Automated text mining comparison of Japanese and USA multi-robot research

R. J. Watts[1], A. Porter[2] & B. Minsk[2]
[1]*U.S. Army Tank-automotive and Armaments Command, 6501 E. 11 Mile Road, Warren, MI 48397-5000, USA*
[2]*Search Technology, Inc. 4960 Peachtree Industrial Blvd., Suite 230, Norcross, GA 30071, USA*

Abstract

Wouldn't it be nice to automatically categorize documented research to distinguish organizational R&D emphases? For example, in the area of multi-robot research, one could surmise that Japanese sources have less emphasis than expected on biological approaches and reconfigurable robots, and more emphasis than expected on human-robot interface, motion coordination and robot learning. Development and analysis of such a capability comprises the subject of this paper.

Over the past decade, we have been developing a software tool for text analysis and classification, the Technology Opportunities Analysis of Scientific Information System (Tech OASIS). The Tech OASIS tool suite contains two information clustering algorithms, both of which evolved from principle components analysis (PCA). Based on the terms contained in an abstracts record field (e.g., abstract noun phrases, keywords, class codes, etc.), Tech OASIS generates topical clusters. We have now developed a capability to assess expectancy (i.e. binomial distribution likelihood) of occurrences. We demonstrate the derived analysis technique on the abstracts of 354 multi-robot research papers, principally, those from Japanese and USA sources.
Keywords: text mining, expectancy measure, principal components analysis, factor mapping, principal components decomposition, multi-robot research, VantagePoint, visualization of statistical measures.

1 Introduction

We first provide a brief background on information profiling and the software tool used to perform this analysis of multi-robot research. We then familiarize


Data Mining V, A. Zanasi, N. F. F. Ebecken & C. A. Brebbia (Editors)
© 2004 WIT Press, www.witpress.com, ISBN 1-85312-729-9

the reader with the multi-robot research data being analyzed and relate the data to expert opinions on the major areas of multi-robot research. The analysis processes are described, after which we present results, observations and conclusions drawn from the multi-robot information analyses. We finish with a discussion of the analysis approach strengths, caveats and future research.

2 Background

Managers need intelligence on emerging technologies. Yet, information available on a given technology exceeds traditional assimilation mechanisms -- namely, reading. For instance, under this tightly focused research subject, multi-robots, 354 unique research abstracts were retrieved from the INSPEC and EI Compendex databases. Full strategic technology planning might well consider information from these and other database sources, both open-source and gray literatures. Information overload abounds! Automated information analysis tools can enhance assessment of the developmental status and future prospects of a given technology.

The software, Technology Opportunities Analysis of Scientific Information System (*TECH OASIS*) or VantagePoint for the commercial market, helps perform "text mining." This can support technology assessments by automating the profiling of open-source R&D. It counts and compiles occurrences of particular terms, providing lists of the most frequent authors, organizations researching particular topics, terms used in the abstracts, etc. Such subject lists (i.e., fields) can be crossed with each other to create matrices. For instance, one might cross the leading keywords against the date of publication to see which keywords are most prevalent in recent years. It might be of interest to identify co-authors for each author (i.e., co-occurrence matrix of authors vs. authors). Or, one could group terms, such as keywords, to see which tend to appear in the same abstracts. van Raan has called these "one-dimensional" (lists) and "two-dimensional" (matrix) analyses [1]. *TECH OASIS* serves the process of "innovation forecasting," by applying bibliometric analyses to augment and enhance traditional technology forecasting techniques [2].

The Technology Opportunities Analysis (TOA) process strives to create knowledge from a "body" of literature beyond that obtainable by digesting individual pieces. *Retrieved text is treated as data* [3]. Text is parsed, counted, and patterns uncovered that speak to analysts' interests and management needs.

3 Multi-robot systems

Research forecasts, technology assessments and state-of-the-art overviews of a given research area often categorize the sub-disciplines or enabling technologies of the larger field to more easily discuss progress and accomplishments. Specifically, for the purpose of this paper, the research overview, "Guest Editorial, Advances in Multirobot Systems [4]," breaks down multi-robot research into seven topic areas:



Data Mining V, A. Zanasi, N. F. F. Ebecken & C. A. Brebbia (Editors)
© 2004 WIT Press, www.witpress.com, ISBN 1-85312-729-9

1. biological inspirations
2. communication
3. architectures, task allocation and control
4. localization, mapping and exploration
5. object transport and manipulation
6. motion coordination; and
7. reconfigurable robots.

This multi-robot review directly references 71 technical papers, and the authors are recognized experts in the field. So, these seven research categories can be considered well founded. Our analysis will support this conclusion and restates the importance of involving subject matter experts during automated information profiling exercises. More importantly, these expert-defined groupings will be an integral part of our publication expectancy analysis of multi-robot research.

