

# **Penggunaan Istilah Teknologi Maklumat dan Komunikasi: Suatu Kajian Berdasarkan Teks Akhbar Harian \***

*Rusli Abdul Ghani* ( rusli@dbp.gov.my )  
*Hazimah Yusof* ( hazimah@dbp.gov.my )

Bahagian Penyelidikan Bahasa  
Dewan Bahasa dan Pustaka ( <http://www.dbp.gov.my> )

## **Abstrak**

Kajian ini meneliti penggunaan istilah teknologi maklumat dan komunikasi (TMK) dalam akhbar berbahasa Melayu yang terdapat di World Wide Web dan mengaitkannya dengan usaha pembentukan istilah TMK oleh Dewan Bahasa dan Pustaka.

Dari segi metodologi, kami memanfaatkan perisian analisis teks untuk memproses sebuah korpus akhbar yang terdiri daripada makalah-makalah TMK yang tersiar dalam ruangan Berita IT (Berita Harian) dan ruangan Megabait (Utusan Malaysia). Pemprosesan ini menghasilkan suatu senarai kata, dan daripada senarai ini kami mengasingkan bentuk-bentuk istilah TMK yang digunakan dalam akhbar.

Kami juga meneliti korpus teks makalah TMK berbahasa Inggeris yang tersiar pada laman Web Wired News (<<http://www.wired.com/news/technology/>>), Zdnet (<<http://www.zdnet.com/zdnn/>>) dan CNET (<<http://news.cnet.com/news/>>) untuk mengenal pasti istilah-istilah mutakhir dalam bidang TMK.

Matlamat utama kajian ini adalah untuk mendapatkan senarai istilah TMK yang digunakan dalam akhbar dan kemudian membandingkan senarai tersebut dengan istilah TMK terbitan DBP. Hasil perbandingan ini akan menjawab persoalan sama ada tata kerja pembentukan istilah yang diamalkan sekarang masih memadai dan mampu untuk menangani bidang TMK yang pesat berkembang, atau sebaliknya.

## PROLOG

"... *Kawai'ae'a, Donaghy, and NeSmith managed to get computers that spewed Hawaiian throughout the classrooms of the Kula Kaiapuni. But once the computers are there, what on earth do you call them? Do you transliterate (as Bingham did to settle on Kristo for Christ), and call the computer kamepiula? Or do you make up a term from components that already exist in Hawaiian - say lolouila, or "electric brain"?*

*As the curriculum of the Kula Kaiapuni expanded, teachers found themselves needing terms unimaginable a century ago, when the language had, in effect, become frozen in time. (What is the Hawaiian word for an adventure story? How do you say "floppy disk"?) Before long, Kawai'ae'a and NeSmith turned to a lexicon committee - comprised of professors and native speakers - that now gathers seven times a year to grow the language. Searching for a Hawaiian term for "upload," the committee settled on ho'ouka, "to load or put up on." Ho'ouka describes how you would put a saddle on a horse. Ho'ouka also gets a suitcase into a trunk. Now it also gets a file onto the Net. In other cases, the committee members resorted to Bingham's transliteration methods: they took an English term and made it Hawaiian sounding. "Line" became laina. "Telephone" became kelepona. "Beep" became pipa ..."* (Hale 1995)

## 1.0 PENDAHULUAN

Terminografi, berdasarkan pengalaman DBP selama lebih dua dekad bergiat dalam ranah ini, boleh disimpulkan sebagai berkisar pada metodologi umum pengumpulan (lazimnya istilah sumber bahasa Inggeris dalam bidang atau sub-bidang tertentu yang menjadi pangsi rujukan untuk pepadanan dengan istilah sasaran), pembentukan atau pepadanan (oleh Jawatankuasa Istilah dengan berpandukan pedoman umum<sup>1</sup> dan pedoman khusus<sup>2</sup>) dan penyebaran istilah (dalam bentuk bahan bercetak seperti senarai, buku, glosari dan kamus istilah dan dalam bentuk helaian Web<sup>3</sup> dan PDF<sup>4</sup> pada laman DBP dan laman Mabbim).

Senario sebenar pembentukan istilah bahasa Melayu tidaklah jauh berbeza daripada apa yang dialami sesiapa sahaja yang bergiat dalam terminografi, sebagaimana pengalaman mereka yang membina bahasa Hawaii dalam Prolog di atas.

Kertas kerja ini tidak bermaksud untuk memberikan huraian tuntas tentang langkah-langkah dalam pembentukan istilah (untuk itu, lihat Felber 1984), tetapi lebih

---

<sup>1</sup> Sebagai contoh, *Pedoman Umum Pembentukan Istilah Bahasa Melayu* (JKTBM 1992).

