7-1-2003

Mining PICMET: 1997-2003 Papers Help You Track Management of Technology Developments

Alan L. Porter

Robert J. Watts
U.S. Army, Development & Research Center

Timothy R. Anderson
Portland State University, tim.anderson@pdx.edu

Mining PICMET: 1997-2003 Papers Help You Track Management of Technology Developments

Alan L. Porter[1], Robert J. Watts[2], Timothy R Anderson[3]

[1]Search Technology, Inc., USA
[2]U.S. Army Tank and Automotive Research, Development & Engineering Center, USA
[3]Department of Engineering and Technology Management, Portland State University, USA

Abstract- PICMET is providing the abstracts from 1997, 1999, 2001, and 2003 to enable you to "text mine" the contents of these four conferences. Using VantagePoint Reader software (available to PICMET attendees at no charge) allows you to locate specific research abstracts of interest. Text mining helps ascertain "who's doing what, when?" The software also enables you to see larger patterns in Management of Technology (MOT) research.

I. INTRODUCTION

This effort derives from exploratory analyses of the 1997, 1999, and 2001 PICMET abstract records [18]. That proved sufficiently rich to experiment with providing the PICMET abstracts as a resource to Conference attendees this year. The abstract records of these four conferences were provided by Portland State University. Search Technology, Inc., applied its *VantagePoint* software to "text mine" those records. These are provided to you along with the *VantagePoint Reader* software* that enables you to find materials of interest.

VantagePoint Reader allows you to browse and select subsets of the 1699 abstract records of interest. This works somewhat analogously to *Acrobat Reader* -- you can see trends, investigate "who's doing what?", and dig down to read particular abstracts that you want. But you can't create new files, perform new analyses, or export record sets. To perform such new analyses you would need the *VantagePoint* software [for sale -- see http://www.theVantagePoint.com; also available at no cost as *TechOASIS* only to U.S. Government personnel]. Beyond the capabilities you have at hand using the *Reader, VantagePoint* is also able to seek relationships based on the co-occurrence of particular terms [2, 5]. Principal components analysis and variants can generate various "maps" -- e.g., of term clusters or authors -- to help perceive activity patterns. At the PICMET tutorial, we illustrate some such analyses.

The next section provides a brief background on text mining in support of R&D management. The following section explains how you can use the PICMET data in support of your research interests.

II. BACKGROUND ON TEXT MINING

This section provides a few pointers to those of you who might want to further explore "bibliometrics" and "text mining" for technology management. These studies of science through looking at its outputs (papers) trace back to Price [13]. Bibliometrics counts such publication (or patent) activity to detect trends and changing patterns [10]. Text mining carries this further to process the content of those papers, aided by natural language processing [7, 8].

We are particularly interested in tracking advances in a field (in this case, "MOT") through profiling research activity changes over time [17, 19].† Such analyses can "map" changes so as to aid R&D management and other aspects of technology management [9, 14, 15]. They can contribute to research evaluation [6]. The scale of such inquiries can range from very specific technical domains [12] to national R&D profiles [4]. Knowing R&D emphases can also help inform competitive technological intelligence [1] and technology forecasting [16]. We are presently writing a book on technology analyses using such approaches [11].

III. MINING THE PICMET ABSTRACTS

A. Limitations

The data themselves were not all compiled using the same format. Hence, you will see that the records differ. Some information is present only for certain conferences. In particular, information on the authors' organization and country is only available for 2001 and 2003. Data for 2003 were provided as of early April.

Processing of the records is not 100% error-free. You may see mistakes in identifying "authors" as being the same or different persons; likewise, for organizations. The "MOT-relevant" phrases are solely the judgment of Porter and Watts. They reviewed the 1000 most frequently occurring phrases from the titles and abstracts (i.e., those occurring 5 or more times) and selected 409 that they thought might be of interest to some MOT colleagues. [Although we tried to use a set of "MOT" keywords from other databases as a thesaurus [3], it did not work well here, probably because MOT terminology is less specialized than in a restricted scientific domain.]

