Pergamon

Technovation 23 (2003) 533–544

technovation

www.elsevier.com/locate/technovation

Growth of embedded software related patents

D.H. McQueen [a,*], H. Olsson [b]

[a] *Firma DMQ Business, Mäster Bengtsgatan 10, S-412 62 Göteborg, Sweden*
[b] *Awapatent AB, Box 11394, S-404 28 Göteborg, Sweden*

Abstract

The distribution of embedded software related patent applications across 118 IPC patent classes has been determined for 1988, 1993 and 1998 using a bibliometric technique. The patent applications were identified using search words based on patent claims concerning various aspects of embedded software. The number of these patent application assignments each year increased at an annual rate of about 17% while the distribution over IPC patent classes narrowed only slightly. In terms of patent classification sections, in 1998 over 41% of all embedded software related patent application assignments were in section G (Physics) while section H (Electricity) contained another 36% of them. In terms of patent classes, the most important patent classes are H04 (Electric communication technique) and G06 (Computing, calculating, counting) accounting for a little more than 40% of all embedded software patent application assignments. In only two patent classes, H04 and G06, is there a majority of embedded software related patent applications.

Since patent application assignments are mainly according to the character of the problem to be solved and its solution, rather than the field of the application of the solution found, these statistics relate to the technical problems solved by the inventions. This is in contrast to economic statistics on the distribution of embedded software over branches of industry that characterizes the application of the problem solutions.

Keywords: Patent; Embedded software; Bibliometric

1. Introduction

The importance of information technology has increased significantly over the last decades. Information technology in the form of embedded software is now to be found in products that traditionally only included mechanical and electrical components. Examples are cars, medical equipment such as patient monitoring systems and home appliances such as freezers and microwave ovens. These examples represent old fields in which the potential of information technology is beginning to be realized. The growth of information technology has also lead to the emergence of new fields based on information technology. Examples are the game industry and simulators. There is a study presenting estimates of the computer software content in six industrial sectors in Sweden (McQueen et al., 1998). The percentage of sales in each sector that can be ascribed to software was used to characterize the software content. According to the estimates in the study the percentage software in the machinery sector (mainly due to CNC machines and industrial robots) increased from essentially zero in 1981 to 8% in 1995 (real annual growth rate, inflation corrected, 24%). The transportation equipment sector (including aviation and rail transportation) had a corresponding increase from 1 to 8% (real annual growth rate 18%). Information technology in the electronics sector (excluding telecommunications and computers) rose from 3 to 15% (real annual growth rate 18%). In the telecommunications sector the percentage software increased from 5 to 60% (real annual growth rate 22%). Even if the methods used to obtain these estimates are not very accurate or totally reliable, it is clear that the importance of information technology is increasing steadily (average annual growth rate in these sectors about 21%) on a wide front. For economic, educational and social reasons, it would be good to know more about this growth.

Since information technology or embedded software is distributed over many branches of industry it would

* Corresponding author. Tel./fax: +46-31-203302.
E-mail address: douglas@minmail.net (D.H. McQueen).

0166-4972/02/$ - see front matter © 2002 Elsevier Science Ltd. All rights reserved.
doi:10.1016/S0166-4972(02)00011-1

be useful to know whether or not there is a correspondingly wide distribution of patent assignments over patent classes and whether this has always been so. This could make a difference in how engineers in different fields are trained to deal with embedded software as well as how intellectual property and patent experts should be trained in different fields. Since information technology is the bearer of the 'new economy' this would tell us more about where this new economy might be found. In what patent classes is the percentage of embedded software patents highest? This sort of information would aid in tracing the historical development of embedded software invention and innovation. How does this compare with the growth of the use of embedded software in different branches of industry? Because patent applications are closely related to development effort and innovation, their number can be used to indicate where investments are being made in embedded software related innovations. By identifying the country of origin of the embedded software patents it would be possible to compare the growth of embedded software related inventions in different countries. It would be interesting to compare data on software related patent applications in general to the patent profiles of firms like Ericsson and Nokia, for example. It would also be interesting to know if the pace of software patenting matches the growth of embedded software and information technology in industrial products or whether fewer and fewer patents protect more and more economic wealth or more and more computational power, for instance.

From the aforementioned, it is clear that there is a considerable interest in analyzing patent documents to obtain more information about the growth and distribution of information technology and/or embedded software. Here we interpret information technology as the 'acquisition, processing, storage and dissemination of information in all its forms (auditory, pictorial, textual and numerical) through a combination of computers, telecommunication, networks and electronic devices' (ENB CAL Project (1989), adapted from MacMillan Dictionary of Information Technology (1985) as found at [www.shef.ac.uk/nni/academic/N-Q/nm/misc/informal/bms/s1main/sld005.htm]). Much information technology cannot be protected by patents, partly because it often does not include a technical contribution to the art, as will be discussed subsequently. 'Computer technology' has been defined as 'the use of computers and telecommunications for the processing and distribution of information in digital, audio, video and other forms' (Academic Press Dictionary of Science Technology, 1996). This is a somewhat narrower definition than that used for information technology but still somewhat broader than the following definition of embedded software. Embedded software might be defined as 'microprocessor-based programs for controlling electromechanical systems'. This more producer/product oriented definition includes an explicit connection to the physical world and thus to technical contributions to the art. Of course, for embedded software to be patentable (statutory) it must also be new and include an inventive step. In the following our discussion based on patent application statistics will more closely correspond to the development and growth of embedded software than to information technology as a whole or to computer technology because of the way patentability criteria are applied by European patent offices.

According to the European Patent Convention Article 52(1) "European patents shall be granted for any inventions which are susceptible of industrial application, which are new and which involve an inventive step." An invention is industrially applicable if it can be used in any kind of industry where 'industry' is to be understood in a broad sense as including any physical activity of technical character. This means that in Europe a patentable invention must have technical character. Embedded software generally fulfills this criterion while information technology may or may not fulfill it.