<sup>2</sup> Misalnya, 'Pedoman Pembentukan Istilah Perubatan' (DBP 1987:xix-lvii).

<sup>3</sup> <<http://dbp.gov.my/glosari.htm>>

<sup>4</sup> Adobe Portable Document Format.

menjurus pada satu aspek khusus pemprosesan istilah sahaja (pemprosesan yang lebih menyeluruh diuraikan dalam Sager 1990).

Tumpuan kami adalah terhadap aspek pengumpulan istilah dengan meneliti penggunaan sebenar istilah dalam teks berbahasa Melayu dan teks berbahasa Inggeris, dengan hasrat untuk menganggarkan keberkesanan dan kepraktisan metodologi penggubalan istilah yang ada sekarang dalam menangani bidang ilmu yang pesat berkembang seperti yang berlaku dalam dunia teknologi maklumat dan komunikasi (ringkasnya TMK). Oleh sebab kepesatan ini jugalah maka kami memfokuskan penelitian terhadap teks akhbar harian.

Berdasarkan penelitian ini, kami akan mencadangkan suatu kaedah umum yang dapat membantu mempercepat proses pembentukan istilah, khususnya pada peringkat pengumpulan istilah.

## **2.0 ASAS TEORETIS DAN MOTIVASI**

Korpus semakin mendapat tempat dalam semua bidang penelitian yang berkait dengan pemprosesan bahasa tabii (Boguraev dan Briscoe 1989; Butler 1992; Wilks et al. 1996; Zaharin Yusoff 1996). Hakikat ini juga seharusnya benar bagi perkamusan dan peristilahan.

Kegiatan perkamusan telah mencapai kejayaan yang memberangsangkan apabila ahli leksikografi mula mengandalkan korpus dalam kerja penyusunan kamus masing-masing (Atkins 1994; Sinclair 1987; Sinclair 1991). Kaedah ini turut diamalkan di DBP untuk penyusunan kamus dengan adanya pangkalan data korpus bahasa Melayu (sebesar sekitar 74 juta perkataan setakat ini).

Namun, alam terminografi Melayu ternyata sepi dan tidak sedikit pun teruja dengan penyelidikan linguistik komputeran yang berasaskan korpus. Biarpun demikian, tidak dinafikan bahawa ahli terminografi ada menggunakan komputer. Malangnya, penggunaan ini terhad pada penstoran data leksikal dan penerbitan bahan istilah sahaja. Komputer belum lagi dimanfaatkan secara langsung untuk mengumpul dan menganalisis data, sama ada secara automatik mahupun separa automatik.

Sebahagian besar kerja terminografi melibatkan pengenalpastian dan analisis butir-butir leksikal yang digunakan khusus dalam bidang ilmu seperti TMK, perubatan, undang-undang, linguistik dan sebagainya. Bagi bidang-bidang yang pesat

berkembang, umpamanya bidang TMK dan genetik, pengenalanpastian dan penghuraian istilah-istilah baru dalam bidang sebegini menjadi kritikal untuk penyebaran ilmu dengan cepat dan berkesan.

Kegiatan terminografi yang berasaskan korpus sudah terbukti mampu mempercepat sebahagian daripada proses kerja peristilahan (sebagai contoh, Bowker 1996). Dagan dan Church (1997)] telah menunjukkan bahawa “ ... *it took about 10 hours to construct a list of 1700 terms extracted from a 300,000-word document.*”

Lantaran itu, ahli peristilahan sudah harus bersedia untuk menggunakan korpus elektronik, paling tidak, untuk mendapatkan butir-butir leksikal yang bakal diteliti dan diperdebat oleh ahli-ahli jawatankuasa istilah yang dilantik. Dalam konteks ini, penelitian ringkas kami ini dilihat sebagai langkah awal dalam pembinaan sistem terminografi yang berasaskan korpus, meskipun kajian penggunaan istilah berdasarkan korpus sudah pernah dilakukan (Rusli dan Zaiton 2001).

### **3.0 METODOLOGI**

Kajian kami ini melibatkan pemanfaatan korpus yang terdiri daripada teks-teks yang dimuat turun dari laman Web. Cara ini merupakan jalan paling mudah dan cepat untuk mendapatkan teks elektronik yang sesuai (dengan pemformatan yang minimum) untuk diproses oleh sistem teks analisis dan sistem-sistem pemprosesan bahasa tabii yang lain.

