

## Question - Answering

(Tanya – Jawab)

## Question-Answering

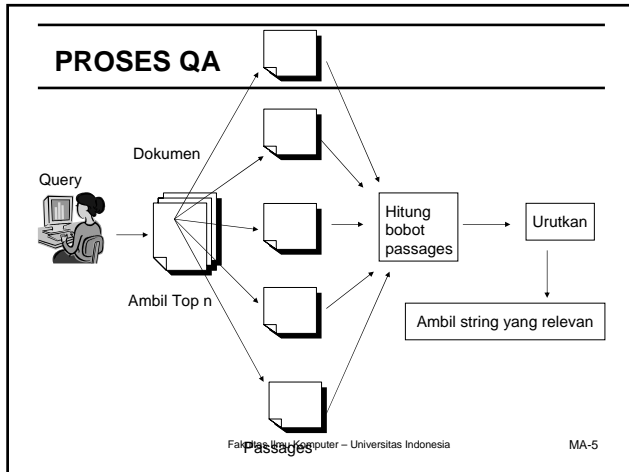
- Query : NLP (bukan keywords)
- Output : Jawaban yang tepat, bukan dokumen
- Teknik QA: Kombinasi IR dan NLP

## Jenis Pertanyaan

- Berdasarkan definisi (fakta)
  - Orang: pekerjaan, peranan, posisi
    - ◆ Who is George W. Bush?
  - Organisasi: nama, kegiatannya dsb.
    - ◆ What is UNICEF?
  - Lokasi
    - ◆ What is the capital of Somalia?
  - Waktu dari suatu kejadian (date)
    - ◆ When did Napoleon die?
  - Ukuran (kuantitas, jarak, lamanya - measure of distance)
    - ◆ What is the world's population?
  - Cara suatu hal bisa terjadi (how)
    - ◆ How did Ayrton Senna die?
  - Pertanyaan bisa tidak mempunyai jawaban

## Evaluasi

- Dilakukan oleh manusia
- Jawaban dinilai dari segi
  - *Reponsiveness* (pasangan jawaban-dokid)
  - Ketepatan (untuk setiap jawaban)
- Pemberian penilaian
  - Wrong (W) : jawaban tidak benar
  - Unsupported (U): jawaban benar tp dok tidak mendukung
  - Inexact (X): jawaban & id benar tp terlalu panjang
  - Right (R) : jawaban & dok benar



### Proses Perolehan Jawaban

- Tahapan QA:
  - Kirim query pada koleksi dokumen
  - Ambil top n dokumen
    - ◆ Bisa juga membagi top n dokumen menjadi *passages* yg berisi beberapa kalimat, sejumlah kata, atau paragraf
  - Terapkan pemberian entitas (nama, lokasi, waktu dsb.) secara otomatis pada *query* dan dokumen. Misalnya menggunakan *Lingpipe* (<http://www.alias-i.com/lingpipe>)

Fakultas Ilmu Komputer – Universitas Indonesia MA-6

### Proses Perolehan Jawaban

- Hitung bobot dari setiap kalimat/*passages*:
  - Buang *passage/kalimat* yang tidak mempunyai tagging yang sesuai dengan *query*. Misalnya *query* mempunyai tagging "lokasi", sehingga *passage* yg tidak mempunyai tagging "lokasi" tidak akan digunakan.
  - Berikan tambahan bobot (mis. 0.5) jika semua kata pada *query* ada pada *passage/kalimat* dan jika semua kata pada *query* muncul pada satu kalimat.

Fakultas Ilmu Komputer – Universitas Indonesia MA-7

### Proses Perolehan Jawaban

- Urutkan kalimat/*passage* sesuai dengan bobot yang telah dihitung
- Hitung jarak antara kata yang mempunyai tagging yang sesuai dengan *query* dengan kata yang muncul pada *query*

Fakultas Ilmu Komputer – Universitas Indonesia MA-8

## Proses Perolehan Jawaban

- Query :
  - What is the capital of <LOCATION> Somalia?
- Passage :
  - Here there is no coordination. <PERSON> Steffan de Mistura – UNICEF representative in the Somali Capital, <LOCATION> Mogadishu, and head of the anti-cholera team – said far more refugees are crowded together here without proper housing or sanitation than during the <LOCATION> Somalia crisis. And many are already sick and exhausted by the long trek from <LOCATION> Rwanda.
  - Jarak antara capital dengan Mogadishu = 1, Rwanda = 38.
- String yang diambil adalah yang mempunyai jarak yang terkecil dan yang paling banyak mempunyai kata sesuai query.

Fakultas Ilmu Komputer – Universitas Indonesia

MA-9

## Teknik QA Yang Lain

- Pencocokan jawaban dengan *snippet* dari mesin pencari
- Mengumpulkan data mengenai fakta-fakta (mis. Nama presiden, nama gunung dsb)
- Membuat *template* dari pertanyaan dan jawaban

Fakultas Ilmu Komputer – Universitas Indonesia

MA-10