Technological Forecasting & Social Change 75 (2008) 1349–1359

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

Structure of research on biomass and bio-fuels: A citation-based approach

Yuya Kajikawa*, Yoshiyuki Takeda

Institute of Engineering Innovation, School of Engineering, the University of Tokyo, 2-11-16 Yayoi, Bunkyo-ku, Tokyo 113-8656, Japan

ARTICLE INFO

Article history:
Received 3 September 2007
Received in revised form 31 January 2008
Accepted 29 April 2008

Keywords:
Biomass
Bio-fuel
Citation network
Bibliometrics
Sustainable energy
Renewable energy

ABSTRACT

Biomass and bio-fuels have gained a growing interest as sustainable and renewable energy. In this paper, we perform a citation network analysis of scientific publications to know the current structure of biomass and bio-fuel research. By clustering and visualizing the network, we revealed their taxonomic structure. Emerging technologies are detected by analyzing the average publication year of clusters. According to the results, bio-diesel and hydrogen production are the most rapidly developing domains among biomass bio-fuel researches. We also analyzed the position of each cluster in the global structure of research. By using citation counts within and out of the cluster, we categorized each cluster into the following four categories: (I) topic specific; (II) domain specific; (III) global link; and (IV) specific & global. For research domains of category (III) or (IV), it is difficult that single technology overcomes the current limitation of bio-energy productions. Research on lignocellulose feedstock is a typical case where knowledge from other scientific disciplines is necessary.

1. Introduction

Energy is a key element for our society and also a key input for economic development. Biomass and bio-fuels have attracted growing interest as sustainable and renewable energy [1–5]. Biomass is a stored source of solar energy initially collected by plants and includes variant organic materials produced from plants and animals. Biomass feedstock such as crops, trees, and food waste can be converted into more convenient energy carriers such as solid fuels (e.g., wood chips, pellets, briquettes), liquid fuels (e.g., methanol, ethanol, bio-diesel, bio-oil), or gaseous fuels (synthesis gas, biogas, hydrogen).

Among them, liquid bio-fuels such as bio-ethanol and bio-diesel are suitable for transportation purposes and they have therefore received substantial attention not only from academia but also in the business scene [6–14]. This is because liquid hydrocarbons are well suited for transport uses because of their high energy density and handling convenience. On the other hand, crude oil-based fuels such as gasoline and diesel fuels are mainly used for vehicles traveling over land, kerosene for aircraft, and heavy fuel oils for ships. But as the increasing industrialization and motorization of the world has led to a steep rise in the demand of petroleum-based fuels, it is becoming a key issue for us to produce supplementary fuels from biomass that can be used without requiring substantial modification of existing vehicles or of the fuel distribution infrastructure.

As crude oil and gasoline prices increase and emission regulations become more stringent, ethanol could be given more attention as a renewable fuel or gasoline additive [6]. Ethanol was first suggested as an automotive fuel in the USA in the 1930s, but was widely used only after 1970. Currently, ethanol for the fuel market is produced from sugar (Brazil) or starch (USA) at competitive prices.

To promote the permeation of bio-energy into the market and enhance R&D efforts, affirmative actions are adopted in the USA, the EU, Japan, and so on. The European Commission plans to substitute progressively 20% of conventional fossil fuels with alternative fuels in the transport sector by 2020, with an intermittent goal set at 5.75% by 2010. In the USA, the Energy Policy Act of 2005 requires blending of 7.5 billion gallons of alternative fuels by 2012, and the US president, in his State of the Union address, set the goal of replacing more than 75% of imported oil with alternative fuels by the year 2025.

---

* Corresponding author. Tel./fax: +81 3 5841 7672.
E-mail address: kaji@biz-model.t.u-tokyo.ac.jp (Y. Kajikawa).

0040-1625/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.techfore.2008.04.007

In the current fast-changing situation, R&D managers and policy makers have to comprehend global political issues, markets, and R&D efforts in an effective and efficient manner. In this paper, we focus on global R&D efforts in bio-fuels among those, because in today's increasingly knowledge-based economy, growth more reliably depends, or is expected to depend, on innovation generated by the application of new science and technology. But it is not a rudimentary task for R&D managers and policy makers to comprehend global R&D structures efficiently because of the flood of information. Generally, there are two approaches to obtaining comprehensive perspectives on R&D [15]. One straightforward manner is the expert-based approach, which utilizes the implicit knowledge of domain experts. The other is the computer-based approach, which analyzes explicit knowledge such as newspapers, magazines, and academic papers. Because the former approach is becoming a highly difficult task due to the increasing rate of information production, the latter approach should be exploited to support the former. A major advance of the past few years in technological forecasting has been using a computer to support the expert-based approach. Data mining (DM) and database tomography (DT) have become practical techniques for assisting the forecaster to identify the taxonomic structure of a research domain [16–20].

In a previous paper, a research map on energy research was investigated [20], and emerging technology trends were detected [20] using the DT approach. But investigations specifically on biomass or bio-fuels are scarce. In a previous report, the number of journal articles and conference proceedings on biomass were counted [21], but further details are not analyzed. The aim of this paper is to offer a global structure of research on biomass and bio-fuels to assist effective policy development for research. We have mainly three objectives. These include the following;

i) to identify emerging technologies and research fronts
ii) to make recommendations on which science and technology areas to develop more rapidly, and
iii) to study the structure of the biomass and bio-fuels literature, along with the inter-thrust dynamics. Before describing our approach, we illustrate the current status surrounding biomass and bio-fuels in the following.

2. Following and opposing winds for bio-fuels

Currently, there are both following and opposing winds regarding bio-fuels. The driving force includes surging global oil demand accompanied by the increasing costs of finding and producing new reserves, the geographical concentration of known petroleum reserves, and growing concerns about atmospheric greenhouse gas (GHG) concentrations. The situation is schematically shown in Fig. 1.

On the first point, it is generally accepted that one of the greatest challenges for society in the 21st century is to meet the growing demand for energy for transportation, heating, and industrial processes, and to provide raw material for the industry in a sustainable way [22]. However, oil production will peak and start to decline as the term “peak oil” is now used to designate these days. Fluctuating on a daily basis, the price of crude oil keeps rising and has approached US$80 per barrel since 2006. Fuels of bio-origin can provide a feasible solution to this worldwide petroleum crisis.

Secondly, the finite reserves of petroleum-based fuels are highly concentrated in certain geographical regions of the world. Therefore, pipelines are needed for distribution from “'area found” to “area needed.” Countries not having these resources have problems in energy security. Hence, it is desirable to look for alternative fuels that can be locally available and produced. Bio-fuels are geographically more evenly distributed than the fossil fuels; thus, the sources of energy will, to a large extent, be domestic and provide security of supply.

Thirdly, the future energy supply must be met with a simultaneous substantial reduction in green house gas emissions. Combustion of fossil fuels has played a dominant role in the emission of GHG into the atmosphere. There is an urgent need to develop highly efficient energy utilization processes and substitute energy sources as a countermeasure against the CO_2 problem. At the Kyoto conference on global climate change, nations the world over committed to significantly reducing GHG emissions. We have several options for sustainable energy: solar, wind, hydro, geothermal, and biomass energy. Bio-energy is a carbon-free process and can contribute to reducing the GHG effect if the emitted CO_2 is captured by plants.

