

**PEMAKAIAN KURVA ZIPF DAN ANALISIS
BIBLIOMETRIKA PADA PENENTUAN SEKTOR-SEKTOR
TEKNOLOGI *RENEWABLE ENERGY* YANG MENARIK
BERDASARKAN INFORMASI PATEN**

Jadi Suprijadi
Departemen Statistika FMIPA Unpad.
jadisuprijadi@gmail.com

Abstrak

Riset terapan dalam bidang *renewable energy* senantiasa dibutuhkan untuk menekan porsi pemanfaatan energi fosil. Pada umumnya riset terapan terpublikasikan dalam bentuk paten. Dalam konteks ini ada kebutuhan menganalisis data paten dengan tujuan untuk menseleksi sektor-sektor teknologi yang menarik untuk dijadikan landasan orientasi riset selanjutnya. Data paten tersebut merupakan data teks yang tersedia dalam pangkalan data. Teridentifikasi masalah yaitu pertama, mentransformasikan data dari format teks kepada data numerik, dan kedua, pengelompokan yang digunakan sebagai landasan penseleksian informasi yang menarik. Untuk menjawab permasalahan tersebut, dalam paper ini dibahas metode pemakaian kurva Zipf untuk menseleksi kelompok sektor teknologi yang menarik dan analisis bibliometrika untuk mentransformasikan dari data teks kepada data numerik. Dalam hal ini, data harus berformat bibliografik dan mengandung *field* IPC (*International Patent Classification*) yang berbentuk kode-kode klasifikasi yang merepresentasikan sektor-sektor teknologi. Data paten dikumpulkan dari pangkalan data international Espacenet. Diperoleh hasil dari 2349 dokumen paten teridentifikasi sebanyak 202 sektor teknologi dan tiga kelompok sektor teknologi yaitu kelompok dengan kategori *trivial*, *interesting*, dan *noise*.

Kata Kunci: bibliometrika; kurva Zipf; *renewable energy*; analisis paten

1. PENDAHULUAN

Indonesia merupakan negara dengan tingkat ketergantungan terhadap energi fosil yang masih tinggi. Upaya pemerintah untuk tujuan menekan porsi pemanfaatan energi fosil dari tahun ke tahun adalah melalui penetapan berbagai peraturan pemerintah dan undang-undang. Misalnya Peraturan Pemerintah (PP) No. 79 tahun 2014, menargetkan penggunaan energi baru atau terbarukan paling sedikit 23% pada tahun 2025 dan paling sedikit 31% pada tahun 2050. Hal tersebut diimbangi oleh target penurunan energi fosil terutama minyak sebesar kurang dari 25% pada 2025 dan kurang dari 20% pada tahun 2050. Kebijakan untuk mencapai target tersebut secara umum adalah keseimbangan penggunaan energi, serta peningkatan penggunaan energi baru atau terbarukan.

Untuk mengorientasikan riset-riset terapan dalam perspektif mewujudkan ketahanan energi dengan memaksimalkan pemanfaatan sumber-sumber energi alternatif yang terbarukan, dibutuhkan informasi mengenai sektor-sektor teknologi yang menjadi area-area riset yang telah dilakukan oleh peneliti-peneliti di berbagai negara. Informasi tersebut dapat diperoleh melalui paten-paten. Menurut Quoniam, informasi paten adalah alat yang

sangat baik untuk mempelajari hal-hal menyangkut pada transfer teknologi, inovasi, dan kreativitas (Quoniam, 2014).

Sumber informasi yang relevan adalah pangkalan data paten international. Saat ini tersedia pangkalan data paten international yang dapat diakses melalui Internet. Penelitian ini dilakukan memanfaatkan sumber informasi ini dengan tujuan untuk mengidentifikasi sektor-sektor teknologi dari sejumlah besar paten dan memperoleh kelompok sektor teknologi yang termasuk kedalam kategori sektor teknologi yang menarik. Sektor teknologi tersebut diterjemahkan dari kode-kode pengklasifikasian IPC (*International Patent Classification*) yang merepresentasikan kandungan teknik dalam suatu dokumen paten. Teridentifikasi masalah pertama, bagaimana mentransformasikan informasi teks menjadi data kuantitatif, karena data paten yang diolah adalah data tekstual. Kedua, mengklasifikasikan sektor-sektor teknologi yang sudah teridentifikasi sehingga diperoleh sekelompok sektor teknologi yang termasuk dalam kategori menarik.

