ANALISIS SUBJEK VERBAL

B. Mustafa
mus@ipb.ac.id atau mustafa_smada@yahoo.com

Pendekatan subjek dalam era elektronik menjadi cara yang utama dalam mencari informasi. Mesin-mesin pencari informasi di internet telah dilengkapi dengan sistem tertentu untuk menanggulangi masalah ini, namun "peselancar" atau penelesur melalui internet tetap menemui banyak kesulitan karena mendapatkan temuan atau "*hit*" yang sangat banyak, bisa sampai ribuan *hit*. Sehingga disadari bahwa pencarian berdasarkan kata kunci ada kalanya tidak memadai lagi. Kenyataannya banyak kata (baik dalam bahasa Inggris maupun dalam bahasa Indonesia) mempunyai banyak arti. Lain dari pada itu banyak kata yang sama dapat digunakan sebagai kata benda, kata kerja, kata sifat atau kata keterangan. Sistem pencari *(search engine)*, misalnya LYCOS, INFOSEEK, dan sebagainya yang memberi kesempatan penelusur mencari melalui istilah alamiah *(natural language)* sering tidak berhasil memberikan temuan yang tepat *(precision)*, malah hanya memberikan temuan *(recall)* yang sangat banyak.

Di lain pihak, ternyata pengarang pun banyak menggunakan istilah yang berbeda-beda untuk konsep/ide/topik yang sama, sementara orang yang menelusur ide/ konsep/topik yang sama tidak menggunakan istilah yang sama. Itulah sebabnya meskipun banyak juga penelitian yang menyatakan bahwa manfaat pengindeksan diragukan, karena pengindeks biasanya tidak konsisten, kiranya kebutuhan akan *assigned* indeks, yaitu pembuatan indeks berdasarkan kosakata terawasi (sebagai lawan dari *derivated index*, yaitu pembuatan indeks berdasarkan istilah bebas) masih diperlukan.

Ditemukan dalam berbagai penelitian bahwa pengawasan kosakata *(controlled vocabulary)* diperlukan untuk memadukan keragaman istilah yang digunakan untuk menggambar suastu konsep/topik/ide dan untuk membedakan beragam arti yang terdapat dalam suatu istilah tertentu.

Proses Analisis subjek

Ada usaha untuk membuat sistem pengindeksan otomatis (automatic indexing), yaitu menggunakan program komputer untuk menentukan subjek suatu dokumen, namun usaha ini belum berhasil dengan baik. Maih banyak kelemahan dari sistem ini. Sehingga disimpulkan bahwa untuk saat ini pengindeksan oleh pengindeks (*human indexing*) untuk menentukan konsep/ide/topik dari suatu dokumen masih diperlukan.

Analisis subjek adalah bagian dari pengindeksan yang mencakup, satu, analisis konsep dari suatu dokumen atau paket informasi. Analisis konsep adalah usaha untuk menentukan apa isi ilmiah (intelektual) dari suatu dokumen atau menentukan mengenai apa suatu dokumen. Kedua, analisis subjek menyangkut usaha menerjemahkan analisis konsep ke dalam kerangka konsep dalam suatu

sistem klasifikasi (DDC) atau sistem tajuk subjek (Daftar Tajuk Subjek atau Thesaurus, misalnya *Library of Congress Subject Headings*).

Langkah pertama dari suatu analisis konsep adalah memeriksa isi subjek dokumen itu sendiri. Dalam banyak hal, biasanya isi subjek dokumen dapat ditemukan dalam dokumen itu sendiri, tetapi mungkin saja isinya harus dicari pada dokumen lain yang menyertainya. Misalnya panduannya *(user manuals)* atau diperoleh dari luar paket informasi itu misalnya melalui informasi dari penerbitnya.

Utamakan mencari informasi dalam dokumen yang akan diindeks. Misalnya carilah pada:

- Judul dan sub-judul
- Daftar isi
- Pengantar
- Istilah atau frase yang biasa dicetak berbeda dari teks lainnya.
- Ilustrasi, diagram atau tabel

Untuk informasi tidak berupa teks (non-teks), pengindeks perlu memperhatikan objeknya, gambar atau informasi lain.