Driven by these following winds, some researchers expect that biomass contribution will reach 7–30% of total energy requirements without affecting food production [3,23]. For example, Kim and Dale estimated that the potential for ethanol production is equivalent to about 32% of the total gasoline consumption worldwide, when used in E85 (85% ethanol in gasoline) for

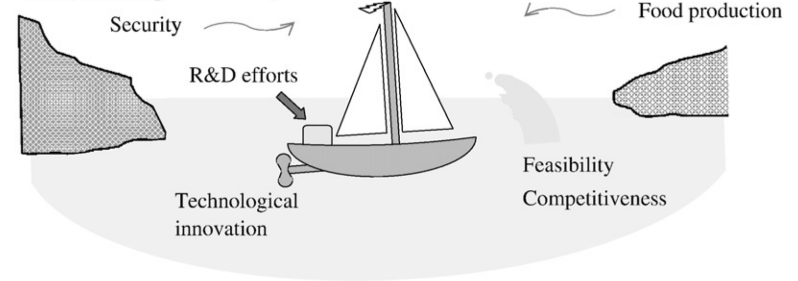

Fig. 1. Following and opposing wind for bio-fuels.

a midsize passenger vehicle [24]. However, such optimistic perspectives on the future of bio-fuels [23–26] are repeatedly criticized as unreasonable [27–30]. Despite the benefits listed above, it seems to take a long time for mass production and utilization of biomass to be realized as renewable energy because of the low energy and economic feasibility.

At first, the energy balance of a bio-energy project is not always favorable. This is particularly the case for some bio-fuels produced from annual energy crops, where energy unit inputs into the overall system can be similar to, or even exceed, the energy unit outputs in the bio-energy product [31,32]. The output-to-input energy yield ratio for ethanol production from corn is only 1.08 [31], while from other oilseed crops such as soybean or rapeseed, it is 3, and from oil palm, which is a highly productive crop and is the world's leading vegetable oil by production and trade volume [33], it reaches 9 [34]. Such an energy balance is still under discussion [14]. Schmidt concisely pointed out that the biggest headache for those developing bio-fuels is the stark contrast between what we have (biomass rich in carbohydrates) and what we want (oxygen-deficient fuels) [35].

Another shortcoming is their relatively high production cost [2,36]. According to an IEA report [2], commercial bio-ethanol production costs, without any agricultural subsidies, direct grants, or other government incentives included, currently range from 0.25 US$ per liter of gasoline equivalent (lge), (sugarcane, Brazil) to 0.80 US$/lge (sugar beet, UK) with corn ethanol around 0.60 US $/lge (USA) and lignocellulosic ethanol from pilot-scale plants claimed to be between 0.80 and 1.00 US$/lge. Bio-diesel costs range from 0.42 US$/l (animal fats, New Zealand) to 0.90 US$/l (oilseed rape, Europe; soybean, USA; palm oil; Malaysia). Technology development and larger-scale plants could lower the production costs of bio-ethanol by 2030 to 0.23–0.65 US$/lge and bio-diesel to 0.40–0.75 US$/lge. Patzek is critical of the fact that the recent growth in ethanol production could only occur because of the massive transfer of money by federal subsidies of corn producers as well as federal and state tax subsidies of ethanol producers from the collective pocket of US taxpayers [32]. Taxpayers' subsidies of the industrial corn-ethanol cycle were estimated to be $3.8 billion in 2004. Parallel subsidies by the US government were estimated to be $1.8 billion in 2004.

In addition to the fact that bio-fuels must compete with other energy sources in cost, bio-fuels compete with the traditional agricultural production of foods and fibers for land, water, and labor use. Some insist that although bio-fuel's contribution as an energy source can be positive, it will remain small, being restricted by the ability of the natural environment to provide both fuel and food for a large and energy-demanding world population [13,27,32,37–39].

Fig. 1 represents a concise summary of this situation surrounding bio-fuels. Energy deficiency, global warming, and the energy security problem assist the development of bio-fuels, while bio-fuels must achieve energy and economic feasibility and gain competitiveness against other complementary energy sources as well as agricultural sectors such as food and textile production. But we must keep in mind that current options of renewable energy such as biomass and solar cells cannot solely meet increasing global energy demand, and biomass and bio-fuels are possible solutions for future energy sources. To realize this, R&D efforts to attain technological innovation are definitely required. And public policy can play an important part in the development and diffusion of bio-energy technologies by crystallizing these efforts and facilitating financing [40]. As noted by Herrera, nobody can blame entrepreneurs, scientists, or engineers this time if bio-fuels fail [41]. As global energy demand is projected to continue to grow, all options should be pursued.

Currently, there is a growing body of research efforts and publications to improve the feasibility of biomass and bio-fuels. Investment for technological innovation is an inevitable first step for future bio-energy systems. For effective research planning and investment, we have to grasp the broader coverage of scientific and technological research. Depicting the global structure of bio-fuels is the inevitable first step for this and can be assisted by a DM and DT approach. In order to visualize the global structure of biomass and bio-fuel research, we perform a citation network analysis of scientific publications.

3. Data and methods

In previous works, citation-based approaches were used to describe the network of energy-related journals using journal citation data [42] or journal classification data [43]. Recently, researchers have explored the possibility of using citation clusters over three time periods to track the emergence and growth of research areas and predict their near-term changes [20,44]. In the citation-based approach, it is assumed that citing and cited papers have similar research topics. By clustering the citation network, we can detect a research front consisting of a group of papers. Clustering of the inter-citation network is also used to detect emerging research areas in energy research, especially focusing on fuel cells and solar cells [20]. In this paper, we also adopted clustering of the inter-citation network.

A schematic illustration of our analysis procedure is shown in Fig. 2. We collected citation data of publications on biomass and bio-fuels from the Science Citation Index (SCI) compiled by the Institute for Scientific Information (ISI). We used the Web of

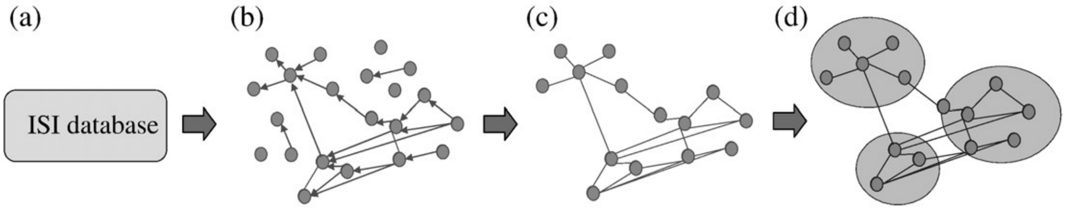

Fig. 2. Schematic illustration of the analysis process.