4 About the data

We combined two literature search sets -- 198 abstracts from the EI Compendex database and 254 abstracts from the INSPEC database -- defined by the search strategy in Table 1. We then removed duplicate abstracts, based on identical titles and source publication years, with fuzzy matches on authors. The combined file contained 354 abstracts, published during the 1998-2003 period, related to collaborative robotics, herein defined as "multi-robots."

Table 1: Multi-robot research literature search strategy.

Set	Items	Description
1	13	COLLABORATIVE()ROBOT?
2	167	DISTRIBUTED()ROBOT?
3	25	(COLLECTIVE() (BEHAVIOR? OR TASK? OR GOAL?)) AND ROBOT?
4	169	(COOPERATIVE ()(BEHAVIOR? OR TASK? OR GOAL?)) AND ROBOT?
5	18	(COORDINATED ()(BEHAVIOR? OR TASK? OR GOAL?)) AND ROBOT?
6	381	S1 OR S2 OR S3 OR S4 OR S5
7	202	S7 AND PY>1997
8	**198**	**RD S7 (unique items) for the EI Compendex Database**
8	**254**	**RD S7 (unique items) for the INSPEC Database**

We created a field of the seven "IEEE multi-robot research groups," using the topic areas defined by Arai, *et al*. The terms used during this edit/find exercise were taken from the text of the topic area discussion in the multi-robot overview paper. Thereby, 310 of the 354 multi-robot abstracts were categorized into one or more of the seven groups. To assess whether any additional topic areas existed, a new dataset from the abstracts that were not contained in any of the seven groups was created. By analyzing the abstract noun phrases and descriptors fields of the new data set, we identified two possible additional topic areas -- robot learning



Data Mining V, A. Zanasi, N. F. F. Ebecken & C. A. Brebbia (Editors)
© 2004 WIT Press, www.witpress.com, ISBN 1-85312-729-9

and human-robot interface. Although these two topic areas captured (i.e., grouped) 42 and 41 abstracts, respectively, they only identified 14 (i.e., 8 and 6 abstracts, respectively) of the abstracts not contained in the original seven topic areas. It could easily be argued that these two research areas are sub-topics of the seven. However, because of specific interests of the authors, the two new topic areas were included in the remaining comparative analysis. The nine resulting topic areas captured 324 of the 354 multi-robot abstracts. The 30 non-grouped abstracts were put into a tenth group entitled "IEEE OTHER."

The co-occurrence matrix, Table 2, shows the number of common abstracts among the 9 topical groupings and total membership numbers for each. This grouping compilation was accomplished manually and at considerable cost in time, taking 8 hours to isolate relevant terms and phrases from the multi-robot overview paper and perform the edit/find groupings. No multi-robot subject matter expert involvement was available for the relevant term selection process, so field experts might argue that some relevant terms have been omitted from the grouping process.

Table 2: Multi-robot research topic abstracts groups' co-occurrence matrix.

		IEEE MultiRobot Groups									
	# Records	142	139	114	61	61	44	42	41	30	12
# Records	IEEE MultiRobot Groups	Architecture Allocation Control	Motion Coordination	Communication	Localization Mapping Exploration	Biological	Transport Manipulation Grasping	Robot Learning	Human Interface	IEEE OTHER	Reconfigurable
142	Architecture Allocation Control	142	55	45	27	27	24	14	9		4
139	Motion Coordination	55	139	38	20	22	17	16	19		4
114	Communication	45	38	114	31	25	10	7	12		2
61	Localization Mapping Exploration	27	20	31	61	9	8	5	8		4
61	Biological	27	22	25	9	61	6	5	5		2
44	Transport Manipulation Grasping	24	17	10	8	6	44	7	6		4
42	Robot Learning	14	16	7	5	5	7	42	4		
41	Human Interface	9	19	12	8	5	6	4	41		2
30	IEEE OTHER									30	
12	Reconfigurable	4	4	2	4	2	4		2		12

For the 354 abstracts in the multi-robot research file, Publication Year was subdivided into three periods: 1998-99, 2000-01 and 2002-03, containing 136, 111 and 107 technical paper abstracts, respectively. We noted that Japanese sources have published more, (even given the English language slant of the two databases!) than U.S. sources over the last 6 years. Albeit, over half of the Japanese abstracts were from the earliest period, 1998-99. This Japan--USA comparison motivated the cross-field expectancy measure comparative analysis.

Data Mining V, A. Zanasi, N. F. F. Ebecken & C. A. Brebbia (Editors)
© 2004 WIT Press, www.witpress.com, ISBN 1-85312-729-9

During each of the periods, 1998-99, 2000-01 and 2002-03, Japanese sources published 66, 24 and 30 (i.e. 120 total) multi-robot research papers and USA sources published 32, 39 and 42 (i.e., 113 total), respectively.