Dalam penelitian ini data paten diunduh dari pangkalan data international paten, berupa informasi teks yang semi-terstruktur dalam format bibliografik. Elemen-elemen bibliografik tersebut misalnya *title*, *inventor(s)*, *applicant(s)*, *International Classification*, *priority numbers*, dan sebagainya. Sehingga metode analisis data yang tepat untuk menjawab tujuan penelitian adalah metode analisis bibliometrika. Bibliometrika adalah instrumen pengukuran dan analisis berdasarkan pemakaian teknik-teknik statistik yang bertujuan untuk membantu dalam membandingkan dan memahami suatu set besar elemen-elemen bibliografik melalui berbagai korelasi (Rostaing, 1996). Penulis ini mendefinisikan Bibliometrika sebagai aplikasi dari metode-metode statistik atau matematik pada sehimpunan referensi bibliografik. Penggunaan bibliometrika pada analisis paten berkembang dengan pesat, Dou mengkreasi perangkat lunak Matheopaten berdasarkan pada teknik pengolahan bibliometrika yang spesifik untuk analisis paten berdasarkan pada pangkalan data Espacenet dan USPTO (Dou, 2004), yang banyak diaplikasikan oleh para peneliti, salah satu diantaranya dalam penelitian Gorin (Gorin, 2013). Secara umum analisis bibliometrika pada data bibliografik dapat direalisasikan menggunakan perangkat lunak Dataview, misalnya penelitian yang dilakukan oleh Tarapanoff yang menganalisis pangkalan data bibliografik tentang disertasi di Perancis pada subjek Brazil. Informasi tentang topik apa yang paling banyak diteliti, dosen pembimbing, dan bidang ilmu diekstrak melalui teknik bibliometrika (Tarapanoff, et al, 2001).

Teknik-teknik bibliometrika meliputi, pertama, elaborasi indikator-indikator univariat, dimana ukuran kuantitatif didasarkan pada penghitungan sederhana dari suatu elemen bibliografik, kedua, elaborasi indikator-indikator relasional, yaitu pengeksplotasian metode-metode analisis data untuk menjelaskan relasi-relasi antar berbagai referensi bibliografik, misalnya analisis *co-citations*, asosiasi kata-kata, *co-classifications*, dan tabel kontingensi (Rostaing, 1996). Dalam penelitian ini penulis menerapkan teknik bibliometrika yang pertama. Teknik lainnya merupakan pemodelan distribusi elemen-elemen bibliometrika yang dalam penelitian ini penulis

mengaplikasikan kurva Zipf untuk pengelompokan unit-unit dari suatu elemen bibliografik. Kurva Zipf tersebut dibentuk berlandaskan kepada frekuensi dari kemunculan kata-kata dalam teks. Menurut Quoniam (dalam Tarapanoff et al, 2001) pada kurva Zipf kita dapat memiliki tiga *zone* yaitu *zone I*, *zone II* dan *zone III* yang masing-masing berisi informasi *trivial*, informasi menarik dan *noise*.

2. METODE PENELITIAN

a. Data

Data paten relatif dalam bidang energi terbarukan diunduh dari database Espacenet dari *European Patent Organization* (EPO) melalui *www.espacenet.com*. Tanpa bantuan perangkat lunak, pangkalan data ini memungkinkan peneliti untuk mengunduh maksimum sebanyak 500 dokumen paten saja, padahal ketersediaan dokumen paten terkait kepada subjek penelitian ini jauh lebih banyak jumlahnya. Oleh karena itu penulis memanfaatkan perangkat lunak *Matheo Paten XE*, dengan menggunakan kata kunci *renewable and energy* hadir dalam judul paten maka terunduh sebanyak 2349 bibliografik paten. Salah satu contohnya adalah sebagai berikut:

-1-
 TI_EN – SUPPLY UNIT FOR POWER AND WATER BASED ON RENEWABLE ENERGY
 PN – AP1995A
 PD – 04/13/2009
 PA – NIEDERER ROBERT (CH); HAUTLE URS (CH)
 IN - NIEDERER ROBERT (CH); HAUTLE URS (CH)
 AN - AP200503442A 20040406
 PR - CH6352003A 20030407; CH2004000215W 20040406
 WIPO - Mechanical engineering/Engines, pumps, turbines;
 Mechanical engineering/Thermal processes and apparatus;
 Electrical engineering/Semiconductors
 CPC - F24J2/542; F03D9/007; F05B2220/61; F05B2220/708; F05B2240/94; F24J2002/5479;
 H01L31/045; H01L31/0583; H01L31/0586; Y02B10/20; Y02B10/30; Y02B10/70;
 Y02E10/465; Y02E10/47; Y02E10/50; Y02E10/72; Y02E60/366; Y02E70/10; Y02E70/20;
 Y10S136/291
 CPC4 - F24J; F03D; F05B; H01L; Y02B; Y02E; Y10S
 CPC7 - F24J2; F03D9; F05B2220; F05B2240; F24J2002; H01L31; Y02B10; Y02E10;
 Y02E60; Y02E70; Y10S136
 IC - F03D9/00; F24J2/54; H01L31/00; H01L31/045; H01L31/058
 I4 – F03D; F24J; H01L
 I7 - F03D9; F24J2; H01L31
 AB_EN - (US7888590B2) The supply unit comprises a box type profile frame (1), whose base rests on wheels (2) and on whose top side (3), a square frame (16), as seen from above, is placed, said frame containing a solar panel (7). This can be swiveled about a horizontal axis (8) and is so connected to one of the top sides of the box type profile frame (1). A peripheral, square frame (5) of same size, each containing a solar panel (6), is connected in a swiveling way to each side of this square frame (16), so that out of the five square frames (16;5), a cube is formed when these frames are swiveled down. The peripherally connected square frames (5) can be swiveled to the plane of the central square frame (16) and can be locked in this position to the central square frame (16). The box type profile frame (1) contains several box

type modules (24-26), which can be inserted like drawers from one side and can be arrested in these inserted positions; these modules are for various functions like accumulation of energy from sun and wind, preparation of drinking water, pumping water, delivery of electric power or Direct Current-Hydrogen generation through fuel cells.

UR -

http://worldwide.espacenet.com/publicationDetails/biblio?CC=AP&NR=1995A&KC=A&FT=D&DB=EPODOC&locale=en_EP

Gambar.1 Contoh salah satu data paten

b. Pemilihan Variabel Penelitian

Mengacu kepada tujuan penelitian, maka pemilihan variabel difokuskan kepada elemen bibliografik yang berisi informasi mengenai kandungan teknik dalam suatu dokumen paten. Dalam dokumen-dokumen paten, *technical content* diklasifikasikan menurut IPC (*International Patent Classification*). Dengan demikian dalam data, *field* yang relevan untuk sektor teknologi adalah *field* yang berisi kode-kode IPC yang direpresentasikan oleh dua buah *field* yaitu "IC" yang berisi kode-kode IPC terdiri atas tujuh digit dan "I4" berisi kode-kode IPC dalam empat digit.

Contoh kode IPC dalam empat digit dan maknanya :

A01B *SOIL WORKING IN AGRICULTURE OR FORESTRY; PARTS, DETAILS, OR ACCESSORIES OF AGRICULTURAL MACHINES OR IMPLEMENTS, IN GENERAL(making or covering furrows or holes for sowing, planting or manuringA01C 5/00; machines for harvesting root crops A01D; mowers convertible to soil working apparatus or capable of soil workingA01D 42/04; mowers combined with soil working implements A01D 43/12; soil working for engineering purposes E01, E02, E21)*

Penulis memilih *field* "I4" sebagai variabel penelitian dengan alasan lebih simpel, maknanya secara umum sudah mencerminkan sektor-sektor teknologi (Jakobiak, 1994).

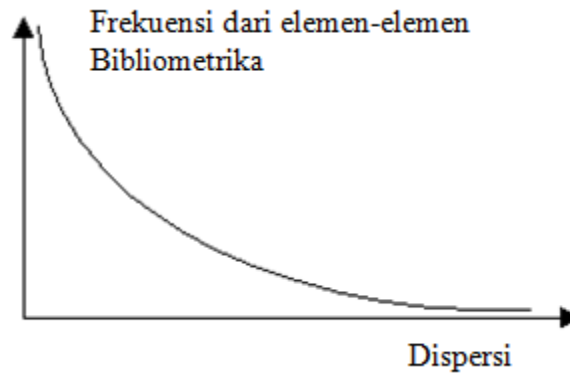
c. Analisis Data

Data paten yang akan dianalisis seperti yang sudah dituliskan di atas merupakan data teks dalam format data bibliografik. Langkah pertama yang dilakukan adalah persiapan data yaitu menghapus sebagian *field* yang tidak diperlukan. Pada analisis data atau pengekstraksian *forms* (yaitu kode-kode IPC dalam empat digit) yang merupakan unit-unit dari elemen bibliografik "I4" penulis menggunakan metode analisis bibliometrika meliputi teknik elaborasi indikator-indikator univariat dengan mengaplikasikan perangkat lunak Dataview. Sedangkan hasil dari pengolahan bibliometrika ini terdapat beberapa opsi (Rostaing, 2000), disesuaikan dengan tujuan penelitian, maka penulis hanya memilih hasil pengolahan meliputi *lists of forms* dan *models of distribution of forms*.

d. Kurva Zipf

Distribusi frekuensi *form* dimana dispersi dari frekuensi-frekuensinya diurutkan dari *form* dengan frekuensi tertinggi hingga *form* dengan frekuensi

terkecil selanjutnya direpresentasikan dalam grafik. Grafik tersebut mengikuti pola kurva *Zipf* yang diperlihatkan pada Gambar.2.