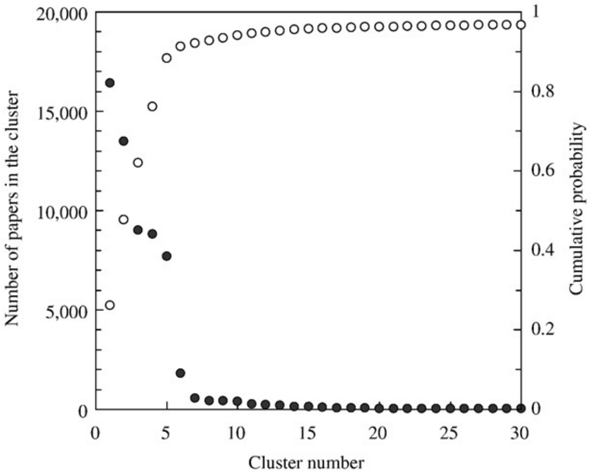

Fig. 3. Cluster size of the top 30 clusters. The black and open dots are the number of nodes in each cluster and the cumulative probability of the number of nodes, respectively.

Science, which is a Web-based user interface of the ISI's citation databases. Bibliographic records of 79,705 papers published in academic journals were collected by using the following query: "biomass* or bio-mass* or biofuel* or bio-fuel* or bioethanol* or bio-ethanol* or biodiesel* or bio-diesel* or biohydrogen* or bio-hydrogen*". By using the collected citation data, we constructed a network consisting of both connected components and isolated nodes as shown in Fig. 2(b). Then, the maximum connected component (Fig. 2(c)), which currently has 62,745 papers (78.7% of the retrieved papers), was extracted. The number of papers in the maximum connected component is less than that of the retrieved papers, because some papers have no citation to other papers in the component and thus omitted. The network was converted into a non-weighted, non-directed network. Finally, the network was divided into clusters using the topological clustering method [45,46]. The network was visualized by using a large graph layout (LGL) [47]. LGL is based on a spring layout algorithm where links play the role of spring connecting nodes. As a result of such layout, a group of papers citing each other is located in closer positions. In our visualization, we hide inter-cluster links and only show the intra-cluster links for each cluster with the same color to grasp the position of each cluster.

After clustering the network, we analyzed the characteristics of each cluster by the titles and abstracts of papers that are frequently cited by the other papers in the cluster as well as the journals in which the papers in the cluster were published. We named each cluster, and calculated the average publication year of the papers in each cluster to know the emerging research field. We also calculated the average number of times that papers in the cluster were cited by other papers. We simply call it "times cited (TC)." We use three ways of counting TC: TC by the other papers in the cluster (TC_c), TC by the papers in the

Table 1
List of journals where papers in the top 6 clusters are published

Cluster #1	# papers		Cluster #2	# papers		Cluster #3	# papers	
HYDROBIOLOGIA	1172	7.1%	SOIL BIOL B	1384	10.3%	BIOMASS BIOENERG	734	8.1%
MAR ECOL-PROGR SER	1005	6.1%	BIOL FERT SOILS	629	4.7%	J GEOPHYS R	734	8.1%
LIMNOL OCEANOGR	584	3.6%	FOREST ECOL MANAGE	607	4.5%	FUEL	329	3.6%
J PLANKTON RES	525	3.2%	PLANT SOIL	566	4.2%	ENERG FUEL	319	3.5%
FRESHW BIOL	487	3.0%	CAN J FOREST RES	396	2.9%	BIORESOUR TECHNOL	291	3.2%
DEEP-SEA RES PT II-TOP ST OCE	399	2.4%	SOIL SCI SOC AMER J	331	2.5%	APPL BIOCHEM BIOTECH	236	2.6%
CAN J FISHERIES AQUAT SCI	356	2.2%	GLOB CHANGE BIOL	311	2.3%	ATMOS ENV	234	2.6%
AQUAT BOT	347	2.1%	APPL SOIL ECOL	297	2.2%	IND ENG CHEM	195	2.2%
ARCH HYDROBIOL	346	2.1%	OECOLOGIA	255	1.9%	J ANAL APPL PYROL	162	1.8%
ESTUAR COAST SHELF S	338	2.1%	AGR ECOSYST ENVIRON	188	1.4%	ENERG SOURCE	159	1.8%
Cluster #4	# papers		Cluster #5	# papers		Cluster #6	# papers	
WATER SCI TECHNOL	696	7.9%	FIELD CROP RES	300	3.9%	CES J MAR SCI	143	7.9%
BIOTECHNOL BIOENG	634	7.2%	PLANT SOIL	234	3.0%	FISH RES	140	7.7%
WATER RES	547	6.2%	AGRON J	227	3.0%	MAR ECOL-PROGR SER	87	4.8%
APPL MICROBIOL BIOT	332	3.8%	WEED SCI	185	2.4%	CAN J FISHERIES AQUAT SCI	73	4.0%
PROCESS BIOCHEM	317	3.6%	OECOLOGIA	174	2.3%	FISH B	62	3.4%
BIORESOUR TECHNOL	271	3.1%	CROP SCI	158	2.1%	B MAR SCI	48	2.7%
ENZYME MICROB TECH	219	2.5%	FOREST ECOL MANAGE	150	1.9%	J APPL ECOL	34	1.9%
APPL ENVIRON MICROB	218	2.5%	AGROFOR SYST	150	1.9%	MAR FRESHW RES	28	1.5%
BIOTECHNOL LETT	201	2.3%	NEW PHYTOL	149	1.9%	BIOL CONSER	26	1.4%
J BIOTECHNOL	200	2.3%	CAN J FOREST RES	124	1.6%	AQUAT LIVING RESOUR	26	1.4%

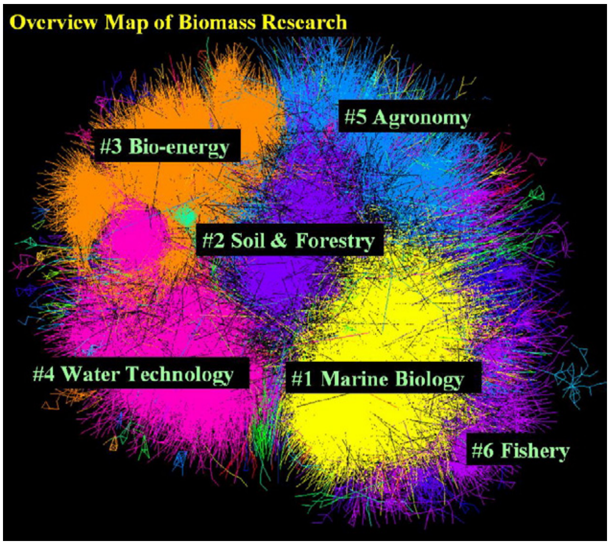

Fig. 4. Overview map of biomass research.

connected component (TC_{in}), and TC by all papers registered by the ISI (TC_{all}). In short TC_{all}, TC_{in}, and TC_c can be calculated on the network data in Fig. 2(a), (b), and (c), respectively. By comparing these ways of counting TC, we can know the possible influence of the research cluster on the other research domains. For example, when $TC_c \sim TC_{in}$, TC_{all}, the cluster is rather isolated and has little influence on and is little influenced by the other clusters. On the other hand, when $TC_c \ll TC_{in}$, TC_{all}, the research topics in the cluster are inter-disciplinary and close to those in other clusters. Innovation in such a cluster can also have a substantial impact on the other clusters.