#### **3.1 Data**

Korpus bahasa Melayu (TMK\_Bahasa Melayu) yang kami kumpulkan merupakan teks makalah tentang TMK daripada ruangan Berita IT dalam akhbar Berita Harian ([http://www.emedia.com.my/Current\\_News/BH/bhindex](http://www.emedia.com.my/Current_News/BH/bhindex)) dan ruangan Megabait dalam Utusan Malaysia (<http://www.utusan.com.my>). Untuk tujuan kertas ini kami sekadar memuat turun makalah yang tersiar pada bulan Februari dan Mac 2001.

Kami juga mengumpulkan korpus teks makalah TMK berbahasa Inggeris (TMK\_Bahasa Inggeris) yang tersiar pada laman Web Wired News (<http://www.wired.com/news/technology/>), Zdnet (<http://www.zdnet.com/zdnn/>) dan CNET (<http://news.cnet.com/news/>) untuk mendapatkan istilah-istilah mutakhir dalam bidang TMK.

### 3.2 Pemprosesan

Kedua-dua korpus TMK\_Bahasa Melayu dan TMK\_Bahasa Inggeris itu diproses dengan sistem analisis teks<sup>5</sup> untuk mendapatkan senarai kata. Maklumat statistik tentang teks kami abaikan kerana tidak relevan dalam penelitian ini.

Senarai kata yang diperoleh terdiri daripada senarai 'Root Word,' 'Unknown Word' dan 'Invalid Word.' Senarai-senarai ini dibincangkan selanjutnya dalam bahagian yang berikut.

**Jadual 1.0 Senarai 'Root Word'**

```

*****
WORD ANALYSIS - ROOT WORD FREQUENCY
*****

Options :
-----

Number of words      :      All
Frequency            :      All

Total Number of Words      21774   100.00000
  Root Words                19026   87.37944
  Affixed Words              176     0.80830
  New Words                  2566   11.78470
  Numbers                    0     0.00000
  Invalid Words              6     0.02756

#####
Root Word                  Count      (%)
#####
dan                        :          604   2.77395
yang                       :          490   2.25039
...
syarikat                   :          163   0.74860
berkata                     :          161   0.73941
...
negara                      :          120   0.55112
bagi                        :          118   0.54193
juga                        :          110   0.50519
tahun                       :          100   0.45926
...
komputer                   :           60   0.27556
maklumat                   :           59   0.27097
...
dunia                       :           56   0.25719
saya                       :           56   0.25719
adalah                      :           55   0.25259
pusat                       :           55   0.25259
aplikasi                   :           53   0.24341
kami                        :           52   0.23882
pelanggan                  :           52   0.23882
dapat                       :           50   0.22963
serta                       :           50   0.22963
#####
2697                        19202   88.18775
#####

```

<sup>5</sup> Sistem MATA (Malay Text Analysis) yang dibangunkan oleh USM dan digabungkan dalam Sistem Korpus yang digunakan di DBP.

## 4.0 PERBINCANGAN

Pemrosesan data korpus TMK\_Bahasa Melayu dan TMK\_Bahasa Inggeris masing-masing menghasilkan tiga senarai kata, iaitu senarai ‘Root Word,’ ‘Unknown Word’ dan ‘Invalid Word.’

### 4.1 Korpus Bahasa Melayu

Bagi korpus TMK\_Bahasa Melayu, senarai “Root Word’ (Jadual 1.0) terdiri daripada kata dasar yang sedia terakam dalam Kamus Dewan Edisi Ke-3 (DBP 1994), ringkasnya KD3, kerana MATA menggunakan entri KD3 sebagai entri untuk kamus sistemnya.

Kata dasar yang belum terakam dalam KD3 akan tersenarai sebagai ‘Unknown Word’ (Jadual 2.0).

**Jadual 2.0** Senarai ‘Unknown Word’

Unknown Word	Count	(%)
internet	113	0.51897
web	44	0.20208
msc	41	0.18830
3d	39	0.17911
...		
oracle	32	0.14696
e-niaga	24	0.11022
ict	22	0.10104
i-mode	21	0.09645
e-dagang	20	0.09185
megabait	17	0.07807
windows	16	0.07348
dell	15	0.06889
...		
cisco	14	0.06430
...		
tele-kesihatan	14	0.06430
akses	13	0.05970
...		
dsl	12	0.05511
scs	12	0.05511
sms	12	0.05511
...		
java	11	0.05052
...		
gsm	9	0.04133
...		
b2b	8	0.03674
dotcom	8	0.03674
e-book	8	0.03674
optikal	8	0.03674
...		
router	4	0.01837
...		
gigabait	1	0.00459
...		
882	2566	11.78470

Satu lagi senarai yang dioutputkan oleh sistem MATA ialah senarai ‘Invalid Word’ (Jadual 3.0).