In the US a patentable invention is new, unobvious and useful. The usefulness or utility requirement is less stringent than the European technical character requirement since computer programs can be useful without necessarily being of technical character. Thus some computer programs or information technology may be patentable in the US but not in Europe. In the present discussion the term embedded software is used in connection with European patents and the terms information technology or computer programs are used in connection with US patents.

In the following we use a bibliometric search technique and the international patent classification (IPC) system to characterize the growth and distribution of embedded software related European Patent Office (EPO) patent applications. The development and characteristics of the search profile used to identify embedded software related patent applications is described in detail. The results of applying this search profile to EPO patent application abstracts are then presented. The results show that most embedded software related patent application assignments are in classes G06 and H04 according to the problems solved by the inventions and that the average annual growth rate of these patent applications is about 17%, comparable to the annual growth rate of the economic value of embedded software in industrial products but significantly higher than the growth rate of EPO patent applications in general, around 10.5% per year.

2. Patent statistics, innovation studies and bibliometric searches

Before going into more detail concerning patents, innovation and bibliometrics it might be useful to briefly outline the context of the discussion. Most inventions are not patented, for several reasons. They may not be technical, new or perhaps inventive. Other patentable inventions may not be patented for economic or other reasons. In spite of all this, patent applications are often taken as indications of inventive activity. Further, most patented ideas never generate much income. The distribution of patent profitability is extremely biased or skewed with a very few patents standing for the great majority of patent related income, as pointed out by Scherer (Scherer, 1965a, b). Still, considerable effort has been expended to relate numbers of patents to corporate sales, value added or profits. Another indicator or measure of innovative activity, at least in large corporations, is R&D effort. Therefore, relationships between patent applications or granted patents and R&D expenditures are often studied.

In the present case it is natural to compare embedded software patenting activity to the growth of the semiconductor market or of the microprocessor market. These markets have grown and continue to grow at high rates, even in terms of constant dollars. Corresponding R&D expenditure (often about 20% of value added in high tech industries) has also grown at a high rate. However, the computing power delivered by the semiconductor industry grows at a much higher rate since memory size and clock speed continue to increase phenomenally. Thus while it seems natural to compare numbers of patent applications to semiconductor or microprocessor R&D expenditures or sales, it should be remembered that this is quite different to comparing numbers of patent applications to delivered computing capacity. This distinction between economic value and technical capacity is especially important in the case of embedded software since computing capacity grows many times faster than semiconductor or microprocessor sales.

In spite of the previously mentioned problems, patents and patent applications have been studied extensively for many years by many researchers and professionals. A large body of published work concerns patents and/or patent applications as indicators of the development and/or spread of particular technologies (Pavitt, 1982; Patel and Pavitt, 1987; Narin and Olivastro, 1988; Pavitt, 1988; Lennon, 1994; Ernst, 1995; Archibugi and Pianta, 1996; Jacobsson et al., 1996; Joly and de Looze, 1996; Liu and Shyu, 1997; Ernst, 1998; Grupp and Schmooch, 1999) or of technology in general (Griliches, 1981; Patel and Pavitt, 1987; Narin and Olivastro, 1988; Pavitt, 1988; Pavitt and Patel, 1988; Griliches, 1990; Chakrabarti, 1991; Archibugi and Pianta, 1996; Kondo, 1999; Grupp and Schmooch, 1999). In their review article Archibugi and Pianta (1996) discussed strengths and weaknesses of patent statistics as measures of technical development, pointing out several of the weak links in the chain from technical idea to industrial innovation and production. For instance, not all inventions are patentable and not all patentable inventions are patented. Propensity to patent can vary with firm size, industry branch and country. They compared the use of patent statistics with innovation surveys as means of assessing technical change, thus making some of the weaknesses of patent statistics in this connection even more clear. For instance, only about a half of all patents are ever used by their owners. In a previous review article Pavitt (1988) discussed assumptions behind various uses of patent statistics and interpretations of the results of their use. He assumed that patent statistics reflect both inventive and innovative activities, an assumption not necessarily shared by others. He pointed out that while patent statistics are difficult to compare with industrial statistics it has been found that the relation between R&D expenditure and patenting activity is strongly dependent on industrial sector. Connections between patenting and sales, exports and other economic measures, both at corporate and industry level, have been established and discussed at length for many years (Pavitt, 1982; Griliches, 1990). In his survey Griliches (1990) discussed difficulties in assigning patent applications to patent classes and in relating patent data to industrial economic data. He also pointed out that patent offices affect the numbers of patent applications that are approved in a given time period through various policy decisions and through variations in their capacity to examine the applications. Thus the number of patents granted need not follow the number of patent applications. By way of illustration of this point it can be noted that during the 1990s the number of EPO patent applications increased at an annual rate of about 10.5% while the number of EPO patents granted decreased at an average annual rate of about 4.9% per year.

Other work concerns the use of patent or patent application statistics, that is, methods of gathering patent/patent application statistics and methods of evaluating them (Pavitt, 1988; Pavitt and Patel, 1988; Griliches, 1990; Ernst, 1995; Archibugi and Pianta, 1996; Joly and de Looze, 1996; Liu and Shyu, 1997; Grupp and Schmooch, 1999). In this context it is clear that patent applications are a different measure of innovative activity than are granted patents (Griliches, 1990). In recent years the EPO databases have become quite valuable for studies of patenting in Europe since patent application examination is the same for applications from all countries instead of country dependent, as previously (Griliches, 1990; Grupp and Schmooch, 1999). Furthermore, European patent applications are made publicly available 18 months after the patent application priority date, in contrast to US patents which are made public only after being granted. Various statistical bases can be used for the studies, among them patent classes, firms and patent citations (patent to patent) (Patel and Pavitt, 1987; Narin and Olivastro, 1988; Chakrabarti, 1991; Lennon, 1994; Ernst, 1995; Archibugi and Pianta, 1996; Jacobsson et al., 1996; Joly and de Looze, 1996; Liu and

Shyu, 1997; Patel and Pavitt, 1997; Ernst, 1998; Malerba and Orsenigo, 1999). Advantages of using patent application statistics rather than patent statistics are that the patent application comes first in time, that is, as near as possible to the moment of invention, and that the statistics are not directly affected by patent office examination policies (Griliches, 1990).