4. Results and discussion

After clustering the citation network, we obtained 329 clusters, while most of the clusters consisted of only a small number of nodes and were peripheral. As shown in Fig. 3, more than 90% of the papers belong to the top 6 clusters. In order to characterize those clusters, we analyzed journals in which their papers are published and named the cluster.

Table 1 is a list of journals in which papers in each cluster are published. The largest cluster (Cluster #1) is dedicated to marine biology. As judged by the journal names, plankton and fisheries in oceans are the prior research targets. The second-largest cluster (Cluster #2) is soil and forestry. Cluster #3 is bio-energy, where biomass and bio-fuels, the main focus of these

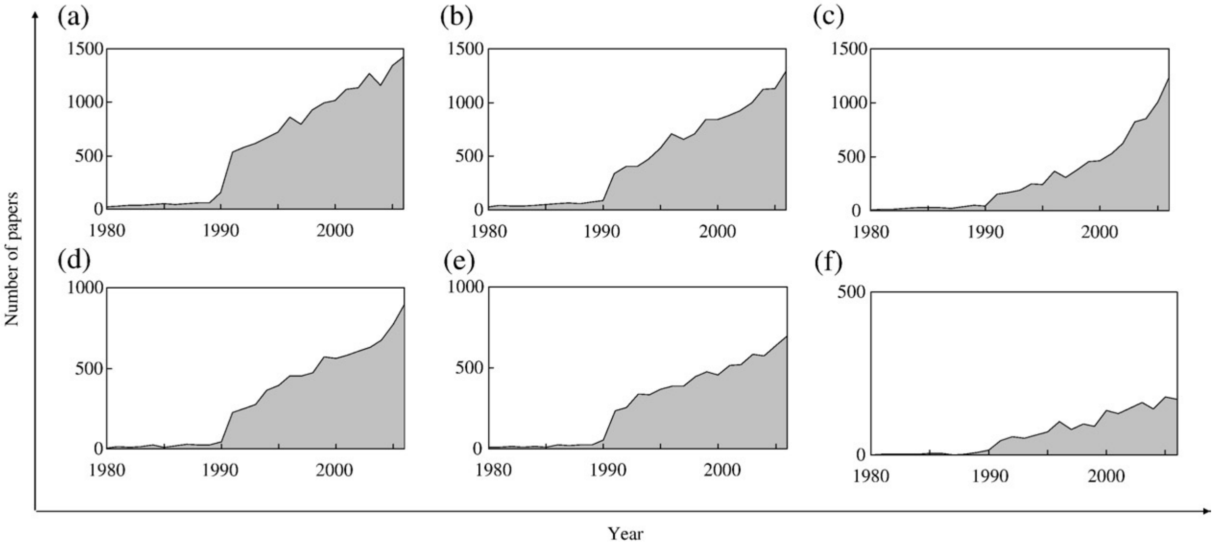

Fig. 5. Publication trend of biomass research. (a)–(f) correspond to Clusters #1–#6.

Table 2
Top 6 clusters in biomass research

Cluster ID	Cluster name	# papers	Year_{ave}	Tc_c	Tc_{in}	Tc_{all}
#1	Marine biology	16,401	1999.4	4.29	4.57	16.6
#2	Soil and forestry	13,496	1999.6	6.50	6.99	17.0
#3	Bio-energy	9015	2001.1	4.75	5.08	12.1
#4	Water technology	8821	2000.0	3.57	3.68	11.2
#5	Agronomy	7693	1999.6	2.41	2.86	11.0
#6	Fishery	1808	1999.9	1.81	2.07	12.4
Component		62,745	1999.8	4.20	4.53	13.9
Others		16,960	1997.2	N.A.	N.A.	5.13
Total		79,705	1999.2	N.A.	N.A.	12.0

papers, are discussed. Clusters #4, #5, and #6 are water technology, agronomy, and fishery cluster, respectively. The topological position of each cluster is visualized by using the LGL algorithm in Fig. 4. As shown in Fig. 4, Cluster #2 is at the central position because of its large size, and common research topics are probably discussed there. Some papers assigned to Cluster #4 are closely positioned to Cluster #3. This is because such papers in Cluster #4 share some citations with papers in Cluster #3.

Fig. 5 shows the number of papers when papers in each cluster were published annually. In all these clusters, the number of publications increased monotonically with some fluctuations. After 1990, the number of papers grew rapidly in these clusters. The number of papers during the 1990s in Cluster #3 is not so large compared to the other clusters, but the growth rate since 1990, especially after 2000, is the most rapid, which might reflect public concern and the resulting increasing funds for bio-fuels.

Table 2 summarizes the analyzed results. Currently, bio-energy (Cluster #3) is the most rapidly developing research domain among biomass researches, while the accumulated number of papers is still high in other domains. Soil and forestry research (Cluster #2) is located in the central position, which reflects a large number of publications and a large TC count. Cluster #6 has a relatively lower value for TC_c and TC_{in} compared to TC_{all} of the cluster. This means that papers in the cluster are strongly related with papers not included in this analysis and therefore they are noisy when we analyze bio-fuels. This is also supported by the peripheral position in the visualization (Fig. 4). On the other hand, soil and forestry (Cluster #2), water technology (Cluster #4), and agronomy (Cluster #5) seem to retain some relationship with bio-energy (Cluster #3) (Fig. 4). Most of the papers retrieved by the set of queries belong to the maximum connected components (62,745 papers/79,705 papers = 78.7%). Papers not included in the component do not have large fraction and their TC_{all} shows a remarkably smaller value than those of papers included in the component. Therefore, we assume that discarding papers from the component as digressional does not affect the analysis results. In the following, we focus on Cluster #3.

In order to analyze Cluster #3 in more detail, we performed the clustering of papers in Cluster #3 again and its obtained subclusters. The topological network of subclusters of Cluster #3 is visualized in Fig. 6. Clusters #31, #32, and #34 occupy the central positions in the network of papers belonging to Cluster #3, while Clusters #33 and #35 are peripheral. By successively clustering the obtained subclusters, we can depict a taxonomic structure of a research domain. In the following, we represent the cluster name after *n*th clustering as *n*-digit. Each digit corresponds to the rank of the cluster in size at each taxonomic level. For

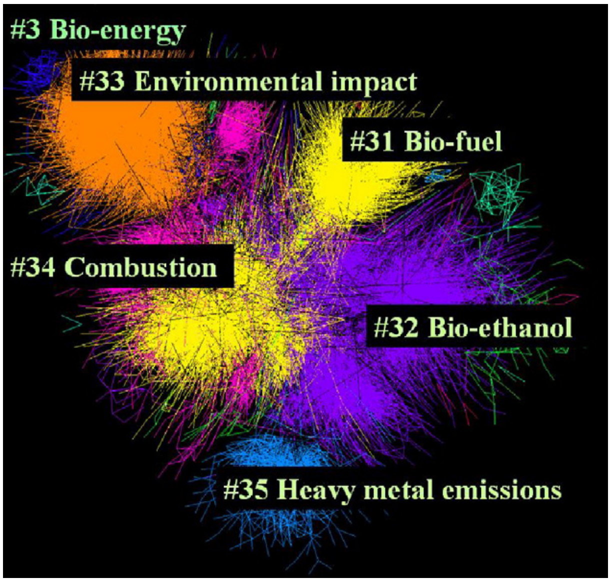

Fig. 6. Overview map of bio-energy research.