5 Automated grouping analysis

The Tech OASIS software extracts latent relationships through an automated algorithm based on principal components analysis (PCA) and can represent the derived relationships graphically. PCA, an inductive approach, does not impose groupings, as we did with the IEEE multi-robot groupings; but, instead, elicits them from the data. PCA factor groupings, however, are not singular for a set of literature abstracts (i.e., user selection of the terms and/or phrases to be clustered and the number of factors to be considered will generate alternative cluster solutions and relationship displays) [5]. To overcome this problem, we have developed a PCA-based algorithm, principal components decomposition (PCD), to standardize the selection of the terms and/or phrases and the number of factors to be generated in the PCA.

The *INSPEC* and EI Compendex descriptors field, a multiple-entry field within the analyzed abstracts, provides the terms and phrases for the cluster analysis documented herein. The descriptors field contains roughly 5 to 8 keywords, generated to reflect the content of each abstracted research paper.

The PCD algorithm applies metrics that provide a standard solution to a min-max problem. This analysis process includes as many of the abstract records in the derived PCA factor groupings as possible. In addition, the PCD algorithm strives to maximize both the number of factor groups and group-defining descriptors (i.e., the high-loading keywords for each PCA factor group). The algorithm also minimizes the duplication of constituent abstracts among the derived factor groupings. This min-max solution approach conceptually, equates to minimizing the entropy and maximizing the cohesiveness of the derived factor groups [5, 6]. For any user, given identical abstract files, the automated PCD analysis will repeatedly derive the same set of factor groups.

6 Expectancy measure

This research strives to determine and display the likelihood of an item in one field having x or more records in a specific category in a second field. Figure 1 illustrates a screen shot of the Tech OASIS view of the Country field (i.e. the central section of Figure 1). This depicts USA and Japan (and other) authored abstracts for the 2002-03 period -- 42 and 30 abstracts, respectively. We have "selected" Japan in the field view. To the left, we see the titles of these 30 papers. If desired, we could "double-click" on a given title to view that abstract record.

Detailed view breakouts appear on the right side of Figure 1 for both the IEEE multi-robot research categories and the PCD-derived groups. The two columns preceding the group names indicate the number of Japanese abstracts in, and the expectancy measures for, the respective group. This expectancy measure



Data Mining V, A. Zanasi, N. F. F. Ebecken & C. A. Brebbia (Editors)
© 2004 WIT Press, www.witpress.com, ISBN 1-85312-729-9

indicates how much the item's number of common records in the detail window departs from expectation based on overall dataset base rates of the main field item (i.e., Japan) and that group (e.g., the first listed category, motion coordination). A high, positive number indicates that the item's number of records in the detail window is higher than expected. Likewise, a large, negative number indicates that the item's frequency is lower than expected. Specifically, from Figure 1, Japanese sources have authored 19 abstracts relating to motion coordination, which has a total group membership of 46 abstracts (not shown). The expectancy measure is 0.98. The frequency is thus higher than expected -- Japanese research appears to emphasize this topic.

The expectancy calculation follows:

Expectancy Measure = 1 – probability that the item has n or greater records.

Similarly, if the number of common abstracts is lower than expected, the probability value is:

Expectancy Measure = - (1 – probability that the item has n or less number of records).

The probability is calculated using the cumulative binomial distribution and uses the detail view item dataset frequency to define the binomial distribution success probability.

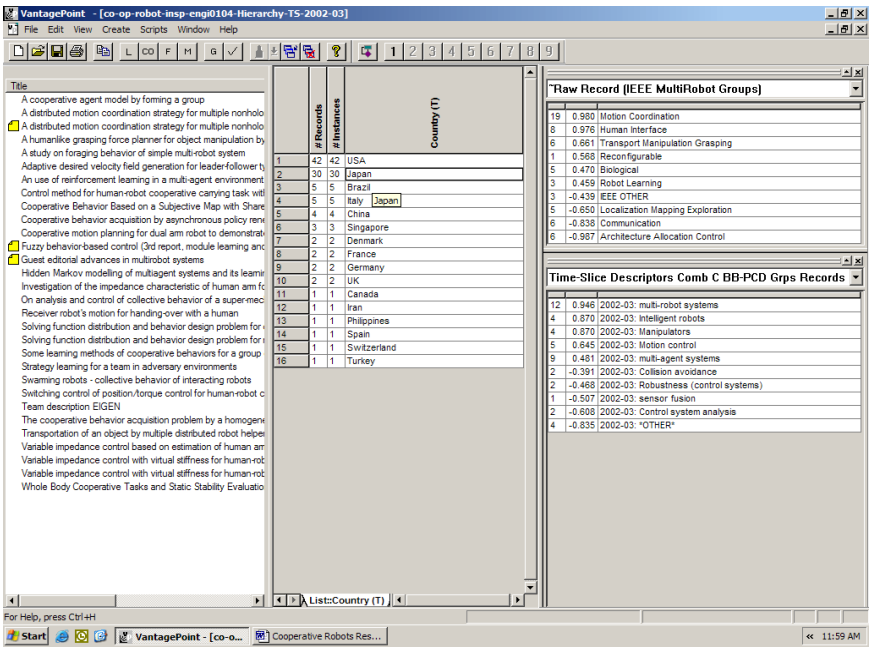


Figure 1: Tech OASIS screen view.