Gambar.2 Kurva Zipf (Tarapanoff, 2001)

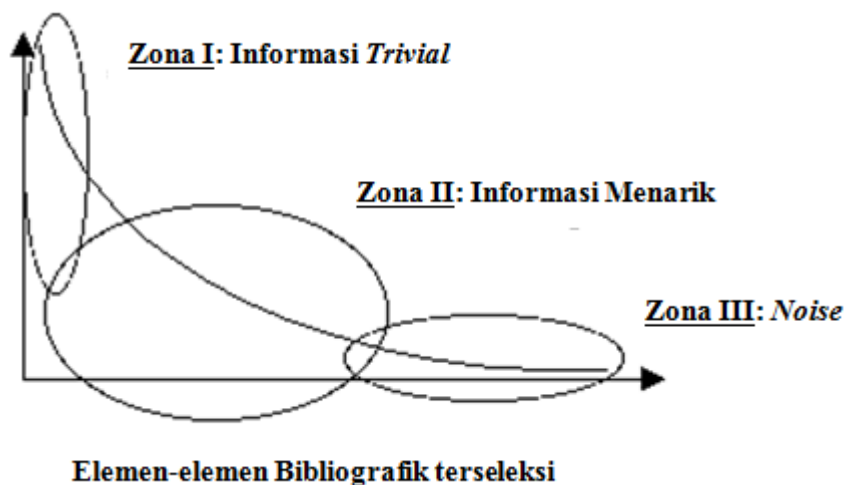
Menurut Quoniam (dalam Tarapanoff et al, 2001), berdasarkan pada kurva Zipf kita dapat memiliki tiga kelompok informasi :

Zone I - Informasi trivial : mendefinisikan tema sentral dari analisis bibliometric.

Zone II - Informasi menarik: ditemukan diantara *zone I* dan *zone II* yang menunjukkan informasi inovatif yang potensial menampilkan topik yang frekuensinya lebih sedikit tetapi sangat berpotensi mengandung informasi yang inovatif. Dalam hal ini transfer teknologi yang berkaitan dengan ide-ide baru harus dipertimbangkan.

Zone III – Noise : mengandung konsep-konsep yang belum muncul dan tidak mungkin dikatakan bahwa konsep-konsep itu akan muncul atau akan tetap merupakan gangguan statistis.

Zone I, II dan III direpresentasikan dalam kurva Zipf diperlihatkan dalam Gambar.3.



Gambar.3 *Zone* distribusi

Dalam memilah *zone* pada kurva Zipf, tidak ada cara yang eksak untuk menentukan pembatas antara *zone* I dan *zone* II maupun antara *zone* II dan *zone* III. Pembatas-pembatas tersebut ditentukan berdasarkan pada pertimbangan peneliti sendiri.

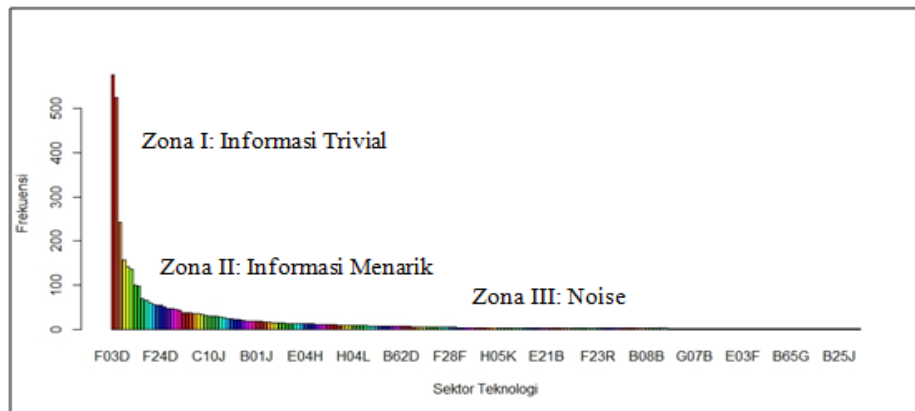
3. HASIL PENELITIAN DAN PEMBAHASAN

Berdasarkan hasil pengolahan bibliometrika pada *field* "I4", pada luaran *models of distribution of forms*, pengekstraksian *form* berhasil mengidentifikasi terdapat 202 *form* yang berbeda. Selanjutnya luaran *lists of forms* disajikan pada Tabel.1. Berdasarkan tabel ini diperoleh diagram (kurva Zipf) yang diperlihatkan pada Gambar.4.