A third group of publications concerning patent and patent application statistics focuses on the results of the statistical analysis, for instance tracing the development of a technology or an industry or even of a country (Pavitt, 1982, 1988; Patel and Pavitt, 1987; Pavitt and Patel, 1988; Chakrabarti, 1991; Lennon, 1994; Ernst, 1995; Archibugi and Pianta, 1996; Jacobsson et al., 1996; Patel and Pavitt, 1997; Malerba and Orsenigo, 1999). Relationships between patent data and R&D data, firm sales and/or number of employees have been investigated (Griliches, 1981; Pavitt, 1982; 1988; Pavitt and Patel, 1988; Griliches, 1990; Ernst, 1995; Archibugi and Pianta, 1996; Jacobsson et al., 1996; Ernst, 1998; Kondo, 1999), often with inconclusive results. Archibugi and Pianta (1996) pointed out that neither R&D data nor patent data give comparable results for small and large firms, partly because standard methods of collecting R&D data underestimate the amount of R&D carried out in small firms. In some fields small firms make relatively large contributions to technical innovation while in other fields they appear to be disadvantaged. Especially in Europe small firms, in general, may be disadvantaged by the high cost of European patents.

From the aforementioned, it is clear that patent statistics are a measure of the innovation process but the exact nature of this measure and its accuracy and reliability are somewhat clouded by a number of factors. This is in spite of an extensive published literature concerning patents as technical and economic indicators of which the previously referenced work is but a very small part and, therefore, to be considered illustrative rather than anything even approaching exhaustive. Four of the works are surveys or reviews: Griliches' 'Patent statistics as economic indicators: a survey' (Griliches, 1990), Archibugi and Pianta's 'Measuring technological change through patents and innovation surveys' (Archibugi and Pianta, 1996), Pavitt and Patel's 'The international distribution and determinants of technological activities' (Pavitt and Patel, 1988) and Pavitt's 'Uses and abuses of patent statistics' (Pavitt, 1988). In the present context the article by Lennon (1994) concerning US patents and computer software is of particular interest. Lennon made use of US patent class 364 (Electrical computers and data processing systems) and US patent class 395 (Information processing system organization). His work suggests that one way of following the recent development of embedded software related patent activity is to trace the number of patents granted in US patent class 395 (which may not include all embedded software oriented patents and may include patents not related to embedded software). Other methods have also been used (see Gregory Aharonian, International Patent News Service, www.patent-software.com).

Most of the previously mentioned work makes use of patent classification systems such as the UPC used by the US Patent and Trademark Office (USPTO), the IPC used by the EPO and most European patent offices or the F term system used by the Japanese patent office (Liu and Shyu, 1997). These classification systems are both systematic and hierarchical. Using them a patent application or patent can be assigned to one or more quite narrowly defined categories, sometimes in quite different parts of the classification tree (that is, in different first level patent classes). The UPC and IPC patent classifications have been related (more or less successfully) to industry branch codes such as the SIC code (Patel and Pavitt, 1987; Archibugi and Pianta, 1996) in the case of the UPC and the German Wirtschaftszweige (WZ) (Greif and Potkowik, 1990) or the NACE codes (Pavitt, 1988) or ISIC codes (Archibugi and Pianta, 1996) in the case of the IPC.

With the possible exception of the US patent class 395, neither the patent classification systems nor the industry branch codes contain categories that properly and inclusively identify embedded software related inventions or products/firms/industries. Computer software, and especially software embedded in industrial products, is present in the widest variety of patent applications and patents as well as in the widest variety of industries and commercial entities in general (McQueen et al., 1998). It is probably a hopeless task to find a way of identifying computer software by class or code within the framework of the classification systems as they are configured at present. However, bibliometric search techniques can be used successfully to find most computer program related patents in patent databases as will be shown in the following.

It has naturally long been possible to apply bibliometric search techniques (key word searches, author searches and the like) to patent data (Liu and Shyu, 1997). It might be surprising that this has not been done more often but perhaps searches based on patent classifications have generally been satisfactory. As mentioned previously, one area in which it is not possible to use the patent classification scheme effectively is embedded software. Nevertheless, there have been estimates of the extent of patenting including computer programs based for instance on certain UPC patent classes (class 364 and 395) (Lennon, 1994). The accuracy of such estimates has been called into question partly because it is difficult to determine how many patents without mention of computer software are included in these patent classes. Another serious difficulty with this data is that there can be many patents including computer software not assigned to class 364 or 395 but to other patent classes.

A similar situation arises in the field of biotechnology. This new generic technology interpenetrates many fields of science and engineering and is, therefore, difficult to capture in a single patent class. Instead biotechnology patents can be found mainly in several IPC classes as well as spread thinly over a wide range of other patent classes (Joly and de Looze, 1996). In order to obtain a good sample of patents in plant biotechnology Joly and de Looze (1996) intersected the results of bibliometric searches using keywords carefully chosen together with experts in plant biotechnology with IPC classes concerning biotechnology. They then used a scientometric method, co-word analysis, to identify 26 more or less coherent research programs within the larger field of plant biotechnology. The research programs revealed useful information concerning the structure of commercial activities in plant biotechnology. There appear to be few studies similar to that of Joly and de Looze (1996) reported in the literature. This should not be taken to mean that these methods are not used by various commercial organizations. Possibly based on experience of such commercial use of keyword searches, Liu and Shyu (1997) discussed some of the problems associated with the use of keywords for patent searches.

3. Development of the search profile

We have used a bibliometric search method based on a carefully chosen list of search words in our attempt to characterize the embedded software patenting during the past two decades. We have chosen to study patent applications rather than granted patents because the applications are nearer the respective inventions than are the granted patents.