B. Using the PICMET Abstract Data

First, let us examine a summary of the data. Figure 1 shows the content on your CD. The fields that you can

---

access are listed. We have parsed the titles and abstracts to list the noun phrases contained in the 1699 papers. We have

names of authors, their organizations, and countries. And we have the year in which the paper was presented.
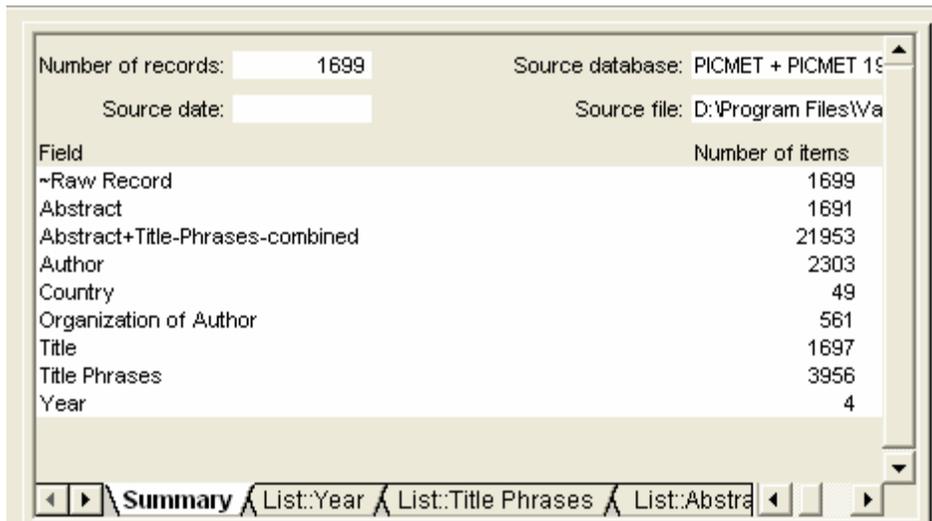

Figure 1. Summary of imported PICMET records.

Figure 2 shows the Author list as a matrix by year. Clicking on an author's name shows each of the paper titles authored. Since this table is ordered by decreasing # of records, we can see that that Hannu Jaakola had the most presentations at PICMET between 1997 and 2003 and is closely followed by David Wilemon and Qingrui Xu. The

table also makes it visible how the number of presentations changed from year to year. When you click on Hannu Jaakola, he is highlighted and all of his 18 paper titles are listed on the left. If you wanted to view just his presentations from one year, clicking on the corresponding cell to the right of his name would show just those papers for that year.
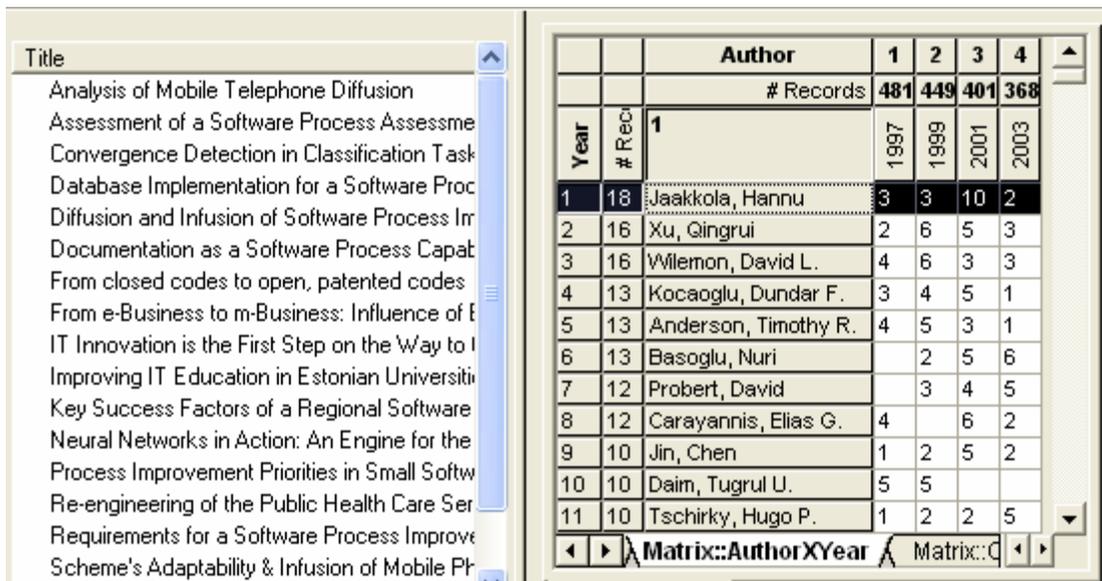
Figure 2. Listing of Authors by Year, sorted by record frequency.