© 2004 WIT Press, www.witpress.com, ISBN 1-85312-729-9

7 Multi-robot research abstracts analysis

We performed PCD time-slice analyses on the cleaned descriptors field. We then created a field from those groups. Viewing the Countries field and having detail views of the time-slice PCD groups and the IEEE multi-robot overview groups fields (as shown in Figure 1), information was extracted, using the Tech OASIS copy command, to construct Tables 3 and 4. Table 3 shows the IEEE multi-robot groups for each time period, each group's number of abstracts and expectancy measures for both Japan and USA sources. Note that Japanese sources' IEEE multi-robot abstract group information (upper right detail view of Figure 1), is incorporated in the lower left portion of Table 3. Similarly, the lower right detail view of Figure 1, listing PCD group statistics for Japanese sources, is incorporated in the lower left portion of Table 4. Table 4 shows the Tech OASIS time-slice PCD categories, categories' number of abstracts and expectancy measures for both Japanese and USA sources.

Table 3: IEEE multi-robot topic areas' emphasis measures for USA and Japan.

Japan Sources					USA Sources				
#	# Grp	Exp.	Metric	Cluster Group	#	# Grp	Exp.	Metric	Cluster Group
1998-99					1998-99				
14	18	0.952	4.28	Human Interface	1	1	0.79	0.79	Reconfigurable
7	9	0.855	3.85	IEEE OTHER	6	18	0.757	1.14	Localization Mapping Exploration
12	18	0.843	2.53	Robot Learning	14	50	0.741	1.03	Architecture Allocation Control
8	13	0.706	1.84	Transport Manipulation Grasping	2	9	-0.356	-1.60	IEEE OTHER
22	44	0.522	1.04	Communication	3	13	-0.366	-1.59	Transport Manipulation Grasping
9	18	0.484	0.97	Localization Mapping Exploration	4	18	-0.42	-1.89	Human Interface
23	50	-0.573	-1.25	Architecture Allocation Control	4	18	-0.42	-1.89	Robot Learning
20	54	-0.926	-2.50	Motion Coordination	12	54	-0.525	-2.36	Motion Coordination
11	34	-0.927	-2.87	Biological	7	34	-0.568	-2.76	Biological
	1			Reconfigurable	9	44	-0.618	-3.02	Communication
2000-01					2000-01				
13	39	0.956	1.43	Motion Coordination	7	9	0.965	4.34	Reconfigurable
3	10	0.631	0.90	Human Interface	5	10	0.727	1.45	Human Interface
4	14	0.641	0.90	Robot Learning	16	38	0.768	1.33	Communication
2	9	0.41	0.53	IEEE OTHER	4	9	0.61	1.10	IEEE OTHER
3	14	-0.359	-1.68	Transport Manipulation Grasping	4	10	0.529	0.88	Biological
7	38	-0.613	-3.33	Communication	7	20	-0.405	-1.16	Localization Mapping Exploration
9	48	-0.638	-3.40	Architecture Allocation Control	4	14	-0.556	-1.95	Robot Learning
2	20	-0.833	-8.33	Localization Mapping Exploration	3	14	-0.742	-3.46	Transport Manipulation Grasping
	10			Biological	15	48	-0.668	-2.14	Architecture Allocation Control
	9			Reconfigurable	10	39	-0.859	-3.35	Motion Coordination
2002-03					2002-03				
8	13	0.976	2.54	Human Interface	21	32	0.995	2.89	Communication
19	46	0.98	1.67	Motion Coordination	20	44	0.759	1.39	Architecture Allocation Control
1	2	0.568	1.14	Reconfigurable	11	23	0.718	1.38	Localization Mapping Exploration
6	17	0.661	1.02	Transport Manipulation Grasping	4	12	-0.515	-1.55	IEEE OTHER
5	17	0.47	0.67	Biological	5	17	-0.676	-2.30	Biological
3	10	0.459	0.66	Robot Learning	5	17	-0.676	-2.30	Transport Manipulation Grasping
3	12	-0.439	-1.76	IEEE OTHER	1	10	-0.914	-9.14	Robot Learning
5	23	-0.65	-2.99	Localization Mapping Exploration	2	13	-0.899	-5.84	Human Interface
6	32	-0.838	-4.47	Communication	15	46	-0.786	-2.41	Motion Coordination
6	44	-0.987	-7.24	Architecture Allocation Control	0	2			Reconfigurable

In both Tables 3 and 4, we have highlighted the six highest emphasis areas based on each of three alternative metrics:

- the absolute number of abstracts
- the expectancy measures and
- an added metric for both countries' research sources.

We do so for comparative purposes.