Table 3
2nd- and 3rd-level subclusters of the bio-energy cluster

Cluster ID	Cluster name	# papers	$Year_{ave}$	Tc_c	Tc_{in}	Tc_{all}
#31	Bio-fuel	2595	2002.1	4.90	5.14	8.72
#311	Pyrolysis	884	2000.9	5.07	5.61	10.2
#312	Bio-diesel	793	2003.5	3.56	3.67	6.73
#313	Gasification	596	2002.6	5.45	6.09	9.07
#314	Gasification	184	2002.2	4.38	5.20	11.4
#32	Bio-ethanol	2417	2000.0	3.53	3.78	9.33
#321	Lignocellulose	706	2000.4	4.57	5.04	12.7
#322	Short rotation forest	571	1999.2	3.58	4.04	8.68
#323	Energy crops	510	2000.1	2.64	2.91	5.91
#324	Future expectations	111	2004.0	1.48	2.23	6.32
#325	Fermentation mechanism	83	1999.9	1.93	2.34	10.3
#33	Environmental Impact	1932	2001.2	7.88	8.01	23.9
#34	Combustion	884	2002.1	3.16	3.78	6.85
#341	Co-combustion	215	2002.0	2.51	2.77	6.14
#342	Emissions	200	2002.8	4.16	4.94	8.59
#343	Indoor-air pollution	160	2001.2	3.28	3.33	9.81
#344	Fluidized bed combustion	109	2002.8	2.21	2.85	5.44
#35	Heavy metal emissions	358	2002.6	2.54	2.72	14.5
Total		9015	2001.1	4.75	5.08	12.1

example, Cluster #31 is the largest subcluster of Cluster #3, and Cluster #312 is the second-largest subcluster of Cluster #31 obtained after clustering the initial network three times. The analysis results are shown in Tables 3 and 4.

The 2nd level of Cluster #3 is shown in Table 3 accompanied by their 3rd-level clusters. The main subclusters of Cluster #3 are bio-fuel (#31), bio-ethanol (#32), environmental impact (#33), combustion (#34), and heavy metal emissions (#35). Two of them, i.e., #33 and #35, concern the negative impact of biomass and bio-energy on human health. They are characterized by a high value of TC_{all} compared to TC_c and TC_{in}. This means that papers in these two clusters are frequently cited by other than the biomass community. Clusters #31, #34, and #35 are young and currently under rapid development as evidenced by the small values of their $year_{ave}$. It is worth noting that while Cluster #33 is located in the peripheral position by the network visualization (Fig. 6), it has a large TC count especially TC_{all}. It seems to mean that we should keep the negative environmental impact of bio-fuels in mind when we develop the use of bio-fuels further.

Third-level subclusters of selected 2nd-level clusters are also shown in Table 3. These subclusters are chosen because they especially focus on bio-fuels. Pyrolysis (#311), bio-diesel (#312), and two gasification clusters (#313 and #314) are the main

Table 4
3rd- and 4th-level subclusters of the bio-energy cluster

Cluster ID	Cluster name	# papers	$Year_{ave}$	Tc_c	Tc_{in}	Tc_{all}
#311	Pyrolysis	884	2000.9	5.07	5.61	10.2
#3111	Flash pyrolysis	294	2000.6	4.46	5.62	10.1
#3112	Kinetic modeling	273	2001.1	4.53	5.91	10.8
#3113	Thermal effect	139	2000.5	2.12	3.99	10.7
#3114	Particle pyrolysis	94	2000.9	3.19	4.83	10.3
#312	Bio-diesel	793	2003.5	3.56	3.67	6.73
#3121	Engine performance	240	2004.1	2.47	3.00	4.85
#3122	Bio-diesel production	202	2004.9	3.65	4.80	6.93
#3123	Lipase-catalyzed synthesis	148	2002.3	3.45	4.12	9.40
#3124	Co-product	119	2002.1	1.65	2.79	7.77
#313	Gasification	596	2002.6	5.45	6.09	9.07
#3131	Elimination of tar	184	2001.7	7.58	9.69	12.8
#3132	Hydrogen production	170	2004.4	4.12	4.56	10.2
#3133	Different biomass materials	134	2002.1	2.09	3.40	5.28
#321	Lignocellulose	706	2000.4	4.57	5.04	12.7
#3211	Economic feasibility and bio-commodity engineering	205	2001.1	4.00	5.85	13.9
#3212	Metabolic engineering	202	2000.9	4.13	5.20	15.6
#3213	Pretreatment and hydrolysis	151	2000.0	2.38	3.66	10.8
#3214	Simultaneous saccharification and fermentation	94	1999.8	1.91	3.34	8.36
#322	Short rotation forest	571	1999.2	3.58	4.04	8.68
#3221	Economic and energy evaluation	168	1999.6	2.85	3.32	7.98
#3222	Poplar	134	1998.1	3.37	4.33	11.5
#3223	Willow	132	1998.2	2.50	3.73	6.87
#323	Energy crops	510	2000.1	2.64	2.91	5.91
#3231	Assessment	94	1999.6	1.77	2.10	5.11
#3232	Switchgrass	89	2002	3.85	4.29	7.12
#3233	Miscanthus	86	2000.7	2.94	3.48	7.99

subclusters of bio-fuel (#31). In the bio-ethanol (#32) cluster, different feedstocks for bio-ethanol are discussed, e.g., lignocellulose (#321), short rotation forest (#322), and energy crops (#323), while the fermentation mechanism (#325) during bio-ethanol processing is also discussed and some researchers discuss future prospects for bio-ethanol and bio-fuels in Cluster #324. Subclusters of combustion (#34) study efficient processing such as co-combustion (#341) of biomass with other fuels and fluidized bed combustion (#344). Other subclusters of combustion (#34), i.e., emissions (#342) during combustion and resulting indoor-air pollution (#343), have environmental concerns.

In Table 4, we show the 4th-level clusters. Several emerging clusters can be detected: kinetic modeling (#3112) of biomass feedstock pyrolysis; engine performance (#3121) of vehicles utilizing bio-diesel; bio-diesel production (#3122) of bio-diesel; hydrogen production (#3132) from biomass feedstock; economic feasibility and bio-commodity engineering (#3211) for lignocellulosic materials; metabolic engineering (#3212) to enhance the fermentation; economic and energy evaluation (#3221) of short rotation forest; and switchgrass (#3232) and miscanthus (#3233) as new energy crops. Among these emerging technological topics, the bio-diesel production (#3122) and hydrogen production (#3132) cluster are the youngest. In the bio-diesel production (#3122) cluster, new process technologies are studied. Examples include supercritical alcohol treatment, alkali and acid catalysis, and non-catalytic transesterification from various feedstocks. Hydrogen is produced from catalytic reforming of biomass-derived hydrocarbons, but no commercial-scale hydrogen process is yet in use [48,49] and building an infrastructure for the distribution of hydrogen is still an issue. We must bear in mind that the economic and energy feasibility of these emerging technologies is still under active discussion. In fact, $\mathrm{year}_{\mathrm{ave}}$ of Clusters #3211 and #3221 shows they are young, which means that these discussions are continuing. For example, Patzek and Pimentel evaluated gigantic tree plantations and concluded that about 500 million hectares (a little more than half of the area of the USA) of new plantations would be needed to replace 10% of the fossil energy used globally [37].