The titles of the presentations for the highlighted author are listed on the left. Clicking on any of the paper titles opens a window (adjustable size) containing the information on the submission. This is illustrated by Figure 3, which

shows the complete information contained in the record as well as any additional notes that might be available or that you might want to add yourself.
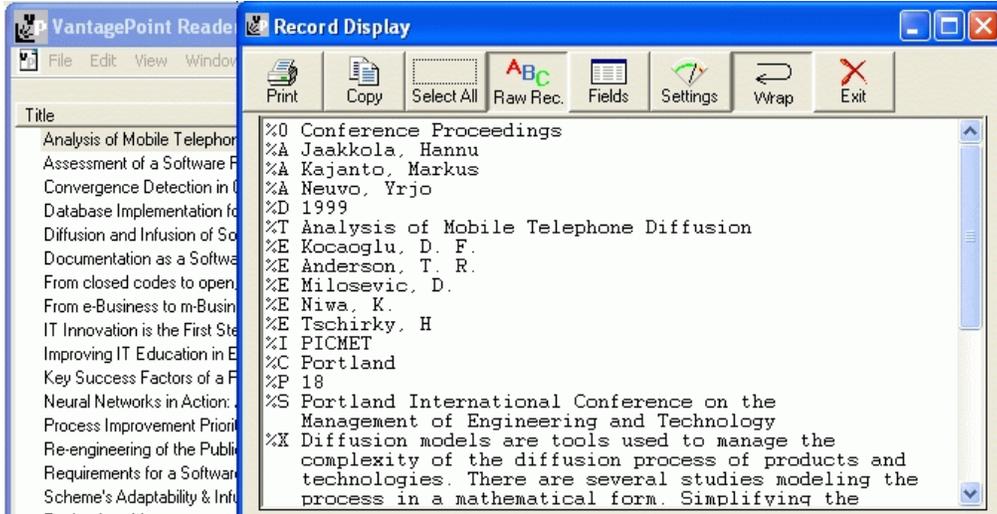
Figure 3. Detailed record/article information.

Figure 4 lists the organizations with the greatest number of records (papers with at least one author affiliated with that organization). Since the 1997 and 1999 data did not include author affiliation, these results are only for the 2001 and 2003 conferences. One interesting thing to note is that the top thirteen most active institutions come from nine different countries. This highlights the international nature of PICMET and the field of Technology Management. In this illustration, the "Electronics and Telecommunications Research Institute," ETRI, is highlighted with its associated presentation titles on the left.

Figure 4. Listing of presentations listed by institution.

Figure 5 shows how we can examine areas to see which technology management terms are the most frequently used and also examine the changing patterns of usage over time. This figure highlights that the terms "options" and "intellectual property" are being used more frequently. This approach could be used to identify subfields of technology management that are quickly developing as well as those that may be on the decline. Such information could help you situate your own research in "ripening" areas.
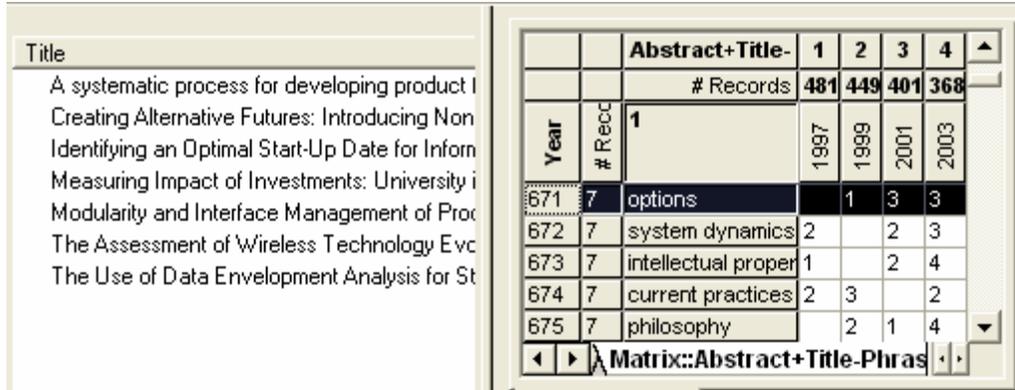
Figure 5. Phrases used in the abstracts and titles of presentations.