Data Mining V, A. Zanasi, N. F. F. Ebecken & C. A. Brebbia (Editors)
© 2004 WIT Press, www.witpress.com, ISBN 1-85312-729-9

Contrast what these metrics say about Japanese and American research emphases. In Table 3, note the six highlighted Japanese groups based on most papers (the first column). Compare with the six highlighted "surprising" expectancy groups -- three are the same and three differ. For instance, the 22 Japanese abstracts relating to multi-robot *Communication* are not exceptional, given the base rates, according to our expectancy measure. Conversely, the 12 abstracts relating to *Robot learning* are more than we would have expected given the overall number of papers on robot learning and the overall number of Japanese authored papers.

Table 4: Tech OASIS time-slice PCD categories' emphasis for USA and Japan.

Japan Sources					USA Sources				
#	# Grp	Exp.	Metric	Cluster Group	#	# Grp	Exp.	Metric	Cluster Group
11	16	0.852	2.73	1998-99: Human computer interaction	8	16	0.971	1.94	1998-99: real-time systems
6	10	0.643	1.61	1998-99: Adaptive control systems	5	12	0.853	1.46	1998-99: Robot learning
7	12	0.636	1.53	1998-99: Robot learning	4	10	0.794	1.32	1998-99: Adaptive control systems
6	11	0.554	1.22	1998-99: Position control	6	21	0.628	0.88	1998-99: *OTHER*
10	21	-0.443	-0.93	1998-99: *OTHER*	3	11	0.515	0.71	1998-99: Position control
21	44	-0.478	-1.00	1998-99: learning (artificial intelligence)	4	16	0.471	0.63	1998-99: Human computer interaction
12	26	-0.502	-1.09	1998-99: Intelligent control	6	26	-0.414	-1.79	1998-99: Intelligent control
18	43	-0.731	-1.75	1998-99: cooperative systems	6	44	-0.932	-6.83	1998-99: learning (artificial intelligence)
6	16	-0.671	-1.79	1998-99: real-time systems	5	43	-0.966	-8.31	1998-99: cooperative systems
5	13	0.858	1.39	2000-01: Robot programming	11	24	0.793	1.46	2000-01: *OTHER*
8	25	0.847	1.25	2000-01: Algorithms	6	13	0.696	1.29	2000-01: Robot programming
6	20	0.745	1.06	2000-01: Computer simulation	7	20	-0.405	-1.16	2000-01: Computer simulation
8	34	0.538	0.70	2000-01: multi-robot systems	10	30	-0.495	-1.49	2000-01: multi-agent systems
3	12	0.511	0.68	2000-01: Distributed parameter control systems	7	25	-0.679	-2.43	2000-01: Algorithms
7	30	0.517	0.67	2000-01: multi-agent systems	3	12	-0.62	-2.48	2000-01: Distributed parameter control systems
4	19	-0.394	-1.87	2000-01: Manipulators	5	19	-0.678	-2.58	2000-01: Manipulators
4	24	-0.617	-3.70	2000-01: *OTHER*	9	34	-0.8	-3.02	2000-01: multi-robot systems
4	7	0.87	2.03	2002-03: Intelligent robots	5	6	0.916	5.50	2002-03: sensor fusion
4	7	0.87	2.03	2002-03: Manipulators	14	24	0.93	2.23	2002-03: *OTHER*
12	27	0.946	1.70	2002-03: multi-robot systems	4	11	-0.436	-1.20	2002-03: Control system analysis
5	14	0.645	1.00	2002-03: Motion control	3	9	-0.475	-1.43	2002-03: Robustness (control systems)
9	31	0.481	0.68	2002-03: multi-agent systems	2	7	-0.524	-1.83	2002-03: Intelligent robots
2	8	-0.391	-1.56	2002-03: Collision avoidance	10	31	-0.709	-2.20	2002-03: multi-agent systems
2	9	-0.468	-2.11	2002-03: Robustness (control systems)	2	8	-0.617	-2.47	2002-03: Collision avoidance
1	6	-0.507	-3.04	2002-03: sensor fusion	8	27	-0.768	-2.59	2002-03: multi-robot systems
2	11	-0.608	-3.34	2002-03: Control system analysis	3	14	-0.817	-3.81	2002-03: Motion control
4	24	-0.835	-5.01	2002-03: *OTHER*	1	7	-0.77	-5.39	2002-03: Manipulators

We generate a special "metric" as a third measure in Tables 3 and 4. The metric strives to add context to the expectancy measure in regards to the amount of research and the level of other sources' research in each group. The metric equals the expectancy measure multiplied by the amount of research in the category (i.e., the number of abstracts in the group) and, then, divided by the amount of outside research, if the expectancy measure is positive, or divided by the number of source abstracts, if the expectancy measure is negative. The premise on applying the metric is that having more abstracts than expected in a research area where there is a large amount of other sources' research documented is less important than in an area having little research from other sources. Conversely, having less research documented than expected in an area in which the source has almost no publications is more important than in an area

Data Mining V, A. Zanasi, N. F. F. Ebecken & C. A. Brebbia (Editors)
© 2004 WIT Press, www.witpress.com, ISBN 1-85312-729-9

where the source has several publications. For instance, in Table 3 we see that use of the metric de-emphasizes the importance of USA 1998-99 *reconfigurable* papers, but raises the importance of USA 2000-01 *human interface* activity. Use of the metric did not change group importance in Table 4 for the USA sources from that based on Expectancy. Research group importance changes for Japanese sources can be observed in Tables 3 and 4. Please note that the importance is based on deviations from expectations. Thorough analysis of Japanese vs. American multi-robot research emphases would also consider absolute number of publications in each category. These results are best used as stimuli for expert discussion and interpretation.