The search profile was assembled explicitly to characterize the distribution of embedded software patent applications in the International Patent Classification system (IPC 6) at different points in time making use of international (WO) patent application abstracts, that is, the title and abstract texts of PCT applications. Their purpose is to briefly inform readers of the contents of the respective patent documents. Patent abstracts usually consist of a short introduction such as a describing name and/or the purpose of the invention followed by a primarily technical description. If it had been possible it would probably have been better to exclude the introductions of the abstracts from the search since a word in the introduction may indicate in what way the invention is intended to be used rather than describe the substance of the invention. We do not believe that it is possible to generate a set of search words that is perfect in the sense that it only picks out the documents that are sought. Rather, there will always be too few or too many documents in the search result.

Some words such as ‘processor’ are commonly used in information technology and in connection with embedded software. However, processor can also be used to describe many inventions with no relation to information technology. Attempts to exclude words of this kind have, therefore, been made. In fact, only words with a very strong exclusive relation to embedded software are in the final search profile. Another category of excluded words is words that have only a very limited bearing to technology, an example being ‘financial’. It would probably be safe to include such words since it is not unlikely that they would relate to embedded software related patent documents for reasons of patentability, at least when considering PCT applications. However, this would probably not improve the search results very much. On the contrary, they might bias the search result instead. Other things to consider here are the use of synonyms and homonyms since the reliability of the search result would increase if synonyms could be included and the ‘wrong’ homonyms excluded.

The first step in developing the search profile was to concatenate patent claims of a selection of appeal board cases as the starting point of a search profile refining procedure. All cases dealing with computer program related inventions from the USA (24 cases) and the EPO (22 cases) included in a previous study dealing with the patentability of computer program related inventions were used (Olsson, 1996). The first US case was from 1972 and the last one from 1994. The first EPO case was from 1985 and the last one from 1994. The inventions concerned include various aspects of information technology such as programs for computers, seismic prospecting and medical applications. There were almost a hundred patent claims presented in the appeal board decisions and they were all included in the concatenation. Both claims from rejected patent applications and claims from granted patent applications were included since focus is on the claims at a word level and not how words are related to each other, which is an issue to address when dealing with the patentability question. The concatenation resulted in a total of approximately 14,700 words.

The first step of the refining procedure was to delete words that cannot have a specific relation to embedded software. The following five word classes (Johansson and Lysväg, 1991) were studied:

- nouns,
- verbs,
- adjectives,
- adverbs,
- lexical words and function words.

Lexical words and function words include determiners (e.g. ‘the’, ‘a’), auxiliary verbs (e.g. ‘be’, ‘must’), prepositions (e.g. ‘for’, ‘of’), conjunctions (e.g. ‘and’, ‘that’) and pronouns (e.g. 'what', 'which'). Words from the last two word classes above were deleted from the list since they are not considered to have any specific relation to embedded software. This step reduced the number of words to about 7000. The next step included identifying and deleting nouns, verbs and adjectives not having a strong relation to embedded software. This step reduced the number of words to 30.

The next step of the refining procedure was to use a US patent database (http://patent.womplex.ibm.com/) to analyze the remaining candidate search words individually to assess their reliability as indicators of embedded software related patent documents. In this database it is possible to choose the option of searching the patent claims and front pages. For each candidate search word the number of patent documents in which the word was found at least once in 1993 was generated as well as a list of the first 500 such documents. (In a few cases this latter list was less than 500 documents long.) The first 10 and the last 10 patent documents were checked to determine whether they indeed related to embedded software. Table 1 summarizes the results. The second and third columns show how many of the 10 first and 10 last documents concerned embedded software. The next column is the average of these two assessments expressed as a percentage. Finally, in the last column the total number of 1993 patent documents containing the candidate search word is given.

Table 1
Indicative ability of embedded software patent search words

Word	Ten first documents	Ten last documents	Relative indication ability (%)	Number of patent hits in 1993
Output	9	5	70	13,887
Input	10	9	95	9996
Data	10	10	100	9975
Image	10	9	95	7021
Memory	10	10	100	6245
Display	10	7	85	5493
Digital	10	10	100	4607
Computer	10	10	100	4403
Average	5	4	45	4296
Calculation	10	10	100	3963
Logic	10	10	100	2685
Bit	9	9	90	2384
Parameters	8	9	85	2372
Register	10	9	95	2343
Network	10	8	90	2314
Coefficient	10	10	100	2087
Analysis	7	5	60	2009
Program	10	10	100	1832
Command	10	9	95	1752
Node	10	9	95	1232
Pixel	10	9	95	1130
Scan	10	7	85	1124
Binary	10	8	90	995
Equation	10	7	85	926
Numerical	10	8	90	621
Arithmetic	10	10	100	493
Interpolation	10	10	100	465
CPU	10	10	100	436
Byte	10	10	100	334
Simulator	6	7	65	142

Table 1 shows that search words such as 'data', 'digital' and 'computer' are used frequently and also that in our test they always picked embedded software related patents (relative indication ability 100%). They are thus identified as good search words in this context. On the contrary, 'simulator' does not contribute very much as a search word as it is not used very often and when it is, then it also picks patent documents that do not concern embedded software.

The final step of the refining procedure was to introduce wildcards to be able to find different grammatical forms of the search words. Thus the final list of search words included arithmetic∗, byte∗, calculat∗, coefficient∗, compute∗, CPU∗, data∗ and digital∗ where ∗ indicates that different forms of the word are used in the search. Also important is that in some cases there may be differences in spelling between UK and US English, for instance in the case of program∗.

These search words were applied to the titles and abstracts of PCT patent applications in the Derwent world patents index database. We chose to study PCT patent applications rather than US patents partly because PCT patent applications are made public 18 months after the date of first submission. On the contrary, US patent applications are kept secret during their processing, which can take several years. Thus PCT patent applications published in March 1993 were submitted in September 1991 while US patents published at the same time can have been submitted many years previously. In the Derwent database the titles and abstracts of the patent applications were re-written in order to make them more informative than the original documents might have been as well as to make it possible to search the documents in a single language (English).

