

## BAB KELIMA

### PENGUMPULAN KORPUS DAN ALAT PENGOLAHANNYA

#### 5.0 Pengenalan

Penyelidikan yang dilakukan dalam tesis ini adalah penyelidikan berdasarkan pada korpus (*corpus-based analysis*). Oleh itu, bab ini secara khas akan membicarakan proses dan metode pengumpulan korpus, di samping perisian komputer yang digunakan dalam proses dan kemudian dalam penganalisisannya. Kerana korpus yang digunakan tidak hanya yang berasal dari buku cetak, tetapi juga berasal dari buku elektronik yang diperoleh dari internet, maka dalam bab ini juga akan dikenalkan laman web yang menyediakan pelbagai buku elektronik yang boleh dimuat turun secara percuma.

#### 5.1 Sumber dan Pengumpulan Korpus

Korpus yang dijadikan objek penyelidikan ini berasal dari buku cetak dan buku elektronik. Buku cetak ada tiga jenis, iaitu:

1. Buku Tafsir *Fī Zilāl al-Qur'ān* berbahasa Arab, terbitan Dār al-Syurūq, Beirut tahun 1980 yang terdiri dari 6 (enam) jilid.

2. Buku Tafsir *Fī Zilāl al-Qur'ān* terjemahan bahasa Melayu-Malaysia oleh Yusoff Zaky Haji Yacob yang diberi tajuk *Tafsir Fi Zilalil Qur'an: Di Bawah Bayangan Al-Qur'an*, terbitan Pustaka Aman Press Sdn. Bhd. Cetakan Pertama Tahun 2000/1420 H. Terdiri dari 17 (tujuh belas) jilid.

3. Buku *Fī Zilāl al-Qur'ān* terjemahan bahasa Melayu-Indonesia dengan tajuk *Tafsir Fi Zhilalil Qur'an: Di Bawah Naungan Al-Qur'an*. Tafsir ini diterjemahkan oleh Drs. As`ad Yasin, Abdul Hayyie al Kattani, Lc., H. Dr. Idris Abdul Shomad, H.

Harjani Hefni, Lc., H. Ahmad Dumyati Bashori, M.A., Abu Ahmad `Izzi, M.A., H. Samson Rahman, M.A., Hidayatullah, Lc., H. Bakrun, M.A., H. Zainuddin Bashiran, Lc., H. Fauzan, Lc., K.H. Mufti Labib, MCL, Tajuddin, Lc., Drs. Muchotob Hamzah, Drs. Syihabuddin, M.A. Editor ahli: Ust. Abdul Aziz Salim Basyarahil dan Dr. Hidayat Nur Wahid, M.A. Diterbitkan oleh Gema Insani, Jakarta. Cetakan Keempat, Muharram 1426 H/ Maret 2005 M.

Sedangkan korpus dalam bentuk elektronik diambil dari laman web <http://www.almeshkat.net/books/> berupa dokumen format *Microsoft Word*. Untuk menyemak keakuratan kandungan buku elektronik ini digunakan perisian al-Maktabat al-Syāmilat yang teks di dalamnya berektensi \*.bok. Setakat ini teks elektronik baru korpus berbahasa Arab, sementara terjemahannya sama ada dalam bahasa Melayu-Malaysia ataupun Melayu Indonesia, penyelidik belum mendapatkannya selain dalam bentuk buku.

### **5.1.1 Sekilas tentang laman web penyedia sumber korpus**

Berkat kemajuan teknologi maklumat pada masa sekarang ini, korpus bahan analisis dapat dengan mudah diperoleh. Buku-buku elektronik dalam format *Microsoft Word*, misalnya dapat dimuat turun dari <http://www.almeshkat.net/books/> yang jumlah dan macam bukunya kerap dikemaskinikan.

Buku elektronik dalam format PDF dapat dimuat turun dari pelbagai laman web, seperti [www.scribd.com](http://www.scribd.com), [www.archieve.com](http://www.archieve.com), [www.esnips.com](http://www.esnips.com), [www.4shared.com](http://www.4shared.com) dan masih banyak lagi yang lainnya.

Khusus untuk buku elektronik yang berformat \*.bok setakat ini hanya dapat dibuka dengan perisian khusus iaitu al-Maktabat al-Syāmilat. Versi terakhir yang ada ketika penyelidikan ini dibuat adalah versi 3.8 yang dapat dikemaskinikan menjadi

versi 3.9. Perisian dapat dimuat turun dari [www.islamport.com](http://www.islamport.com). Al-Maktabat al-Syāmilat Versi 3.8 terdiri dari 6688 buku. Perisian diinstall secara mandiri (*portable*) sehingga dapat dijalankan tanpa bergantung pada *Operating System* seperti Windows.

## 5.2 Penyiapan Data

Seperti telah dikemukakan di atas bahawa teks elektronik Tafsir *Fī Zilāl al-Qur'ān* diperoleh dari <http://www.almeshkat.net/books/>. Untuk memudahkan pengumpulan data, teks elektronik tafsir ini dikelompokkan menurut surah, dan untuk memudahkan pencarian letak data kolokasi dan terjemahannya, teks elektronik ini diberi halaman mengikuti halaman dalam buku.

### 5.2.1 Ukuran Data<sup>1</sup>

Data berbahasa Arab dalam bentuk buku sebanyak 4012 halaman.

Data berbahasa Arab dalam bentuk elektronik sebanyak 1.905.227 kata.

Data berbahasa Melayu-Malaysia 10.482 halaman.

Data berbahasa Melayu-Indonesia 4924 halaman.