**Jadual 3.0** Senarai ‘Invalid Word’

```
#####
Invalid Word                               Count           (%)
#####
pemaparnya                               :                3           0.01378
mengehantar                              :                1           0.00459
menyelusurinya                           :                1           0.00459
sethi                                     :                1           0.00459
#####
4                                          6           0.02756
#####
```

Senarai ini biasanya boleh diabaikan kerana mengandungi bentuk kata yang tidak terleraikan oleh sistem, misalnya bentuk kata dengan ‘-nya’ atau berpartikel ‘-kah’ atau ‘-lah’.

Senarai yang relevan dengan penelitian ini ialah senarai ‘Unknown Word.’ Walaupun bentuk kata yang terdapat dalam senarai ini mungkin belum terakam sebagai entri KD3, namun ini tidak bermakna bahawa kata itu tiada dalam senarai, buku, atau daftar istilah yang diterbitkan DBP.

Oleh itu, sebelum sistem ini dapat digunakan untuk mencari istilah ‘baru’ kamus sistem itu sendiri perlu dikemaskinikan dengan istilah ‘lama.’ Dengan cara ini semua istilah yang belum terakam dalam terbitan DBP tetapi telah digunakan dalam teks akhbar, majalah dan sebagainya dapat disaring sebagai calon istilah dan disalurkan kepada jawatankuasa istilah masing-masing untuk pengesahan dan penyelarasan.

Dalam analisis awal kami dengan menggunakan sistem MATA (tanpa pengemaskinian kamus sistem), bentuk-bentuk istilah yang ‘baru’ sudah dapat disaringkan daripada kata-kata umum. Ini setidaknya-tidaknya dapat mengasingkan bentuk kata umum daripada bentuk-bentuk istilah. Dan jika senarai calon istilah ini

dibandingkan dengan senarai istilah DBP, maka suatu senarai istilah baru akan diperoleh. Perhatikan Jadual 4.0 di bawah.

**Jadual 4.0** Calon Istilah Baru

<b>Istilah DBP</b> (daripada Buku Istilah DBP <sup>6</sup> )	<b>Istilah Baru</b> (daripada Korpus TMK_Bahasa Melayu)
	SDSL (Talian Pelanggan Digital Simetriks)
	PDA (alat bantuan digital peribadi)
	TFT (paparan transistor filem nipis)
memuat turun	memindah turun
	sistem mesej pendek (SMS)
hab	hub

Senarai ini kemudian dijadikan bahan untuk penelitian Jawatankuasa istilah TMK yang berkenaan.

#### **4.2 Korpus Bahasa Inggeris**

Keadaan ini berbeza sedikit dengan korpus bahasa Inggeris kerana sistem MATA menggunakan kamus bahasa Melayu dan rumus morfologi bahasa Melayu dalam pemrosesannya.

Oleh itu, kamus dan rumus morfologi sistem perlu diubah suai terlebih dahulu untuk menangani morfologi bahasa Inggeris dan kemudian semua istilah Inggeris yang menjadi istilah sumber dalam pembentukan istilah bahasa Melayu dijadikan entri dalam kamus sistem. Dengan cara ini, pemrosesan teks bahasa Inggeris akan menghasilkan senarai istilah yang belum digarap oleh mana-mana jawatankuasa istilah DBP.

---

<sup>6</sup> Termasuk juga Glosari Teknologi Maklumat (DBP 2000a) dan Glosari Teknologi Maklumat Jilid 2 (DBP 2000b)



Oleh sebab, sistem MATA belum diubah suai maka kami menggunakan sistem konkordans bahasa Inggeris<sup>7</sup> yang dimuat turun dari WWW untuk penelitian ini.

Pemprosesan dengan Concordance menghasilkan suatu senarai kata yang perlu dibandingkan dengan senarai istilah DBP untuk mendapatkan calon istilah baru yang perlu diberikan padanan.

Antara bentuk istilah baru bahasa Inggeris yang kami peroleh dengan cara ini termasuklah *instant messaging, e-commerce, dot-com incubator*.

## **5.0 KESIMPULAN**

Hasil peristilahan, seperti juga hasil kerja perkamusan, dilihat sebagai 'ketinggalan' atau '*outdated*' sejeurus sahaja diterbitkan. Untuk tidak jauh terkebelakang, suatu kaedah pembentukan istilah yang lebih inovatif harus diikhtiarkan setidak-tidaknya dalam konteks pengumpulan istilah untuk dikerjakan.