4. Search results

The search words were used to determine the numbers of patent applications in each IPC patent subclass, for instance G06C or G06D. To indicate the number of patent applications in patent class G06 the numbers for each subclass were summed. Finally, in order to get an indication of the number of applications in patent section G the numbers for all the patent classes belonging to section G were summed. On the surface this looks like a very innocent procedure, but in fact it is more subtle since a given patent can be classed in two or more patent subclasses. Thus when the numbers for the subclasses are summed the result can be larger than the total number of patent applications in the patent class to which they belong. The same situation obtains when summing the numbers for patent classes in a given patent section. In the present case typically each patent application was classed in between two and three patent subclasses.

To limit the number of patent applications to be investigated and thus the cost of the patent search and subsequent data analysis only patent applications published from January to April in 1988, 1993 and 1998 were included.

The distribution of the patent section assignments for embedded software related patent applications over the eight IPC sections is shown in Table 2. There are non-zero entries in all eight sections in all three years even if the numbers for sections D and E are very small. The calculated growth rates for these two classes are, therefore, unreliable and enclosed in parentheses. Assuming exponential growth the total annual volume of these patent assignments has grown about 17% per year. This can be compared to the growth in numbers of EPO patent applications in general during the 1990s, around 10.5% per year. Patent sections G and H dominate the picture with H growing faster than G. Sections B and F are characterized by low numbers and low growth rates while sections A and C show low numbers but high growth rates. At this level the sums of the patent application assignment numbers are a little greater than the total number of patent applications (patent families) since a given patent can be assigned to two or more patent classes.

If the growth rates shown in Table 2 are used to predict corresponding numbers of patent application assignments in 2003 the result shows that section H will surpass section G.

The next step was to investigate the distribution of patent application assignments over the 118 IPC patent classes. The results for the patent classes with the most application assignments in 1998 are shown in Table 3. These 20 classes contain more than four of every five patent application assignments identified in the study. The seven patent classes for which average growth rates are shown contain more than three of every five patent application assignments.

Class H04 with more than a fourth of all embedded software patent assignments in 1998 and a high rate of growth corresponds to the telecommunications industry. This is followed by class G06 having to do with both analog and digital computing and including more than a sixth of all embedded software patent assignments in 1998. Finally, G01 concerning measuring and testing contributes a little less than a 10th of the assignments. Compared to the first two, this third class shows relatively low growth. Extrapolation of the data in Table 3 to the year 2003 indicate that the order of the first seven patent classes will remain unchanged.

Table 2
The number of times each of the eight IPC sections was chosen in patent applications dealing with computer software published in the months of January, February, March and April of 1988, 1993 and 1998

IPC section	1988	1993	1998	Average annual growth (%)
A	67	168	379	19
B	122	271	347	11
C	43	132	289	21
D	11	23	18	(6)
E	4	33	43	(29)
F	48	98	126	10
G	532	1084	2205	15
H	251	722	1930	23
Sum	1078	2531	5337	
Average				17

Section A: Human necessities; section B: Performing operations, transporting; section C: Chemistry, metallurgy; section D: Textiles, paper; section E: Fixed constructions; section F: Mechanical engineering, lighting, heating, weapons, blasting; section G: Physics; section H: Electricity.

Table 3
The number of times each of the 20 most populous (in terms of patent application assignments in 1998) of the 118 patent classes was chosen in the months of January, February, March and April of 1988, 1993 and 1998

IPC class	1988	1993	1998	Average annual growth (%)
H04	129	472	1417	27
G06	158	381	902	19
G01	162	293	485	12
A61	49	117	251	18
H01	32	85	204	20
G11	49	82	200	15
H03	50	87	186	14
G08	14	54	116	
G07	24	33	100	
G09	36	44	96	
C07	10	33	93	
G02	24	45	92	
B60	9	52	86	
G05	40	85	85	
H02	26	49	76	
C12	8	29	61	
G03	10	28	53	
C08	5	18	52	
G10	9	22	51	
A63	7	10	48	
Sum	1078	2531	5337	
Average				17

Class H04: Electric communication technique; class G06: Computing, calculating, counting; class G01: Measuring, testing; class A61: Medical or veterinary science, hygiene; class H01: Basic electric elements; class G11: Information storage; class H03: Basic electronic circuitry.

The previous statistics describe the growth of the numbers of embedded software related patent application assignments in absolute terms but they say nothing about which patent classes contain high concentrations of these assignments. For instance, it turns out that these assignments are relatively highly concentrated in patent classes H04 and G06 but they are relatively rare in patent class G01. To carry out this analysis it was first necessary to eliminate patent classes with low numbers of embedded software related patent application assignments since this data contain much statistical noise. It is also of little real importance since the numbers are low. Thus only patent classes that contain at least one percent of all assignments in the given year were included in the analysis. Then the numbers were divided by the total numbers of patents applied for in that class in the first four months of that year, as determined in a separate study of patenting frequency, this time at the patent class level. The results for the 20 most embedded software intensive patent classes are shown in Table 4.

The number of embedded software based patent application assignments exceeds the total number of patent applications in two patent classes, H04 and G06. This underlines the fact that a patent application can be assigned to two or more subclasses and thus counted two or more times in the procedure used here. In the present case the number of embedded software patent application assignments exceeds the number of embedded software patent applications (patent families) by a factor of about 2.25 on the average so to estimate the actual percentage of patents based on embedded software in any class one could divide the number in Table 4 by 2.25. Thus only in two cases, H04 and G06, do embedded software-based patent applications ever exceed about half of all patent applications in a given IPC class.