Kedalaman Pengindeksan *(Exhaustivity)*

Dalam proses pengindeksan dikenal istilah kebijakan pengindeksan *(indexing policy)*. Ini berarti sampai sejauh atau sedalam mana pengindeks melakukan analisis dan penentuan istilah indeks terhadap suatu dokumen yang diindeks. Dikenal misalnya **pengindeksan mendalam** *(Depth indexing)* dan **pengindeksan ringkas** *(Summarization Indexing)*.

Pengindeksan mendalam berusaha mengambil semua konsep utama dari dokumen, termasuk subtopik dan subtema. Sedangkan pengindeksan ringkas hanya mengeluarkan konsep utama dari suatu dokumen. Hanya mengambil konsep yang ada dalam tema utama.

Dalam dunia PUSDOKINFO pada umumnya yang dilakukan adalah pengindeksan ringkas. Pengindeksan mendalam biasanya hanya dilakukan oleh lembaga-lembaga penerbit informasi, misalnya pengindeks jurnal. Jadi di perpustaskaan, pengindeks biasanya hanya berusaha mencari satu konsep utama untuk menentukan isi dari suastu dokumen. Penerbit indeks jurnal biasanya menetapkan konsep yang diambil dari sub-sub bab dari suatu buku atau dari isi suatu artikel pada jurnal.

Identifikasi Konsep

Beragam tipe konsep dapat digunakan sebagai subjek dari suatu dokumen atau paket informasi. Berikut antara lain tipe konsep yang dapat digunakan:

- Topik
- Nama-nama:

    - Orang (person)
    - Badan Koorporasi
    - Nama Geografi
    - Nama entitas (misalnya: bangunan, situs dsb.)
- Periode (waktu)
- Bentuk *(form)*

Keragaman Pengindeksan

Masalah keragaman dan ketidakkonsistenan pengindeks dalam pengindeksan topik antara lain disebabkan karena:

- Latar belakang pendidikan dan pengetahuan
- Waktu pengidneksan
- Perkembangan ilmu pengteahuan
- Informasi bukan dalam bentuk teks

Topik Sebagai Subjek

Nama sebagai Subjek

- Nama orang bisa menjadi subjek dari dokumen. Misalnya dokumen berbentuk biografi.
- Nama badan koorporasi dapat menjadi subjek dari suatu dokumen, misalnya lembaga perminyakan (Pertamina)
- Nama geografi dapat menjadi subjek dari suatu dokumen. Misalnya tarian di Tanah Batak.
- Entitas nama dapat pula menjadi subjek dari suatu dokumen. Misalnya masjid Istiqlal.

Kronologis (waktu) sebagai Subjek

Informasi mengenai waktu dapat menjadi konsep penting dari suatu dokumen. Periode waktu biasanya membatasi cakupan topik pembahasan dalam suatu dokumen, karena itu berarti menunjukkan isi dokumen. Misalnya dokumen yang memuat infromasi mengenai akses informasi menggunakan komputer pada tahun 1970-an, sudah barang tentu tidak mencakup akses informasi mengenai internet atau mengenai SGML.

Konsep Bentuk sebagai Subjek

Kebanyakan pengindeks tidak menganggap konsep bentuk sebagai subjek dari suatu dokumen. Buku mengenai tomat atau jurnal yang memuat artikel mengenai tomat dianggap sama saja. Karena bentuk dokumen bukan mengenai isi dokumen. Tetapi dalam kenyataannya kedua wujud informasi itu tidak pernah ‘berkumpul’ menjadi satu.

Namun demikian pemisahan subjek dari bentuk semakin menjadi penting seiring dengan semakin banyaknya informasi berbentuk non-tekstual. Misalnya dalam

dunia musik, teknologi informasi (map digital). Dengan memisahkan antara subjek dengan bentuk, akan membantu dalam rancangan sistem temu-balik yang memungkinkan menelusur melalui bentuk.

Penerjemahan Konsep ke Istilah Indeks

Setelah analisis konsep dilakukan, tahap selanjutnya adalah menerjemahkan konsep-konsep itu ke dalam istilah yang terawasi *(controlled vocabulary)*, sehingga menjadi istilah indeks. Beberapa alat bantuk untuk menerjemahkan istilah indeks tersebut adalah Thesaurus, Daftar Tajuk Subjek bahkan bagan klasifikasi seperti DDC pun dapat digunakan.

