1999). "Between Texts and Contexts: Advances in Theories of Citation". *Scientometrics*, 44 (1999), 169-182

Leydesdorff, Loet; Cozzens, Susan E. & van den Besselaar, Peter (1994). "Tracking Areas of Strategic Importance using Scientometric Journal Mappings. *Research Policy*, 23 (1994), 217-29.

OECD (1999). *Internet Infrastructure Indicators. OECD, Paris.*
<http://www.oecd.org/dsti/sti/it/cm /prod/tisp98-7e.htm>.

Rousseau, Ronald (1997). "Sitations: an exploratory study". *Cybermetrics*, 1 (1997), Issue 1, Paper 1 at
<http://www.cindoc.csic.es/cybermetrics/articles/v1i1p1.html>.

Rousseau, Ronald (1999). "Daily time series of common single word searches in AltaVista and NorthernLight" *Cybermetrics*, 2/3 (1999), Issue 1, Paper 2. <http://www.cindoc.csic.es/cybermetrics/articles/v2i1p2.html>.

Van Els, W.P.; Jansz, C.N.M. & le Pair, C. (1989). "The citation gap between printed and instrumental output of technological research: the case of the electron microscope". *Scientometrics*, 17 (1989), 415-425.

Whitley, Richard D. (1984). *The Intellectual and Social Organization of the Sciences.* Oxford University Press, Oxford, etc., 1984.

Notes:

1 The Powersearch engine of *Northern Light* offers also search abilities for combining time delimitations with domain specifications, but the latter are less clear in terms of the analytical organization.

2 We found the following values for 1998 (delimited with domain:br on 15/12/1999):

	Free text	title	text	link
Indústria	30632	796	30623	196
Industria	32436	836	32297	305
OR	36048	865	35955	305

Received 20/March/2000
Accepted 31/May/2000