8 Observations

We first turn our attention to the IEEE expert defined categories of multi-robot research. Japanese sources show less emphasis than expected on *biological approaches* and *reconfigurable robots* and more emphasis than expected on *human-robot interface, motion coordination* and *robot learning*. This is based on the expectancy measure, the metric and the near-absence of abstracts for biological approaches and reconfigurable robots (Table 3). The US sources' research shows less emphasis than expected on *robot learning, motion coordination, human interface* (i.e., in 2002-03) and *transport manipulation grasping*. USA sources have more emphasis than expected on *reconfigurable robots, communication allocation control* and *localization mapping exploration*.

Turning to the PCD groups shown in Table 4, we observe the 2000-01 and 2002-03 *OTHER* groups have more USA sources' abstracts than expected. Prior research showed different, sometimes cutting edge, research could often be found in *OTHER* PCD derived group papers [7]. The PCD cluster groups are based on consensus usage (i.e., relational analysis) of terms. Abstracts with non-consensus descriptors were put into the *OTHER* group. The Japanese sources' *OTHER* categories had high negative expectancy. What does this mean? One might surmise that the non-consensus, different, Japanese research has not been published in English; or possibly, not at all. Another interpretation might be that Japanese research is more focused, an indicator of successful innovation [8]. Or, a third possibility, perhaps this pattern reflects a "close follower" national R&D strategy. Expert opinion would be needed to make a final judgment on the observed anomalies, since only open-source information has been included in this analysis.

Table 3 compares Japanese and American research emphases based on imposed categories (a "deductive" approach). Table 4 compares their research emphases based on an "inductive" approach -- PCD. We see advantages in this dual perspective, offering additional chances to generate useful insights. One can peruse these tables looking for:

- General pattern similarity (Do these two leading research nations appear to pursue similar multi-robot developments?)
- Pattern differences (Where do we differ from them?)

Data Mining V, A. Zanasi, N. F. F. Ebecken & C. A. Brebbia (Editors)
© 2004 WIT Press, www.witpress.com, ISBN 1-85312-729-9

- Gap analyses (Are there areas of special emphasis in the other country that might warrant research stimulation in our country?
- Consider the Japanese lack of research in the first two time slices (but not in the most recent) on biological and reconfigurable robots.
- Topical strength (For a given topic area, which research community appears most active?)

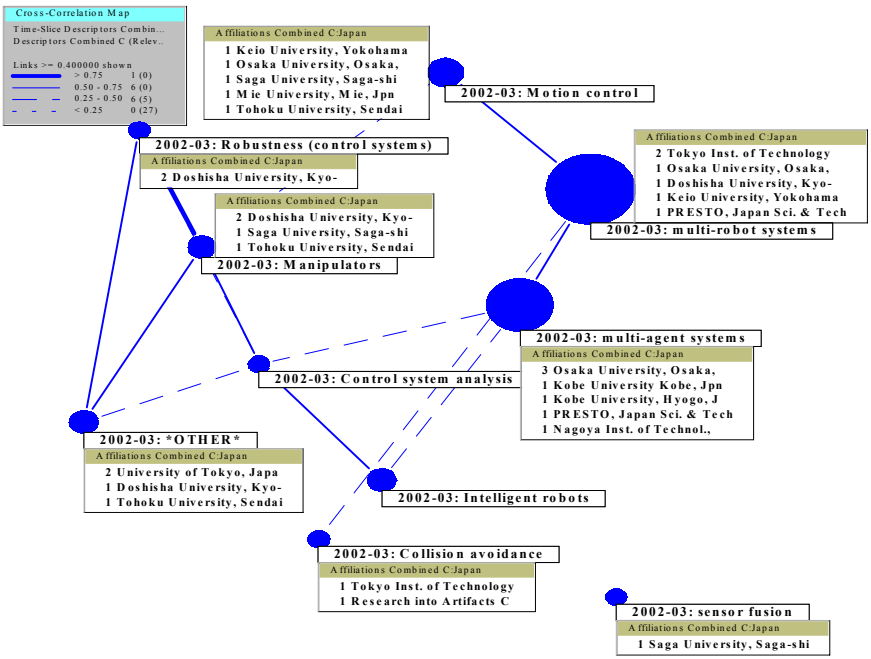


Figure 2: Japanese sources 2002-03 multi-robot research.