Metodologi yang kami lakarkan ini hanya sekadar melibatkan pengumpulan bahan teks yang berkait dengan sesuatu bidang khusus daripada WWW, penyarian calon istilah daripada teks tersebut, perbandingan dengan istilah yang sudah dibentuk, dan perkongsian analisis dan dapatan dengan pakar bidang.

Justeru itu, kajian penggunaan istilah dalam teks ini harus dilihat sebagai langkah awal dalam pembinaan sistem terminografi yang dapat berfungsi secara separa automatik, kalau tidak pun sepenuhnya automatik, untuk mempercepat proses pengumpulan, pembentukan dan penyebaran.

## **EPILOG**

*What a thrill to visit the UH-Hilo campus and to hear everyone speaking exclusively in the beautiful language that so nearly perished. And it's not just the language they are bringing back, it's the culture. Every child I meet there greets me not in the mainland way, by shaking hands, but in the Hawaiian way, with "aloha" and a kiss on the cheek. "A hui ho!" they call after me as I leave: Until we meet again! (Hale ibid.)*

---

<sup>7</sup> Versi demo Concordance (<http://www.rjcw.freeseerve.co.uk/version200.htm>).

## RUJUKAN

- Boguraev, B. dan Briscoe, T. (editor) 1989. *Computational Lexicography for Natural Language Processing*. Wiley: New York & Longman: London.
- Bowker, L 1996. "Towards a corpus-based approach to terminography." *Terminology*, 3(1): 27-52.
- Butler, C.S. (editor) 1992. *Computers and Written Texts*. Blackwell: Oxford.
- Dagan, I. dan Church, K. 1997. "Termight: Coordinating humans and machines in bilingual terminology acquisition." *Machine Translation*, 12:89-107.
- DBP 1987. *Istilah Perubatan Pengajian Tinggi*. DBP: Kuala Lumpur.
- DBP 1994. *Kamus Dewan Edisi Ketiga*. DBP: Kuala Lumpur.
- DBP 2000a. *Glosari Teknologi Maklumat*. DBP: Kuala Lumpur.
- DBP 2000b. *Glosari Teknologi Maklumat Jilid 2*. DBP: Kuala Lumpur.
- Felber, H. 1984. *Terminology Manual*. Paris: UNESCO General Information Programme and UNISIST.
- Hale, C. 1995. "How do You Say Computer in Hawaiian?" *Wired Archive*. [http://www.wired/archive/3.08/hawaii\\_pr.html](http://www.wired/archive/3.08/hawaii_pr.html). Capaian Mutakhir 12 April 2001.
- JKTBM 1992. *Pedoman Umum Pembentukan Istilah Bahasa Melayu*. DBP: Kuala Lumpur.
- Klavans, J. dan Tzoukermann, E. 1995. "Combining corpus and machine-readable dictionary data for building bilingual lexicons." *Machine Translation*, 10:185-218.
- Meyer, I. dan Mackintosh, K. 1996. "The Corpus from a Terminographer's Viewpoint." *International Journal of Corpus Linguistics*, Vol. 1, No. 2.
- Pearson, J. 1998. *Terms in Context*. Amsterdam/ Philadelphia: John Benjamins.
- Rusli Abdul Ghani dan Zaiton Darois 2001. 'Istilah Sains dalam Teks Bacaan Umum.' Dalam *Rampak Serantau*, Bil. 8/2001:67-83.
- Sager, J. 1990. *A Practical Course in Terminology Processing*. Amsterdam/ Philadelphia: John Benjamins.
- Sinclair, J. 1991. *Corpus, Concordance, Collocation*. Oxford University Press: Oxford.

- Sinclair, J.M. (editor) 1987. *Looking Up. An Account of the COBUILD Project in Lexical Computing*. Collins: London/Glasgow.
- Wilks A.Y, Sator B.M., dan Guthrie L.M. 1996. *Electric Words: Dictionaries, Computers, and Meanings*. MIT Press: Cambridge, Mass.
- Zaharin Yusoff 1996. *Cintailah Bahasa Kita: Suatu Tanggapan Linguistik Berkomputer*. Syarahan Pelantikan Profesor. Siri Syarahan Umum USM. Universiti Sains Malaysia: Pulau Pinang.

\* Dibentangkan dalam Seminar Cabaran Penulisan Sains dan Teknologi Dalam Alaf Baru di Kelab Rekreasi UKM, Bangi, pada 25 - 26 April 2001.