Table 4
The ratio of the number of embedded software-based patent application assignments to the total number of patent applications in the 20 most embedded software intensive of the 118 patent classes in the months of January, February, March and April of 1988, 1993 and 1998

IPC class	1988	1993	1998	Average
H04	0.88	1.11	0.87	0.96
G06	1.12	0.98	0.82	0.97
G11	0.77	0.66	0.63	0.68
G07	0.86	0.55	0.59	0.67
H03	0.70	0.60	0.55	0.62
G08	0.40	0.55	0.52	0.49
G05	0.57	0.54	0.41	0.51
G09	0.55	0.35	0.37	0.42
G01	0.43	0.33	0.29	0.35
G03	0.00	0.16	0.25	0.14
G02	0.29	0.22	0.19	0.23
H02	0.29	0.26	0.18	0.24
H01	0.13	0.15	0.14	0.14
B60	0.00	0.18	0.13	0.10
A61	0.10	0.06	0.07	0.08
C12	0.00	0.04	0.04	0.03
C07	0.00	0.03	0.04	0.02
G12	0.00	0.00	0.00	0.00
G21	0.00	0.00	0.00	0.00
H05	0.20	0.17	0.00	0.12

Class H04: Electric communication technique; class G06:Computing, calculating, counting; class G11: Information storage; class G07: Checking devices; class H03: Basic electronic circuitry; class G08: Signaling; class G05: Controlling, regulating.

Patent classes H04, G06, G11 and H03 belong to both the seven most populous and the seven most concentrated patent classes in terms of embedded software patent application assignments with the first two significantly more populous and more concentrated than the latter two. Clearly classes H04 and G06 are the main IPC embedded software patent classes.

Concerning class H04, in 1998 embedded software patent application assignments were fairly evenly distributed over the five subclasses B (transmission), L (transmission of digital information), M (telephonic communication), N (pictorial communication) and Q (selecting) with L a little stronger than average and M a little weaker. In 1988 subclass Q was relatively very weak and subclass M was weak compared to the other three subclasses. Thus there has been some leveling out in favor of transmission and pictorial communication here. Concerning class G06 the situation is somewhat simpler since the great majority of embedded software patent application assignments have been in subclass F (digital computers) with most of the rest in subclass K (recognition, handling and recording data). Here, there has been some concentration since the dominance of subclass F increased from 1988 to 1998.

To get a measure of the spread of embedded software based patent application assignments across all patent classes it is natural to multiply the number of patent application assignments in a given class by the concentration of those assignments in the same class, thus combining Tables 3 and 4. The result for the 20 most important classes is shown in Table 5. Apart from some changes in order, at the top of the table the same patent classes appear in all three years. The concentration of assignments to classes H04 and G06 increased steadily from 54% in 1988 to 69% in 1993 and to 75% in 1998. More generally, with time there has been a steady increase in the concentration of embedded software patent application assignments to fewer and fewer patent classes. This is in stark contrast to the spread of embedded software across industrial sectors with time.

5. Comparison with other data

Care is necessary when comparing the results of this bibliometric search concerning European patent appli-

Table 5
The distribution of embedded software patent applications across the 20 most important patent classes (number of patent application assignments times their concentration) for the months of January, February, March and April of 1988, 1993 and 1998

IPC class	1988	IPC class	1993	IPC class	1998
G06	177	H04	524	H04	1236
H04	114	G06	375	G06	736
G01	70	G01	98	G01	140
G11	38	G11	54	G11	126
H03	35	H03	53	H03	101
G05	23	G05	46	G08	61
G07	21	G08	29	G07	59
G09	20	G07	18	G09	35
H02	7	G09	15	G05	35
G02	7	H02	13	H01	28
B41	6	H01	12	G02	17
G08	6	G02	10	A61	17
A61	5	B60	10	H02	14
H01	4	G10	9	G03	13
B23	3	F02	9	B60	11
H05	3	A61	7	C07	4
F02	2	H05	5	C12	3
B65	1	G03	4	G12	0
G10	0	B41	4	G21	0
B60	0	B23	2	H05	0

cations with patent data obtained by other means. It is necessary to include as great a portion of all embedded software oriented patent applications as possible without including patent applications that are not based on embedded software. Further, it is important to distinguish between patent applications and granted patents for several reasons. The latter are only a fraction of the former. Patents can be granted any number of years after the application is filed whereas patent applications are published 18 months after the priority date in the European patent system. The number of patents granted need not correlate well with the number of patent applications, as illustrated by EPO patent trends during the 1990s.

One easily accessible source of data that may be comparable to the present results is the US patent class 395. According to Lennon (1994) this class was restructured in 1991 to include many US patents concerning software related inventions. Whether it includes all such patents and only such patents is not quite clear. It contains many subclasses, of which by far the most populous is subclass 500 ‘Compatibility, simulation, or emulation of system components’. It is easy to get the numbers of patents published annually in class 395 from www.delphion.com/cgi-bin/patsearch. During the beginning of the 1990s the numbers of patents assigned to this class increased significantly in a transient response to the change in classification policy. Then the annual rate of increase in the numbers of patents leveled out at around 28% before turning negative in 1999. This downturn may be an artifact. The annual growth rate of 28% can be an overestimate as the initial transient response may not die out completely for many years. It is significantly higher than the 17% annual growth rate for embedded software-based patent applications that characterizes our European bibliometric data.

This brings up the question whether these two results are indeed compatible. A way of beginning to answer the question is to carry out a bibliometric analysis of the US patents in the same way as the analysis of European patent applications was made. The result could also be used to compare the rates of growth of embedded software-based patents in Europe and the US. A better method might be to repeat the bibliometric analysis of European patent applications while distinguishing between the countries of origin of the applications. This could potentially give interesting information on the relative strengths of different countries in this important area.

It would be natural to compare the growth in frequency of embedded software-based patent applications with R&D effort devoted to embedded software. However, it is unlikely that data on R&D effort devoted to embedded software can be obtained except in specific relatively rare cases. There is some evidence that the number of patents is closely related to the number of trained engineers (Schmookler, 1957; Pavitt, 1982). Such a relationship should also hold in the case of embedded software related patents and patent applications. It would, therefore, be instructive to compare the numbers of engineers trained in information technology or computer science or closely related subjects with the numbers of embedded software related patent applications in various countries.