In making these observations, one must remember that any conclusions reached would only relate to the search subject, multi-robot research, as defined by the search strategy in Table 1. However, the expectancy measure calls out research areas that are not as expected and might warrant attention by field researchers. In making such investigations, other measures of the research information and other information sources should be considered too.

One such additional approach would be to evaluate the changes of group titles, which depict possible conceptual content, over time. The PCD time-slice groups' names provide information to help gauge the maturation of the multi-robots research. In Table 4, we see *position control* as a group in 1998-99, as opposed to *motion control* (i.e., a higher level of technology maturity) in 2002-03. In 1998-99, research clusters in a generic research area, *adaptive control systems*. In the latter periods, research groups in a more specific research area, *distributed parameter control systems*, in 2000-01, and then in *control system*

© 2004 WIT Press, www.witpress.com, ISBN 1-85312-729-9

analysis and *robustness (control systems)* in 2002-03. Analysis of the conceptual content can provide a subjective measure of the maturity level of a field of research and should be considered along with other measures.

In prior research, we have found that technology maturity could be gauged by an increasing similarity of research descriptors among constituent technologies (i.e., cluster groups) within a field of research. The descriptors reflect the content of the research. Theoretically, as a technology matures, the content (i.e., descriptors) will reflect increasing systems technology integration issues. As a technology matured, map linkages (depicting increased relatedness of each cluster groups' composite set of descriptors to those of the other groups) would increase. As stated earlier, the PCD algorithm strives to group abstracts such that the number of common abstracts among the groups is minimized. That serves to reduce the relatedness among groups.

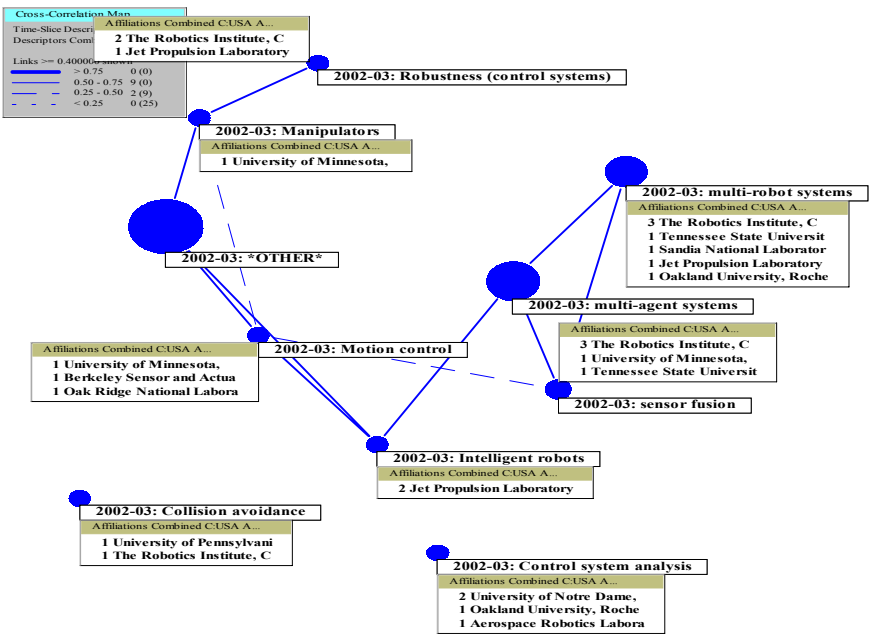


Figure 3: USA sources 2002-03 multi-robot research.

As a corollary hypothesis, assume competing technologies appear as separate clusters on a map. Then, increased linkages over time to one, rather than the other, might point to the wining technology. Such clues can be critical in making astute choices of which technological approach to pursue for future applications.

Figures 2 and 3 depict the 2002-03 PCD group relatedness maps for Japanese and USA sources, respectively (based on descriptors -- i.e., keywords). Both maps reflect a moderate level of maturity, having many cluster groups linked, but at moderate relatedness levels. In the upper left corner of each map, is a

© 2004 WIT Press, www.witpress.com, ISBN 1-85312-729-9

summary box showing the number of links for each of 4 relatedness levels. Japanese sources have only one relatedness link above 0.75, that being between *manipulators* and *robustness (control systems)*. The *manipulators* group had a high expectancy measure for Japanese sources. Does this indicate that Japanese manipulator research is more advanced than American? The USA sources have 9 links greater than .5, as compared to 6 for Japanese sources. Each country's sources have 11 links greater than .25. Does this indicate that multi-robot research is about equally mature in both countries? Probably so.

We can also explore relatedness concerning a topic of particular concern to us. Note for USA sources (Figure 3), *sensor fusion*, a group with a high expectancy measure (Table 4), is related to *multi-robot systems*, *multi-agent systems* and *motion control*. *Control system analysis* and *collision avoidance* (Figure 2) are more related to the other multi-robot research groups for Japanese sources than for USA sources (Figure 3). Is American multi-robot, multi-agent research more focused on sensor fusion and intelligent robots? Whereas, is Japanese research more focused on motion control, control system analysis and collision avoidance? These questions can spark creative dialog among field researchers.