As described above, we analyzed the citation network of biomass and bio-fuels and revealed their global research structure. We clarified the taxonomic structure of bio-fuel research as shown in Tables 3 and 4. Next, we plotted $\mathrm{TC_c/TC_{in}}$ and $\mathrm{TC_{in}/TC_{all}}$ of the 3rd- and

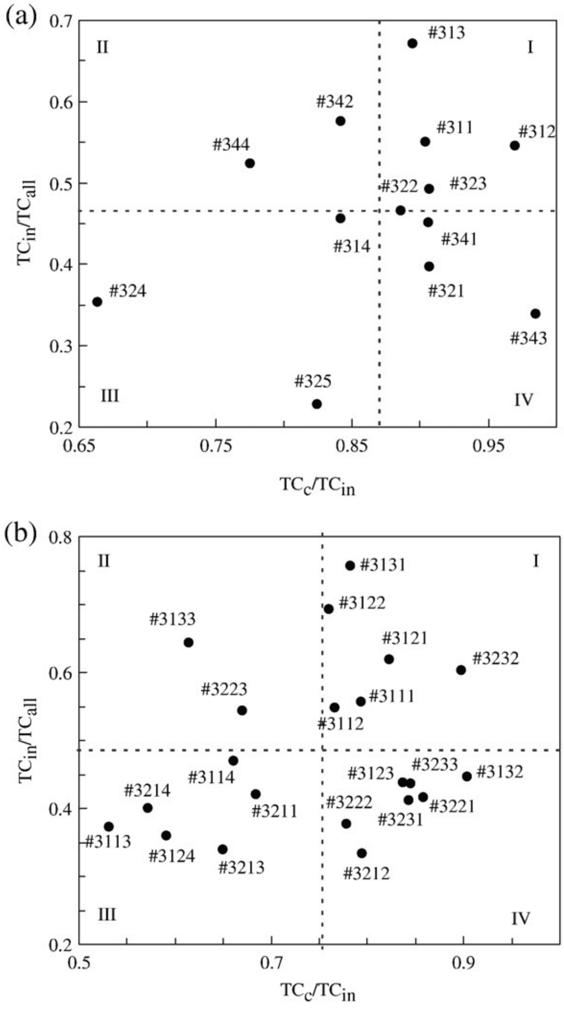

Fig. 7. TC count of (a) 3rd-level clusters and (b) 4th-level clusters. The dashed lines represent the average values for the TC counts. (I) Topic specific; (II) domain specific; (III) global link; (IV) specific & global.

Table 5
3rd- and 4th-level subclusters of the bio-energy cluster categorized by TC count

(I) Topic specific			
#311	Pyrolysis	#3111	Flash pyrolysis
#312	Bio-diesel	#3112	Kinetic modeling
#313	Gasification	#3121	Engine performance
#322	Short rotation forest	#3122	Bio-diesel production
#323	Energy crops	#3131	Elimination of tar
		#3232	Switchgrass
(II) Domain specific			
#342	Emissions	#3133	Different biomass materials
#344	Fluidized bed combustion	#3223	Willow
(III) Global link			
#321	Lignocellulose	#3113	Thermal effect
#341	Co-combustion	#3114	Particle pyrolysis
#343	Indoor-air pollution	#3124	Co-product
		#3211	Economic feasibility and bio-commodity engineering
		#3213	Pretreatment and hydrolysis
		#3214	Simultaneous saccharification and fermentation
(IV) Specific & global			
#314	Gasification	#3123	Lipase-catalyzed synthesis
#324	Future expectations	#3132	Hydrogen production
#325	Fermentation mechanism	#3212	Metabolic engineering
		#3221	Economic and energy evaluation
		#3222	Poplar
		#3231	Assessment
		#3233	Miscanthus

4th-level clusters in order to understand the relationships between these clusters and the others (Fig. 7). When TC_{in}/TC_{all} is small, it means that papers in the cluster are cited by papers other than biomass research, and therefore retain wide coverage of interest for other scientific domains. When TC_c/TC_{in} is small, it means that papers in the cluster are cited by papers in the other biomass clusters, not by those in that cluster, and therefore retain wide coverage of interest within biomass research, while large TC_c/TC_{in} means that the cluster deals with rather specific and narrow topics. Based on TC_{in}/TC_{all} and TC_c/TC_{in}, we categorized 3rd- and 4th-level clusters into the following four categories: (I) topic specific; (II) domain specific; (III) global link; (IV) specific & global (Fig. 7).

In Table 5, we show the list of clusters categorized in each category. For example, pyrolysis (#311) and its subclusters, flash pyrolysis (#3111), and kinetic modeling (#3112) as well as bio-diesel (#312) and its subclusters, engine performance (#3121), and bio-diesel production (#3122) are categorized into (I) topic specific. Citations of the papers in these clusters are closed within them. Therefore, researchers engaged in these researches might not have to notice the research trends of other researches. Some other clusters such as emissions (#342) and different biomass materials (#3133) are categorized into (II) domain specific, which is reasonable because these topics are common in biomass research.

The lignocellulose (#321) cluster and its subclusters, economic feasibility and bio-commodity engineering (#3211), pretreatment and hydrolysis (#3213), simultaneous saccharification and fermentation (#3214), and co-combustion (#341) are categorized into (IV) specific & global. Bio-ethanol, whose cluster ID is #32, can be produced by fermentation of sugarcane. Compared to energy crops, lignocellulosic raw materials minimize the potential conflict between land use for food (and feed) production and energy feedstock production. The raw material is less expensive than conventional agricultural feedstock and can be produced with lower input of fertilizers, pesticides, and energy. Plant lignocellulosic biomass is renewable, cheap, and globally available at 10–50 billion tons per year [50]. It is also advocated that 2nd-generation bio-ethanol from lignocellulosic feedstocks instead energy crops are indispensable to realizing sustainable bio-energy [39,50,51]. But to improve economic and energy feasibility, not only fermentation technology but also knowledge of other scientific disciplines such as chemicals, physics, and engineering is necessary, which leads to low TC_{in}/TC_{all} of these clusters. It is also valid for combustion technology, which needs knowledge of chemical engineering and kinetics. It might be worth noting that combustion technology should be simultaneously developed with fermentation technology because ethanol production from sugarcane is driven by burning cane leftovers, bagasse, and parts of attached cane tops and converting their combustion heat to steam, electricity, and shaft work.