Figure 5 showed how term usage changes by year. Additional *VantagePoint Reader* file tabs (that open other worksheets, much like *MS Excel* worksheets) would allow the researcher to examine the usage of these or other terms by author, organization, or country. Both industry analysts and academicians could find this to be an important tool to highlight opportunities for rich new areas of research, point out potential collaborators, and warn of competitors.

The issue of collaboration is an interesting one. To investigate this, a matrix showing the co-authorship relationship among organizations is given in Figure 6. The seven organizations with the most presentations in PICMET 2001 and 2003 are shown. It is interesting to note that Bogazici University and Yeditepe University have had an impressive amount of collaboration, accounting for most of Yeditepe University's presentations and all of Bogazici's.

Figure 6. Matrix showing co-authorship of presentations by organization.

Lastly, Figure 7 illustrates a factor map. Using a Principal Components Analysis, presentations were clustered by their usage of common terms in the title and abstracts. The right-most blue circle in Figure 7 corresponds to the "Technical Problem-Solving" group. The terms listed in the box on the right-hand side of Figure 7 are those that show the strongest statistical association with technical problem-solving in the collection of 1699 presentations. These form one cluster of keywords tending to appear together, named after the most central of those keywords (highest loading on this factor).

It is surprising to see that closest statistical linkage to the *technical problem-solving* term is *market place*. The map

uses a special multi-dimensional scaling routine to locate nodes roughly in proportion to their relatedness. However, in this example there are 19 nodes, so accurate scaling in 19 dimensions cannot be retained in a 2-D representation. To better depict distance among the nodes, links are shown. These reflect a path-erasing algorithm to indicate greater or lesser similarity. To the left you can see the list of papers associated with the general topic of Technical Problem-Solving with a relevance score for each (ignore the sign; larger absolute values mean stronger relationship -- i.e., use of more of the high-loading terms that constitute the factor).
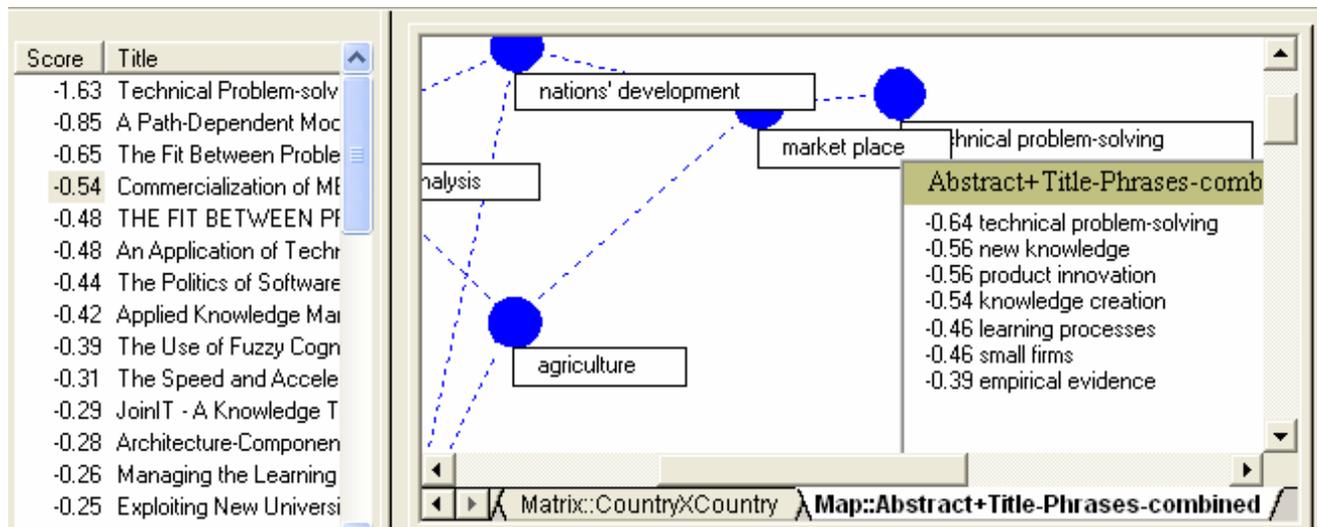
Figure 7. Factor map illustrating papers within the Technical Problem-Solving cluster.

