

Factoid Spørsmål–Svar

INF5820 – H2008

Jan Tore Lønning

Institutt for Informatikk
Universitetet i Oslo

25. september

Outline

- 1 Grunnleggende
- 2 Prosessering av spørsmålet
- 3 Finn passasjer
- 4 Prosesser svar
- 5 Noen alternativ

Outline

- 1 Grunnleggende
- 2 Prosessering av spørsmålet
- 3 Finn passasjer
- 4 Prosesser svar
- 5 Noen alternativ

Factoide

Merriam-Webster

- 1 an invented fact believed to be true because of its appearance in print
- 2 a briefly stated and usually trivial fact

Wikipedia

- A factoid is a spurious — unverified, incorrect, or fabricated — statement formed and asserted as a fact, but with no veracity.
- (Mot slutten:) The word *factoid* is now sometimes also used to mean a small piece of true but valueless or insignificant information, in contrast to the original definition.

Factoide Spm.-svar

Enkle spørsmål som kan besvares med ett eller noen få ord.

- *Hvem er president i Finland?*
- *Når ble Finland en selvstendig stat?*
- *Hvilke land grenser til Svartehavet?*

Trinnene

- 1 Prosesser spørsmålet (reformuler, normaliser, ekspander, ...)
- 2 Finn relevant informasjon for svaret (“passage retrieval”):
 - Finn dokumenter (med enkle IR-metoder)
 - Finn relevant informasjon i dokumentene (evt. med mer avanserte metoder)
- 3 Formulér svaret

Outline

- 1 Grunnleggende
- 2** **Prosessering av spørsmålet**
- 3 Finn passasjer
- 4 Prosesser svar
- 5 Noen alternativ

Spørsmålsformulering

Kan inngå:

For store dokumentmengder (web):

- Bare trekke ut termer fra spørsmålet og søke på disse
- Dette kan overlates søkemaskinen hvis vi bruker en standard web-søker
- Evt. formulere fraser som vi venter å finne i svaret, f.eks.
 - *president i Finland er,*
 - *Finland ble selvstendig stat*
- mer presise søk, mindre søkerom

Ved mindre dokumentmengder:

- Ekspander søket ved f.eks. å ta med hyponymer

Spørsmålsklassifkasjon

- En klassifisering er til hjelp for å finne hva slags svar vi er ute etter
- Svartypetaksonomi, el. spørsmålsontologi
- Se på spørreordet (*hvem, hva, hvor, når, hvilken*) og på **hovedordet** (*hvilken by...*) for å klassifisere spørsmålet i forhold til ontologien.

Outline

- 1 Grunnleggende
- 2 Prosessering av spørsmålet
- 3 Finn passasjer**
- 4 Prosesser svar
- 5 Noen alternativ

To trinn:

- 1 Finn relevante dokumenter med vanlig IR-teknikker
- 2 Se på disse dokumentene og finn relevante passasjer
 - Utelukk de som ikke kan gi et svar, f.eks.
 - Bruk NER (named entity recognition) og utelukk passasjene som ikke inneholder noen av rett type ut i fra spørsmålsontologien
- 3 Ranger aktuelle passasjer, kombiner informasjon
 - Antall navngitte entiteter
 - antall spørsmålsnøkkelord
 - lengst eksakte match mellom spørsmålsformulering og svar
 - m.fl.

Outline

- 1 Grunnleggende
- 2 Prosessering av spørsmålet
- 3 Finn passasjer
- 4 Prosesser svar**
- 5 Noen alternativ

Metoder

- 1 Mønstergjenkjenning
 - NE-type
 - mønstergjenkjenning
- 2 n-gram “tiling”

NE-type

- Ved spørsmålsklassifikasjonen kan vi ha funnet ut hva slags type entitet vi er på utkikk etter, f.eks. HUMAN eller DATO.
- Vi kjører NER på de innsamlede passasjene og ser om vi finner fraser (navn) av riktig type.

Mønstergjenkjenning

Example

Eksempel	Mønster
<i>Ask er et løvtre</i>	<QP> is <AP>
<i>... ask, et løvtre...</i>	<QP>, <artikkel> <AP>
<i>... et løvtre slik som ask...</i>	<AP> slik som <QP>

- Mønster kan konstrueres for hånd, eller
- læres

Mønsterlæring

- 1 Start med en håndbygget liste av korrekt par, f.eks.
Gandi:1869, Mozart:1756
- 2 Finn dokumenter (på web) som inneholder begge termene i et par.
- 3 Trekk ut setninger som inneholder begge termene.
- 4 Finn mønster mellom par av term.
- 5 Se på de mønsterne som er funnet, hvilke som gjentar seg osv.

N-gram “tiling”

Litt for kort forklart. Vi lar det ligge.

Outline

- 1 Grunnleggende
- 2 Prosessering av spørsmålet
- 3 Finn passasjer
- 4 Prosesser svar
- 5 Noen alternativ**

Nye søkemotorer

Example

Hakia

- www.hakia.com
- factoide-søk

Example

PowerSet

- www.powerset.com
- lingvistisk basert (factoide) søk
- Parser Wikipedia med LFG (Lexical-Functional Grammar)
- En viss semantisk analyse

Example

- *Hvem er ordfører i Norges hovedstad?*
- Finner ikke svare direkte, men kan kombinere informasjonsbiter:
 - *Oslo er Norges hovedstad.*
 - *Fabian Stang er ordfører i Oslo*
- En form for slutning (“textual entailment”)
- Vi er ikke ferdig med å finne dokumenter. Først når vi har funnet at Oslo er Norges hovedstad, kan vi lete etter dokumenter for Oslos ordfører.