The fact that lignocellulose (#321) is categorized into (IV) specific & global can be explained by the bio-refinery concept [52–54]. Just as from oil, gas, and coal, a range of chemical products can be co-produced from biomass. The concept of developing a bio-refinery to produce multi-products from a single feedstock similar to an oil refinery has promise. The production costs for bio-ethanol is on average around 250 US$ per ton, for ethylene, it is around 700 US$ per ton, and for acetic acid, it is around 650 US$ per ton [10]. Looking into the amounts and production costs of the respective chemicals, it seems realistic to produce, for example, acetic acid from bio-ethanol with the proper catalytic reaction pathway. In the USA, a joint venture between DuPont and Tate and Lyle is started to produce polymer for clothing, carpeting, and many other uses [54]. Therefore, topics discussed there are diverse and closely related with other scientific domains.

Our analysis showed that in the subcluster of lignocellulose (#321), metabolic engineering (#3212) is categorized into (IV) specific & global. This means that while papers in the cluster are closely related with other scientific domains, they also develop a dense network within the cluster. At present, biomass is converted to fermentable sugars for the production of bio-fuels using pretreatment processes that disrupt the lignocellulose and remove the lignin, thus allowing the access of microbial enzymes for cellulose deconstruction. Both the pretreatment and the production of enzymes in microbial tanks are expensive. Recent advances in plant genetic engineering could reduce biomass conversion costs by developing crop varieties with less lignin, crops that self-produce cellulase enzymes for cellulose degradation and ligninase enzymes for lignin degradation, or plants that have increased cellulose or an overall biomass yield. The metabolic engineering (#3212) cluster, whose year_{ave} is 2000.9, is rather young but seems to have become an independent research domain.

It is not plausible that a single technology can overcome the current limitation of bio-energy productions especially for clusters categorized into (III) global link and (IV) specific & global. Research categorized there should be developed by utilizing knowledge from other scientific domains and involving researchers and engineers from those domains.

5. Conclusion

Biomass and bio-fuels have gained growing interest as sustainable and renewable energy. They currently provide a significant amount of global consumer energy but mainly for traditional domestic cooking and heating in developing countries. Currently, there are both following and opposing winds regarding bio-fuels. The driving force includes surging global oil demand, increasing attention to the energy security problem caused by the geographical concentration of known petroleum reserves, and growing concern about atmospheric greenhouse gas concentrations. For the mass production and utilization of biomass as a renewable energy, the lack of energy and economic feasibility against other complementary energy sources as well as agricultural sectors such as food and textile production must be overcome by new and improved modern bio-energy technologies.

To realize this, R&D efforts to attain technological innovation are definitely required. Policy makers should allocate valuable and shrinking research funds to the development of bio-fuel as a feasible alternative to oil while comprehending the global structure of research. In this paper, we performed a citation network analysis of scientific publications to know the structure of biomass and bio-fuel research. We have mainly three objectives. These include the following;

i) to identify emerging technologies and research fronts
ii) to make recommendations on which science and technology areas to develop more rapidly, and
iii) to study the structure of the biomass and bio-fuels literature, along with the inter-thrust dynamics.

In order to achieve the first objective, we performed topological clustering of the citation network. By the hierarchical clustering, we can identify the emerging technologies and research fronts in a feasible manner. The following clusters were extracted as main clusters in biomass research: marine biology; soil and forestry; bio-energy; water technology; agronomy; and fishery. We revealed the taxonomic structure of biomass research especially focusing on bio-fuel and bio-energy research.

The second objective was solved by the publication years of papers as the clue, because emerging and rapidly developing domains can be characterized by the fast speed of publications. And as a result, the average publication year of papers in each cluster should be young. According to the analyzed results, bio-diesel and hydrogen production are the most rapidly developing domains.

The third objective, inter-thrust dynamics, was analyzed by using citation counts, i.e., times cited (TC) within and between clusters. This is based on the assumption that inter-disciplinary topics should have more inter-cluster citations than monodisciplinary topics. Based on the proposed ways of counting TC, we categorized research domains into the following four categories: (I) topic specific; (II) domain specific; (III) global link; and (IV) specific & global. It is not plausible that a single technology can overcome the current limitation of bio-energy production especially for clusters categorized into (III) global link and (IV) specific & global. Research on lignocellulose feedstock is a typical case where knowledge from other scientific disciplines is necessary. Research in categories (III) and (IV) should be developed by utilizing knowledge from other scientific domains and involving researchers and engineers from those domains.

Our methodology is based on clustering to identify technology. The disadvantage of such an approach is to miss the isolated paper. In other words, we disregard plausible promising technologies until the year when they bundle a couple of papers. For example, a radically new technology published in the most recent year might be completely isolated from a clustering perspective and might not have had sufficient time to accumulate citations, and therefore be neglected. Another shortcoming of citation-based approach is the existence of non-cited influential papers and the exclusion of some type of publications such as non-English papers, textbooks, conference proceedings, data sheets, and technical reports from the analysis. To overcome such a limitation, other techniques such as text-based approach or expert-based approach and careful re-examination of the corpus might be helpful. But our approach can work, at least, in a complementary method in recent knowledge-based but information-flood era.

Acknowledgment

This research was partially supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Young Scientists (B), 18700240, 2006.

References

[1] IEA, Biofuels for Transport: an International Perspective, International Energy Agency, Paris, 2004.
[2] IEA, Bioenergy Project Development Biomass Supply, International Energy Agency, Paris, 2007.
[3] D.O. Hall, Biomass energy in industrialized countries—a view of the future, For. Ecol. Manag. 91 (1997) 17–45.
[4] D. Morris, The next economy: from dead carbon to living carbon, J. Sci. Food Agric. 86 (2006) 1743–1746.