IV. CONCLUSION

This paper demonstrates how text mining software can be used to examine trends and patterns of research in the fields of technology management using software developed specifically for these types of knowledge mining applications. Specific instructions on installing and using the VantagePoint Reader are provided in the Appendix.

REFERENCES

[1] Ashton, W.B., Johnson, A.H., and Stacey, G.S., "Monitoring Science and Technology for Competitive Advantage," *Competitive Intelligence Review* 7: S115-S126, 1996.

[2] Callon, J., Courtial, J.P., and Laville, F., "Co-word Analysis as a Tool for Describing the Network of Interactions Between Basic and Technological Research -- The Case of Polymer Chemistry," *Scientometrics* 2(1), 155-205, 1991.

[3] Chen, H., Lynch, K.J., Koushik, B., and Ng, T.D., "Generating, Integrating, and Activating Thesauri for Concept-based Document Retrieval," *IEEE Expert*, 25-34, 1993 (April).

[4] Cunningham, S.W., *The Content Evaluation of British Scientific Research*. PhD. Thesis. Science Policy Research Unit. University of Sussex, Brighton, UK, 1997.

[5] Kostoff, R.N., "Co-word Analysis," in Bozeman, B., and Melkers, J. (Eds.), *Evaluating R&D Impacts: Methods and Practice*, Kluwer, Boston, 63-78, 1993.

[6] Kostoff, R.N., "Research Impact Assessment: Problems," Progress, Promise, in *Fourth International Conference on Management of Technology*. Miami, FL, 423-433, 1994.

[7] Kostoff, R.N., and Geisler, E., "Strategic Management and Implementation of Textual Data Mining in Government Organizations," *Technology Analysis & Strategic Management* 11(4), 493-525, 1999.

[8] Kostoff, R.N. "Technology Watch and Evaluation," http://www.onr.navy.mil/sci_tech/special/technowatch/, March 13, 2003.

[9] Leydesdorff, L., "Various Methods for the Mapping of Science," *Scientometrics* 11, 295-324, 1987.

[10] Narin, F., Loivastro, D., and Stevens, K.A., "Bibliometrics -- Theory, Practice and Problems," *Evaluation Research* 18(1), 65-76, 1994.

[11] Porter, A.L., and Cunningham, S.W., *Tech Mining*, Wiley, New York, to appear.

[12] Porter, A.L., Kongthon, A., Lu, J-C., "Research Profiling: Improving the Literature Review," *Scientometrics* 53, 351-370, 2002.

[13] Price, D. de S., *Big Science, Little Science*, Columbia University Press: New York, 1963.

[14] Rip, A., "Mapping of Science: Possibilities and Limitations," in van Raan, A.F.J. (Ed.), *Handbook of Quantitative Studies of Science and Technology*, North-Holland, New York, 1988.

[15] Tijssen, R.J.W. , and van Raan, A.F.J., "Mapping Changes in Science and Technology," *Evaluation Review* 18(1), 98-113, 1994.

[16] Watts, R.J., and Porter, A.L., "Innovation Forecasting," *Technological Forecasting and Social Change* 56, 25-47, 1997.