Kosakata Terawasi *(Controlled Vocabulary)*

Prinsip kosakata terawasi adalah pemilihan suatu istilah tertentu yang digunakan mengenai suatu konsep dan mengatur istilah sinonim, keterkaitan dan hubungan dengan istilah lain yang juga digunakan atau tidak digunakan sebagai istilah mengenai konsep tersebut.

Masalah dalam Kosakata Terawasi

Dalam penentuan istilah terawasi, meskipun ada buku panduan yang dapat membantu, masih terdapat beberapa kesulitan yang dihadapi pengindeks, antara lain adalah:

- **Khusus Vs Umum**
  Tingkat kekhususan perlu diperhatikan dalam penentuan istilah yang digunakan. Misalnya pemilihan istilah kucing saja atau kucing angora untuk suatu dokumen mengenai kucing angora.

- **Konsep Sinonim**
  Dalam hampir semua bahasa, termasuk Indonesia dan Inggris, istilah sinonim tidak selamanya mempunyai pengertian yang persis sama. Sesungguhnya ada perbedaan kecil atau nuansa di antara kedua istilah sinonim tersebut.

- **Bentuk Kata untuk Istilah Satu Kata**
  Perhatikan istilah ***online, on-line*** dan ***on line*** walau sama hurufnya karena perbedaan pengejaan, bisa bermakna sama dalam suatu dokumen tertentu, tetapi bisa juga bermakna lain dalam suatu dokumen lainnya.

- **Istilah Bentuk Frase**
  Dalam suatu kosakata terawasi, ada istilah yang terdiri dari lebih satu suku kata, atau disebut juga frase. Hal ini sering menimbulkan perbedaan pengertian. Lihat misalnya frase *Energy conservation* dan *Conservation of energy resources*. Sebenarnya bisa bermakna sama, tetapi mungkin di lain saat orang bisa mengartikan berbeda.

- **Homograf dan Homofon**
  Homograf adalah kata yang teksnya sama tetapi mempunyai makna yang berbeda. Misalnya '*Mercury*', bisa berarti cairan metal, suatu planet, sebuah merek mobil atau dewa Romawi. 'Bunga' bisa berarti bagian pucuk dari suatu tanaman atau bonus/tambahan uang karena menabung pada sebuah bank. Homofon adalah kosakata yang dieja berbeda tetapi diucapkan sama. Fenomena ini kebanyakan dalam bahasa Inggris. Dalam bahasa Indonesia tidak banyak terjadi.

- **Kualifikasi Istilah**
  Kualifikasi istilah adalah menambahkan keterangan di belakang suatu istilah. Umumnya dengan menulis keterangan atau penjelasan dalam tanda kurung setelah istilah itu. Salah satu cara menanggulangi masalah homograf adalah dengan menambahkan kualifikasi istilah. Misalnya: Mercuri (Planet) dan Mercury (Dewa Romawi).

- **Singkatan dan Akronim**
  Keberadaan singkatan dan akronim sering menjadi masalah dalam penentuan kosakata terawasi. Misalnya AIDS, apakah merupakan bantuan *(aids)* atau nama penyakit. Karena itu singkatan dan akronim perlu diberi penjelasan agar jelas maksudnya.

- **Istilah Popular dan istilah Teknis**
  Kalau suatu istilah sama antara istilah teknis dan istilah populer, maka perlu diberi penjelasan oleh pengindeks Misalnya istilah *Neoplasms* dan *Cancer*. Pengertiannya sama, tetapi keduanya digunakan untuk kalangan pembaca berbeda. Kalangan umum mengenalnya sebagai cancer (kanker), tetapi kalangan kedokteran mengenalnya sebagai neoplasms.

- **Subdivisi Istilah**
  Subdivisi digunakan dalam pengawasan kosakata untuk mengatur istilah dalam tahap pra-koordinasi. Misalnya: untuk membedakan melalui bentuk/genre, contoh: Kimia - Kamus. Untuk membedakan antara batas geografi atau waktu, contoh: Perekonomian-Indonesia-Masa penjajahan.