### 5.2.2 Penyuntingan Data

Setelah dilakukan penyemakan terhadap data elektronik, diketahui bahawa banyak halaman yang hilang. Untuk itu, dengan berpedoman kepada buku Tafsir *Fī Zilāl al-Qur'ān* yang berbentuk buku, halaman-halaman yang kurang dapat dilengkapi dengan mengambil dari perisian al-Maktabat al-Syāmilat. Dengan demikian

---

<sup>1</sup> Data dalam jumlah kata hanya dapat diperoleh dari buku elektronik, sedangkan dari buku cetak biasa penulis mendapat kesulitan untuk menghitungnya.

penambahan dapat dilakukan dengan mudah dan cepat, tak payah mentaip, kerana dalam beberapa surah, kekurangan tersebut dapat mencapai puluhan halaman.

### **5.3 Perisian Komputer Yang Digunakan dalam Penyelidikan ini**

Perisian utama yang digunakan dalam penyelidikan ini adalah *Microsoft Word* 2003. Selain untuk menyunting data elektronik, perisian ini juga digunakan untuk mengubah data berformat \*.doc menjadi \*.txt. Data dengan ekstensi \*.txt ini diperlukan dalam pencarian konkordansi dengan menggunakan perisian *aConcorde*.

Selain *Microsoft Word*, perisian *Microsoft Excel* juga digunakan untuk mengumpulkan sampel kolokasi yang dikelompokkan mengikut jenisnya.

#### **5.3.1 Sekilas tentang Perisian *aConcorde***

Perisian pencarian konkordansi *aConcorde Version* 0.4.1 diperoleh dengan percuma dari <http://www.comp.leeds.ac.uk/andyr/software/aConCorde/>. Perisian dibuat oleh Andrew Robert, pakar komputer dari *School of Computing*, University of Leeds, Leeds, United Kingdom yang memperoleh bantuan dari Latifa Al-Sulaiti, Bayan Abu Shawar dan Soleh Al-Osaimni dalam menterjemahkannya ke bahasa Arab.

Dalam proses pengumpulan data, *aConcorde* ini hanya dapat digunakan untuk mencari konkordansi kata-kata tertentu sahaja. Misalnya, kata *المَعْرَكَة* /al-ma‘rakat/ (lihat rajah di bawah). Sementara banyak jenis kolokasi harus dilakukan secara manual dengan membaca secara teliti dan mengelompokkannya menurut jenis masing-masing.



Rajah 5.1: Contoh hasil pencarian kata *المعركة* /al-ma'rakat/ dengan *aConcorde*

#### 5.4 Kamus-kamus Yang Digunakan:

Bagi mengetahui secara pasti bahawa kombinasi kata-kata yang telah ditemukan sama ada dengan menggunakan *aConcorde* atau pun yang didapatkan secara manual adalah kolokasi, penulis memastikannya dengan merujuk kepada kamus-kamus khas tentang kolokasi dan idiom, seperti: *المُعْجَمُ السِّيَاقِي لِلتَّعْبِيرَاتِ* /al-mu'jam al-siyāqī li al-ta'bīrāt al-iṣṭilāḥiyyat/ 'A Contextual Dictionary of Idioms' (1996) karya Maḥmūd Isma'īl Ṣīnīy, Mukhtār al-Ṭāhir Ḥusayn, dan Sayyid 'Iwaḍ al-Karīm al-Dawṣy yang diterjemahkan ke bahasa Melayu-Indonesia oleh Drs. Basuni Imamuddin M.A dan Dra. Nashiroh Ishaq dengan tajuk *Kamus Idiom Arab-Indonesia: Pola Aktif* (2005), *مُعْجَمُ الْحَافِظِ لِلْمُتَصَاحِبَاتِ الْعَرَبِيَّةِ* /mu'jam al-ḥāfiẓ li al-mutaṣāḥibāt al-'arabiyyat/ 'Al-Hafiz Arabic Collocations Dictionary' (2004) karya Dr. al-Ṭāhir bin 'Abd al-Salām Ḥāsyim Ḥāfiẓ, dan *مُعْجَمُ التَّعَابِيرِ الْإِصْطِلَاحِيَّةِ فِي الْعَرَبِيَّةِ*

المُعَاصِرَة /mu‘jam al-ta‘ābīr al-iṣṭilāḥiyyaṭ fī al-‘arabiyyaṭ al-mu‘āṣiraṭ/ ‘Kamus Idiom dalam Bahasa Arab Kontemporari’ (2007) karya Dr. Wafā’ Kāmil Fāyid, pensyarah linguistik di Fakulti Sastera, Universiti Kaherah.

## 5.5 Penutup

Korpus elektronik yang menjadi bahan penyelidikan ini memang belum terlalu sempurna sehingga masih lagi diperlukan penyuntingan sehingga isinya sesuai dengan korpus dalam bentuk buku yang dicetak. Demikian pula dengan perisian yang digunakan dalam mengolah korpus tersebut, seperti *aConcorde* ini. Tidak semua data dapat diperoleh dengan menggunakan perisian percuma ini.

Sampel data kolokasi yang dilakukan secara manual, diambil dari bahagian mukadimah, tafsir surah al-Fatihah ditambah tafsir tujuh surah panjang, iaitu al-Baqaraṭ, Ali ‘Imrān, an-Nisā’, al-Ma’idaṭ, al-An‘am dan al-A‘rāf. Pencarian secara manual adalah untuk kolokasi unsur-unsur bahasa – terutama partikel – yang letaknya tidak berdampingan, dan karena itu tidak mungkin dilakukan untuk seluruh pentafsiran surah. Adapun perisian *aConcorde* hanya dapat digunakan untuk mencari kolokasi unsur-unsur bahasa yang letaknya memang berdampingan dan ini dapat dilakukan untuk seluruh data dalam tafsir *Fī Zilāl al-Qur’ ān*.