[17] Watts, R.J., and Porter, A.L., "Requirements-based Knowledge Discovery for Technology Management," *Portland International Conference on Management of Engineering and Technology (PICMET)*, Portland, OR, 2001.

[18] Watts, R.W., and Porter, A.L., "Tracking the Evolution of Management of Technology (MOT), " *International Association for the Management of Technology*, Miami, 2002.

[19] Watts, R.W. , and Porter, A.L., "R&D Cluster Quality Measures and Technology Maturity," *Technological Forecasting and Social Change*, to appear.

APPENDIX: HOW TO USE THIS RESOURCE:

First, obtain the VantagePoint Reader from Search Technology, Inc. The reader is available to PICMET 2003 participants free of charge by e-mailing picmet2003@searchtech.com providing a) your name, b) your current corporate or educational affiliation, and c) your current location (city/state/country).   Second, open the "PICMET-97-99-01-03" file from the 2003 PICMET Conference CD-ROM.

You now have access to a set of sheets -- rather analogous to worksheets in *MS Excel.*

Look at the "Summary" sheet -- it indicates the available fields derived from the records and the number of entities in each. Want to know how many papers were presented at PICMET?   Here's an answer.   After the summary, the following "worksheets" provide a series of lists and matrices (combining any 2 lists).

You can use this information to get a "bird's eye" perspective of what is happening in our research arena. For instance, from the **lists**, you can generate:

- "who" information -- which authors, organizations, or countries have been most active at PICMET?
- "what" information -- which terms & phrases have been most popular, and
- "when" information -- looking at changes over time (conference **year**).
  [But note the caveat on some information not being available for all 4 years.]

The **Record View Window** enables you to "dig down" into the information to locate details on topics of interest. For example, you can select one or more entries from a list, or one or more cells in a matrix. To the left, you will see the titles of the abstracts that pertain (**Title Window**). If you double-click on a title, you will open up a record window to see that abstract.

Matrices combine any two of the lists:
1. authors
2. organizations
3. countries
4. abstract & title phrases
5. abstract & title phrases deemed most interesting for MOT (409 phrases)
6. year

For instance, you might want to see who from South Korea has addressed "supply chain management" at PICMET. In the *"Matrix: Abstract+Title PhrasesXCountry"* sheet, you can sort the entries in the "South Korea" column (by double clicking in this column's header) and locate "supply chain management."

Or, to check how "hot" supply chains are at PICMET, you could go to the *"Abstract+Title Phrases X Year"* matrix. Double-clicking on the Abstract+Title Phrases column alphabetizes.   Double-clicking again reverses the order, making it easy to find the supply chain items and see that interest remains strong.

"**Detail Views,**" on the right side of the screen, show aspects concerning entities you select "click on". Under the "View" menu, be sure "Details Window" is checked.   Then check "Create Details Window." You can have several Detail Views open and you can reposition them as you want (e.g., two columns of two views each).

Suppose you're planning to visit Japan and want to interact with someone active in the product development area. Go to the *"Matrix: Abstract+Title Phrases-combined: MOT terms X Country"* sheet. Sort the entries in the Japan column. Go to the "Product Development" row cell -- 7 abstracts are there. In the Detail View to the Right, pull down to see whatever field you want -- for instance, "Author" shows Seino with 2 related articles. If you click on Seino in the Detail View, the titles of his/her articles are highlighted in the Title View (to the left) and you can double-click on those to read the abstracts.

You can add more Detail Views. For instance, a second Detail View might show "Year" while the first Detail View showed "Author." You can resize windows as you like.

Note *VantagePoint Reader* "**Help**" on the top menu. Menu pull-downs also enable you to perform various usual functions. Under the **"File"** menu, you can:

- Open and Save the PICMET abstracts file, incorporating changes you make
- Print

Under **"Edit"** you can:

- Copy -- to paste sheets or portions -- e.g., say you wanted to copy the "Top 10" MOT terms X Year into *MS Excel* to generate trends in MOT emphases
- Find

Under **"Window"** you can:

- Have 2 versions of the file open to facilitate comparisons (either as 2 windows or as a cascade or tile)
- See the list of available worksheets (and reorder them as desired)