At the other extreme, one of the best measures of the commercial value of inventive activity is probably granted patents that have been maintained for a few years. Lacking that, granted patents are probably a better measure of commercial value than are patent applications (Griliches, 1990). In the present case the data for US patent class 395 for the middle 1990s indicates a growth rate of software related patenting up to about 28% per year. The annual growth rate of embedded software related patent applications according to this bibliometric study is only about 17%. Concerning commercial value, a Swedish study indicated that the economic value of Swedish embedded software grew at about 21% per year during the late 1980s and the early 1990s (McQueen et al., 1998). This result may be a little higher than the corresponding result for all of Europe. According to world semiconductor trade statistics (obtained from www.semichips.org/stats) the real rate of growth in semiconductor trade in the middle 1990s was about 12–13%. Based on this data (which should be complemented with other, more detailed data) it is unclear whether the field is becoming ‘crowded’ in the sense that more and more patents protect less and less economic value. On the other hand, it is definitely true that fewer and fewer patents protect more and more computational capacity since the amount of computer power and speed delivered at a given price has increased drastically each year. What this really means to the software industry in terms of the value and role of patents it not clear.

6. Discussion and conclusions

In interpreting the previous results it should be kept in mind that perceptions of patentability, especially the patentability of software, play a key role here. In Europe patentability requirements include novelty, inventive step and technical character. Computer programs per se cannot be patented since they do not per se have technical character. They must be explicitly embedded in a technical environment. In the beginning of the 1980s European interpretation of this technical character requirement effectively disqualified the great majority of patent applications including computer programs. In the middle of the 1980s this interpretation was changed so that patents including computer programs and explicit reference to a technical environment can be granted. However, today it is still not uncommon to find inventors who (wrongly) believe that computer programs cannot be patented and, therefore, do not apply for patents on inventions based on computer programs. This reduces the numbers of embedded software related patent applications found in our bibliometric search compared to the real number of patentable embedded software related inventions.

The situation is a little different in the US where a patent must be 'useful' rather than have technical character. In many respects the range of patentable computer program related inventions has long been wider in the US than in Europe. For this reason in the above the terms 'information technology' or 'computer programs' have been used in connection with US patents while 'embedded software' has been the preferred term in connection with European patent applications. Thus caution is required when comparing US patent statistics on information technology with European patent application statistics on embedded software related inventions.

It is also important to remember that far from all inventions are patentable and not all patentable inventions are patented. There are alternative ways of having exclusive use of intellectual property, for instance being first to market without disclosing know-how. Also, especially in Europe patenting is expensive and inventors can decide not to patent for economic reasons. More generally, the perceived need to patent must exceed the various costs of patenting in order to be profitable to submit a patent application to a patent office. This need to patent depends on the characteristics of the owner of the invention and the nature of the market and technical competition he faces (Pavitt, 1985; Olsson and McQueen, 2000). In any case, the number of embedded software related patent applications made is reduced by these 'barriers to patenting'. Whether this affects some patent classes more than others is not known, but such biases are expected to be small.

It is also possible to apply for a patent solely in order to prevent others from using an invention. This increases the number of patent applications found in our bibliometric search without a corresponding increase in the number of embedded software related inventions actually realized.

Perhaps more importantly, there are trends in patenting. For instance, patenting frequencies have increased during the 1990s both in the US and Europe. This can be due to public campaigns in favor of patenting, to business cycles or to a perceived widening set of technical opportunities, for instance. In any case a trend toward increased patenting inflates the annual patent application growth rates found in our bibliometric search. How much this really amounts to is not known.

From the aforementioned and as shown in the literature survey it should be clear that counting numbers of patent applications gives an indication of inventive activity and innovative effort, but exactly what it measures and how well it is measured is not perfectly clear. The long-standing questions involved cannot be answered here.

In developing the search profile considerable effort was expended to identify search words that pick out only embedded software related patent applications. The number of search words was limited at the risk of not finding all embedded software related patent applications. The probable result is that more embedded software related patent applications were missed than wrongly included in the search results.

Notwithstanding the caveats above, the most important conclusions that can be drawn from the results presented previously are the following:

(a) The number of embedded software related patent application assignments per year has grown by about 17% per year during the 10 year period from 1988 to 1998. The growth rate was a little higher during the period from 1988 to 1993 than from 1993 to 1998. This may be a transient due to the liberalization of patenting rules concerning software related patents that took place at the EPO around 1985.
(b) During the period from 1988 to 1998 there has been a slight concentration of embedded software related patent application assignments to fewer patent classes. However, there are only two IPC patent class in which a majority of the patents is embedded software related, H04 and G06.
(c) More than three of every four patent application

assignments are in IPC patent sections G and H while very few patent applications are assigned to sections D, E and F (together around 5% and decreasing). The two most important IPC patent classes are H04 and G06 which together account for a little more than 40% of all embedded software related patent application assignments. The top seven patent classes account for two-thirds of embedded software related patent application assignments. This situation is expected to remain unchanged during the next few years.

(d) The fact that the distribution of embedded software related patent application assignments over patent classes has changed so little over the decade (slight concentration can be detected) means that patent statistics can provide little or no information about the spread of this technology. On the other hand, this also means that time series (longitudinal studies) can be informative.

The 17% growth rate is significantly lower than that obtained from US patent class 395 (about 28% per year), which may be less reliable because the time base is shorter, patents rather than patent applications are reported and the range of patentable inventions is wider than that for European patents. It is significantly higher than the rate of increase of world semiconductor sales (about 12–13%) but less than the growth in the economic value of embedded software in Swedish industry (about 21% per year). Since the amount of computational power that can be bought for a given price has grown very quickly, this does not necessarily mean that patent protection is becoming ‘more dense’.

