

# T-(538|725)-MALV, Málvinnsla Sameining markara

Hrafn Loftsson<sup>1</sup> Hannes Högni Vilhjálmsson<sup>1</sup>

<sup>1</sup>Tölvunarfræðideild, Háskólinn í Reykjavík

September 2007

**1** Sameining markara

**2** Kosning

**1** Sameining markara

**2** Kosning

# Sameining markara (e. combination of taggers)

- Markarar sem byggja á mismunandi aðferðum (þekkingu) hafa tilhneigingu til að gera mismunandi villur.
- Hægt er að nýta sér þennan mun með því að:
  - Láta fleiri en einn markara markaða sama texta og kjósa svo á milli markanna sem þeir úthluta (kallað “voting”).
  - Þjálfa “flokkara” (e. classifier) í að velja mörk að gefnu úttaki úr einum eða fleirum mörkurum (kallað “stacking”).

## Áhrif á nákvæmni

- Sameining markara hefur oftast í för með sér meiri nákvæmni samanborið við notkun einstakra markara.
- Að bæta við markara í sameiningarsafnið er yfirleitt árangursríkt.
  - Jafnvel þó að bætt sé við “ónákvæmum” markara.
  - Hins vegar hjálpar yfirleitt ekki til að bæta við markara sem er á einhvern hátt svipaður þeim sem fyrir eru.

## Hvenær fýsilegt?

- Gera má ráð fyrir að tíminn til að meðhöndla texta með  $N$  mörkurum sé a.m.k.  $N$  sinnum meiri heldur en sá tími sem fer í að marka texta með einum markara.
- Því er fýsilegt að sameina markara þegar keyrslutíminn skiptir miklu minna máli en hin aukna nákvæmni sem fæst með sameiningunni.
- Dæmigerð notkun er við mörkun málheilda.
  - <http://www.lexis.hi.is/malheild.htm>

# Outline

1 Sameining markara

2 Kosning

## Einföld kosning (e. simple voting)

- Sérhver markari fær að kjósa með einu atkvæði.
- Markið með flest atkvæði hlýtur kosningu.

## Vegin kosning (e. weighted voting)

- Atkvæði markara fá mismunandi vigt.
- “Nákvæmur” markari fær t.d. meiri vigt heldur en aðrir markarar.
  - T.d. vigt sem er jöfn útreiknaðri nákvæmni markarans.
- Markari sem vitað er að er nákvæmur fyrir tiltekið mengi af mörkum gæti fengi meiri vigt fyrir mörk úr því mengi.



## Einföld kosning (e. simple voting)

- Sérhver markari fær að kjósa með einu atkvæði.
- Markið með flest atkvæði hlýtur kosningu.

## Vegin kosning (e. weighted voting)

- Atkvæði markara fá mismunandi vigt.
- “Nákvæmur” markari fær t.d. meiri vigt heldur en aðrir markarar.
  - T.d. vigt sem er jöfn útreiknaðri nákvæmni markarans.
- Markari sem vitað er að er nákvæmur fyrir tiltekið mengi af mörkum gæti fengi meiri vigt fyrir mörk úr því mengi.

## Orðtíðnibók: Einföld kosning

Markari	Óþekkt orð	Þekkt orð	Öll orð
TBL*	66.30%	91.90%	90.15%
TnT*	72.75%	92.53%	91.18%
IceTagger	75.09%	92.74%	91.54%
Einföld kosning	76.76%	93.77%	92.61%

\* Breytt útgáfa af viðkomandi markara.

Hrafn Loftsson (2006). Tagging Icelandic text: an experiment with integrations and combinations of taggers.

## Orðiðibók: Einföld kosning

Markari	Óþekkt orð	Þekkt orð	Öll orð
MXPOST	62.29%	91.00%	89.03%
MBT	59.40%	91.47%	89.28%
TBL*	66.30%	91.90%	90.15%
TnT*	72.75%	92.53%	91.18%
IceTagger	75.09%	92.74%	91.54%
Einföld kosning	78.70%	94.36%	93.29%

\* Breytt útgáfa af viðkomandi markara.

Hrafn Loftsson (2006). Tagging Icelandic text: an experiment with integrations and combinations of taggers.