[5] A.A. Vertès, M. Inui, H. Yukawa, Implementing biofuels on a global scale, Nature Biotechnol. 24 (2006) 761–764.
[6] L. Petrus, M.A. Noordermeer, Biomass to biofuels, a chemical perspective, Green Chem. 8 (2006) 861–867.
[7] F. Ma, M.A. Hanna, Biodiesel production: a review, Bioresour. Technol. 70 (1999) 1–15.
[8] J.M. Marchetti, V.U. Miguel, A.F. Errazu, Possible methods for biodiesel production, Renew. Sustain. Energy Rev. 11 (2007) 1300–1311.
[9] A.C. Hansen, Q. Zhang, P.W.L. Lyne, Ethanol–diesel fuel blends—a review, Bioresour. Technol. 96 (2005) 277–285.
[10] J. Rass-Hansen, H. Falsig, B. Jørgensen, C.H. Christensen, Bioethanol: fuel or feedstock? J. Chem. Technol. Biotechnol. 82 (2007) 329–333.
[11] A. Demirbas, Progress and recent trends in biofuels, Pror. Energy Combust. Sci. 33 (2007) 1–18.
[12] J.R. Mielenz, Ethanol production from biomass: technology and commercialization status, Curr. Opin. Microbiol. 4 (2001) 324–329.
[13] R.H. Sims, A. Hastings, B. Schlamadinger, G. Taylor, P. Smith, Energy crops: current status and future prospects, Glob. Chang. Biol. 12 (2006) 2054–2076.
[14] J. Hill, Environmental costs and benefits of transportation biofuel production from food- and lignocellulose-based energy crops. A review, Agron. Sustain. Dev. 27 (2007) 1–12.
[15] R.N. Kostoff, R.R. Schaller, Science and technology roadmaps, IEEE Trans. Eng. Manage. 48 (2001) 132–143.
[16] R.N. Kostoff, H.J. Eberhart, D.R. Toothman, Hypersonic and supersonic flow roadmaps using bibliometrics and database tomography, J. Am. Soc. Inform. Sci. 50 (1999) 427–447.
[17] R.N. Kostoff, R. Tshiteya, K.M. Pfeil, J.A. Humenik, Electrochemical power text mining using bibliometrics and database tomography, J. Power Sources 110 (2002) 163–176.
[18] R.N. Kostoff, R. Tshiteya, K.M. Pfeil, J.A. Humenik, and G. Karypis: Science and technology text mining: electric power sources, DTIC Technical Report No. ADA421789, 2004. Springfield, VA: National Technical Information Service.
[19] R.N. Kostoff, R. Tshiteya, K.M. Pfeil, J.A. Humenik, G. Karypis, Power source roadmaps using bibliometrics and database tomography, Energy 30 (2005) 709–730.
[20] Y. Kajikawa, J. Yoshikawa, Y. Takeda, K. Matsushima, Tracking emerging technologies in energy research: Toward a roadmap for sustainable energy, Technol. Forecast. Soc. 75 (6) (2008) 771–782.
[21] S.M. Thomas, The evaluation of plant biomass research: a case study of the problems inherent in bibliometric indicators, Scientometrics 23 (1992) 149–167.
[22] G.P.J. Dijkema, P. Ferrão, P.M. Herder, M. Heitor, Trends and opportunities framing innovation for sustainability in the learning society, Technol. Forecast. Soc. Change 73 (2006) 215–227.
[23] S.E. Koonin, Getting serious about biofuels, Science 311 (2006) 435.
[24] S. Kim, B.E. Dale, Global potential bioethanol production from wasted crops and crop residues, Biomass Bioenergy 26 (2004) 361–375.
[25] A.J. Ragauskas, C.K. Williams, B.H. Davison, G. Britovsek, J. Cairney, C.A. Eckert, W.J. Frederick Jr., J.P. Hallett, D.J. Leak, C.L. Liotta, J.R. Mielenz, R. Murphy, R. Templer, T. Tschaplinski, The path forward for biofuels and biomaterials, Science 311 (2006) 484–489.
[26] A.E. Farrell, R.J. Plevin, B.T. Turner, A.D. Jones, M. O'Hare, D.M. Kammen, Ethanol can contribute to energy and environmental goals, Science 311 (2006) 506–508.
[27] D. Connor, I. Mínguez, Looking at biofuels and bioenergy, Science 312 (2006) 1743.
[28] T.H. Deluca, Looking at biofuels and bioenergy, Science 312 (2006) 1743–1744.
[29] C.J. Cleveland, C.A.S. Hall, R.A. Herendeen, Energy returns on ethanol production, Science 312 (2006) 1746.
[30] N. Hagens, R. Costanza, K. Mulder, Energy returns on ethanol production, Science 312 (2006) 1746.
[31] S. Ulgiati, A comprehensive energy and economic assessment of biofuels: when “green” is not enough, Crit. Rev. Plant Sci. 20 (2001) 71–106.
[32] T.W. Patzek, 'Thermodynamics of the corn-ethanol biofuel cycle, Crit. Rev. Plant Sci. 23 (2004) 519–567.
[33] C. Carter, W. Finley, F. James, D. Jackson, L. Willis, Palm oil markets and future supply, Eur. J. Lipid Sci. Technol. 109 (2007) 307–314.
[34] Y. Basiron, Palm oil production through sustainable plantations, Eur. J. Lipid Sci. Technol. 109 (2007) 289–295.
[35] L.D. Schmidt, P.J. Dauenhauer, Hybrid routes to biofuels, Nature 447 (2007) 914–915.
[36] M. Bender, Economic feasibility review for community-scale farmer cooperatives for biodiesel, Bioresour. Technol. 70 (1999) 81–87.
[37] T.W. Patzek, D. Pimentel, Thermodynamics of energy production from biomass, Crit. Rev. Plant Sci. 24 (2005) 327–364.
[38] M. Giampietro, S. Ulgiati, Integrated assessment of large-scale biofuel production, Crit. Rev. Plant Sci. 24 (2006) 365–384.
[39] B. Hahn-Hägerdal, M. Galbe, M.F. Gorwa-Grauslund, G. Lidén, G. Zacchi, Bio-ethanol — the fuel of tomorrow from the residues of today, Trends Biotechnol. 24 (2007) 549–556.
[40] R.H. Wiser, S.J. Pickle, Financing investments in renewable energy: the impacts of policy design, Renew. Sustain. Energy Rev. 2 (1998) 361–386.
[41] S. Herrera, Bonkers about biofuels, Nat. Biotechnol. 24 (2006) 755–760.
[42] R. Dalpé, F. Anderson, National priorities in academic research–strategic research and contracts in renewable energies, Res. Policy 24 (1995) 563–581.
[43] R.J.W. Tijssen, A quantitative assessment of interdisciplinary structures in science and technology: co-classification analysis of energy research, Res. Policy 21 (1992) 27–44.
[44] H. Small, Tracking and predicting growth areas in science, Scientometrics 68 (2006) 595–610.
[45] M.E.J. Newman, Fast algorithm for detecting community structure in networks, Phys. Rev. E 69 (2004) 066133.
[46] M.E.J. Newman, M. Girvan, Finding and evaluating community structure in networks, Phys. Rev. E 69 (2004) 026113.
[47] A.T. Adai, S.V. Date, S. Wieland, E.M. Marcotte, LGL: creating a map of protein function with an algorithm for visualizing very large biological networks, J. Mol. Biol. 340 (2004) 179–190.
[48] E. Chornet, S. Czernik, Renewable fuels — Harnessing hydrogen, Nature 418 (2002) 928–929.
[49] S.S. Penner, Steps toward the hydrogen economy, Energy 31 (2006) 33–43.
[50] M. Sticklen, Plant genetic engineering to improve biomass characteristics for biofuels, Curr. Opin. Biotechnol. 17 (2006) 315–319.
[51] C. Schubert, Can biofuels finally take center stage? Nat. Biotechnol. 24 (2006) 777–784.
[52] B. Kamm, M. Kamm, Principles of biorefineries, Appl. Microbiol. Biotechnol. 64 (2004) 137–145.
[53] T. Bridgwater, Biomass for energy, J. Sci. Food Agric. 86 (2006) 1755–1768.
[54] J.R. Mielenz, Bioenergy for ethanol and beyond, Curr. Opin. Biotechnol. 17 (2006) 303–304.

Yuya Kajikawa is an assistant professor at the Institute of Engineering Innovation at the School of Engineering at the University of Tokyo. His research interests include technology management, knowledge management, and structuring knowledge. He has a PhD in chemical system engineering from the University of Tokyo.

Yoshiyuki Takeda is an assistant professor at the Institute of Engineering Innovation at the School of Engineering at the University of Tokyo. His research interests include information retrieval, natural language processing, and network analysis. He has a PhD in computer science from the Toyohashi University of Technology.