Tabel.1 Distribusi frekuensi dari *forms*

<i>Frequency</i>	<i>Occurrence</i>	<i>Hapax</i>	<i>Form</i>
576	576	211	F03D
525	525	244	H02J
243	243	100	F03B
156	156	26	F24J
141	141	77	G06Q
136	136	36	F03G
99	99	48	G06F
98	98	19	H01L
70	70	21	H02M
66	66	22	C02F
60	60	21	H01M
56	56	18	C10L
54	54	18	F25B
54	54	16	F24D
50	50	27	G05F
47	47	11	H02N
47	47	21	C12M
44	44	7	F01K
42	42	15	H02K
38	38	4	F04B
⋮	⋮	⋮	⋮
1	1	0	A23K

Dari 2349 dokumen paten yang dianalisis, teridentifikasi terdapat 202 sektor teknologi. Sektor teknologi yang paling banyak diaplikasikan pada paten-paten relatif dalam bidang *renewable energy* adalah F03D (*Wind Motors*) dengan frekuensi sebanyak 576 diikuti oleh H02J (*Circuit Arrangements or Systems for Supplying or Distributing Electric Power; Systems for Storing Electric Energy*) dengan frekuensi sebanyak 525 dan F03B (*Machines or Engines for Liquids*) dengan frekuensi sebanyak 243.



Gambar.4 Kurva Zipf sektor-sektor teknologi.

Selanjutnya berdasarkan kurva Zipf dan pertimbangan peneliti diperoleh tiga kelompok informasi sektor-sektor teknologi yang masing-masing tercakup dalam *zone 1*, *zone 2* yang ditampilkan pada Tabel. 2.

Tabel.2. *Zone I dan Zone II* Sektor Teknologi

Sektor Teknologi	Sektor Teknologi	Sektor Teknologi
<i>Zone I: Informasi Trivial</i>		
F03D	H02J	F03B
<i>Zone II: Informasi Menarik</i>		
F24J	C02F	H02N
G06Q	H01M	C12M
F03G	C10L	F01K
G06F	F25B	H02K
H01L	F24D	
H02M	G05F	

Sektor-sektor teknologi yang tidak tercakup dalam *zone I* dan *zone II* dikategorikan ke dalam *zone III* atau *noise*.

4. SIMPULAN

Berdasarkan analisis terhadap 2349 dokumen paten relatif dalam bidang *renewable energy* terdapat 202 area teknologi dimana penemuan-penemuan baru dipatenkan oleh para peneliti. Sektor-sektor teknologi yang dikategorikan menarik sebagaimana tercantum pada *zone II* dalam Tabel.2 dipertimbangkan sebagai landasan area-area teknologi untuk mengembangkan ide-ide baru, mengorientasikan program riset dalam bidang *renewable energy* maupun transfer teknologi. Namun demikian penetapan kelompok sektor teknologi yang menarik ini harus dikomunikasikan atau perlu kehadiran pakar di bidang energi.

5. DAFTAR PUSTAKA

- Dou Henri, J. M. (2004). Benchmarking R&D and companies through patent analysis using free databases and special software: a tool to improve innovative thinking. *World Patent Information* 26, 297–309.
- Goarin, C. D. (2013). Patent analysis, detection of new markets foremployment. HAL Id: hal-00913373 <https://hal.archives-ouvertes.fr/hal-00913373>
- Jakobiak, F. (1994). *Le brevet source d'information*. Paris, Dunod.
- Quoniam, L., Baaziz, A. (2014). Patents used by NPE as an Open Information System in Web 2.0 – Two mini case studie. *Journal of Intelligence Studies in Business* Vol 4, No 2 ,51-60.
- Rostaing, H. (2000). *Guide d'utilisation de Dataview: logiciel bibliométrique d'aide à l'élaboration d'indicateurs de tendances*. Marseille: CRRM.
- Rostaing, H. (1996). *La bibliométrie et ses techniques*. Toulouse, Sciences de la Société.
- Tarapanoff, K. (2001). Intelligence obtained by applying data mining to a database of French theses on the subject of Brazil. *Information Research*, 7(1). Available at: <http://InformationR.net/ir/7-1/paper117.html>