The fact that the distribution of embedded software related patent applications has sharpened only slightly over the past 10 years means that there is a sort of constancy here that may be useful. When a patent is classed in a patent office prime attention is paid to the problem to be solved and how it is solved. In this sense during the whole time period studied the same sorts of problems have been attacked and solved by inventors in this field. On the other hand, the coupling of the patented solutions to technical milieux, which gives the inventions technical character, does not affect the classification much. On the contrary, when the invention is used commercially it is its technical effect that is important and which is the basis for classification by business branch in national economic statistics. Thus we seem to be in possession of a potentially very valuable tool since patent statistics can be used to track embedded software related inventive effort in time with little interference due to changes in classification distributions while economic statistics can be used in cross-sectional studies to characterize how this inventive effort is utilized.

Thus a good way to go forward might be to use patent data to describe the temporal development of embedded software related invention and innovation and to use economic data to describe the spread of the technology. It might be particularly instructive to compare different countries using this type of bibliometric search of patent applications.

References

Academic Press Dictionary of Science Technology, 1996. Academic Press.

Archibugi, D., Pianta, M., 1996. Measuring technological change through patents and innovation surveys. Technovation 16, 451–468.

Chakrabarti, A.K., 1991. Competition in high technology: analysis of patents of US, Japan, UK, France, West Germany, and Canada. IEEE Transactions on Engineering Management 38, 78–84.

Ernst, H., 1995. The use of patent data for technological forecasting: the diffusion of CNC-technology in the machine tool industry. Small Business Economics 9, 361–381.

Ernst, H., 1998. Industrial research as a source of important patents. Research Policy 27, 1–15.

Greif, S., Potkowik, G., 1990. Patente und Wirtschaftszweige. Carl Heymanns Verlag KG, Köln.

Griliches, Z., 1981. Market value, R&D, and patents. Economics Letters 7, 183–187.

Griliches, Z., 1990. Patent statistics as economic indicators: a survey. Journal of Economic Literature 28, 1661–1707.

Grupp, H., Schmooch, U., 1999. Patent statistics in the age of globalisation: new legal procedures, new analytical methods, new economic interpretation. Research Policy 28, 377–396.

Jacobsson, S., Oskarsson, C., Philipson, J., 1996. Indicators of technological activities comparing educational, patent and R&D statistics in the case of Sweden. Research Policy 25, 573–585.

Johansson, S., Lysväg, P., 1991. Understanding English Grammar, third edition. Universitetsforlaget, Gjàrvik Trykkeri, Oslo.

Joly, P.-B., de Looze, M.-A., 1996. An analysis of innovation strategies and industrial differentiation through patent applications: the case of plant biotechnology. Research Policy 25, 1027–1046.

Kondo, M., 1999. R&D dynamics of creating patents in the Japanese industry. Research Policy 28, 587–600.

Lennon, M.J., 1994. Patents and computer software in the United States Patent World (July/August), 6–15.

Liu, S.-J., Shyu, J., 1997. Strategic planning for technology development with patent analysis. International Journal of Technology Management 13, 661–680.

MacMillan Dictionary of Information Technology, 1985. Macmillan.

Malerba, F., Orsenigo, L., 1999. Technological entry, exit and survival: an empirical analysis of patent data. Research Policy 28, 643–660.

McQueen, D.H., Wilhelmsson, L., Emanuelsson, G., 1998. Microprocessor-compatible engineers and embedded software in industrial products. European Journal of Engineering Education 23, 365–381.

Narin, F., Olivastro, D., 1988. Technology indicators based on patents and patent citations. In: van Raan, A.F.J. (Ed.), Handbook of Quantitative Studies of Science and Technology. Elsevier, Amsterdam (Chapter 15).

Olsson, H., 1996. Patentability and Computer Software in the USA and Europe, Previous study. Chalmers University of Technology.

Olsson, H., McQueen, D.H., 2000. Factors influencing patenting in small computer software producing companies. Technovation 20, 563–576.

Patel, P., Pavitt, K., 1987. Is Western Europe losing the technological race? Research Policy 16, 59–85.

Patel, P., Pavitt, K., 1997. The technological competencies of the world’s largest firms: complex and path-dependent, but not much variety. Research Policy 26, 141–156.

Pavitt, K., 1982. R&D, patenting and innovative activities. Research Policy 11, 33–51.
Pavitt, K., 1985. Patent statistics as indicators of innovative activities: possibilities and problems. Scientometrics 7, 77–99.
Pavitt, K., 1988. Uses and abuses of patent statistics. In: van Raan, A.F.J. (Ed.), Handbook of Quantitative Studies of Science and Technology. Elsevier, Amsterdam (Chapter 16).
Pavitt, K., Patel, P., 1988. The international distribution and determinants of technological activities. Oxford Review of Economic Policy 4, 35–55.
Scherer, F.M., 1965a. Corporate inventive output, profits and growth. The Journal of Political Economy 73, 290–297.
Scherer, F.M., 1965b. Firm size, market structure, opportunity and the output of patented inventions. The American Economic Journal 55, 1097–1125.
Schmookler, J., 1957. Inventors past and present. The Review of Economics and Statistics 39, 321–333.

Henrik Olsson was born in Göteborg, Sweden, in 1964. In 1990 he was awarded a masters degree in physics by Chalmers University of Technology. In 1994, after language studies at the University of Göteborg and after working as a computer programmer, he started as a doctoral student in the field of patents, patent policy and computer software. He worked part time for a patent law firm for three years before joining Awapatent AB.

Douglas H. McQueen was born in Philadelphia, Pennsylvania, USA, in 1945. In 1966 he was awarded a bachelors degree in physics by MIT and in 1971 he was awarded a PhD, also in physics, by the University of North Carolina at Chapel Hill. Subsequently he earned the docent degree at Chalmers University of Technology in 1973 and an undergraduate degree from the University of Gothenburg in 1979. From 1979 to 2000 he worked mainly in the Chalmers Innovation Center. He is presently a private consultant dividing his time between studies of technical innovation and research in the physical sciences, especially materials science.