Tech OASIS automates the clustering/categorization of information to enable and improve cognition of broad fields of research, elicit research questions from noted anomalies and promote innovation through expert involvement in the analysis.

9 Conclusions

We provide a brief background on research profiling and the software tool, Tech OASIS. We apply this form of text mining to multi-robot research data. We focus on newly developed expectancy measures that help ascertain which topical activity levels are relatively high or low. We then present results drawn from the multi-robot research information analyses.

Comparing the high and low expectancy-based-metric research categories for the inductive (i.e., PCD) cluster groups, Table 4, and deductive (based on the expert IEEE review) topic areas, Table 3. Research categories, deserving of subject matter expert investigation, include Japanese research in human interface, transport manipulation grasping, and robot learning. The cause of the low expectancy measures for Japanese research in the "OTHER" categories also begs for further investigation. Is the Japanese multi-robot research analysis a depiction of a close-follower strategy, or selective publication of only mainstream research?

In assessing the implications of this research information profiling, a holistic approach must be stressed. Pervasive results from multiple tools must be judged more meaningful, in comparison to the output from a single tool. For example, the USA 2000-01 and 2002-03 *multi-robot systems* groups both have low expectancy measures. However, both groups are among the USA sources' top six classification groups, as observed in Table 4, based on the absolute number of

Data Mining V, A. Zanasi, N. F. F. Ebecken & C. A. Brebbia (Editors)
© 2004 WIT Press, www.witpress.com, ISBN 1-85312-729-9

publications. Other dichotomies can be observed and are best resolved through the involvement of technical experts.

Because the group names remain the same for all time periods, it is easier to generalize observations using the IEEE groupings. The IEEE multi- robot group names might also be more consistent with experts' perceptions of the field of research. However, the human defined group names would most probably be biased toward the current research period. As we demonstrated, the evolution of the PCD time-slice group names' terminology provides information to help gauge the maturation of the multi-robots research. In addition, the PCD derivation process strives to minimize cross-group relatedness; thereby, making time-sequenced relatedness maps meaningful for comparative purposes. In the future, an alternate analysis approach might be to derive the PCD cluster categories using only the abstracts from authors having multiple publications in the research field (i.e., the field experts). The categories could then be transposed into the complete file for the sources comparative analyses.

We suggest these tools and the new expectancy measures enrich the opportunities to derive useful technological intelligence from text mining [9]. The expectancy measures help identify unusual patterns of research emphasis. Those can help compare the activity of one country or organization with its competitors. This can fuel gap analysis to identify weaknesses in one's R&D portfolio, leading toward strategies to remedy same. It can also suggest collaborative opportunities. Note also that these text mining tools enable both profiling of the main activity emphases and discerning "outlier" activity. That can help alert one to novel approaches and emerging opportunities.

References

[1] van Raan, A.F.J. (1992). "Advanced Bibliometric Methods to Assess Research Performance and Scientific Development: Basic Principles and Recent Practical Applications," Research Evaluation, 3(3): 151-166 See also website: http://sahara.fsw.leidenuniv.nl/cwts/nnmap0.html).

[2] Porter, A.L., Roper, A.T., Mason, T.W., Rossini, F.A., and Banks, J.: *Forecasting and Management of Technology*. Wiley, New York, NY, 1991.

[3] Losiewicz, P., Oard, D.W., and Kostoff, R.N.: Textual data mining to support science and technology management, *Journal of Intelligent Information Systems* 15(2), 99-119 (2000).

[4] Arai, Tamio, Pagello, Enrico and Parker, Lynne: "Guest Editorial, Advances in Multirobot Systems," IEEE Transactions on Robotics and Automation, Vol. 18, No. 5, October 2002.

[5] Börner, K., Chen, C., & Boyack, K. W. (2003). Visualizing knowledge domains. *Annual Review of Information Science and Technology* 37, 179-255.

[6] Steinbach, M., Karypis, G., Kumar, V., "A Comparison of Document Clustering Techniques" University of Minnesota, Technical Report #00-034 (2000). http://www.cs.umn.edu/tech_reports/.



© 2004 WIT Press, www.witpress.com, ISBN 1-85312-729-9

[7] Watts, R.J., Courseault, C.R., Kaplin, S.J., "Identifying Unique Information Using Principal Component Decomposition," International Association for Management of Technology, Miami, 2000.
[8] Souder, W. E., *Managing New Product Innovations*, Lexington Books, New York, 1987
[9] Porter, A.L., Cunningham, S.W., Tech Mining: Technology Management through Information Mining, Wiley, New York, to appear.

Data Mining V, A. Zanasi, N. F. F. Ebecken & C. A. Brebbia (Editors)
© 2004 WIT Press, www.witpress.com, ISBN 1-85312-729-9