



Orðabók Háskólans

Mörkuð íslensk málheild

Söfnun efnis fyrir Markaða íslenska málheild

Eyrún Ellý Valsdóttir



1. Um efnissöfnunina
2. Hverju hefur verið safnað?
3. Samþykki



Upphaf efnissöfnunar í júní 2006

- Nýsköpunarverkefni á vegum OH
 - verkefnisstjórar SH og Ásta Svavarsdóttir
 - Einblínt á bloggtexta, tölvupósta og ræður/erindi
- Fyrst og fremst óútgefið efni
- Form textanna: Pdf-form eða textaform
- Efni fyrst safnað efni af netinu og síðan óskað leyfa.



Staða söfnunar í lok árs 2007

- 9 milljónum orða hefur verið safnað
 - Dagblöð: 2,1 milljón orð
 - Tímarit: 1,2 milljónir orða
 - Ýmislegt útgefið: 3 milljónir orða
 - Ýmislegt óútgefið: 200 þúsund orð
 - Til upplestrar: 500 þúsund orð
 - Óflokkað: 2 milljónir orða
- Bæst við: 5 milljónir orða frá Mbl. og efni frá útgáfufélaginu Birtíngi.



Morgunblaðið

FRÉTTABLAÐIÐ

- **Dagblöð**

- Alls nú um 7 milljónir orða
- Morgunblaðið og Fréttablaðið 2000-2007
- Leysa þarf upp úr pdf-skjölum
 - Dálkaskipting torveldar úrvinnslu



- Morgunblaðið

- um 5 milljónir orða
- mjög fjölbreytt efni
- ca. 9 tbl. tekin á hverju ári og dreift yfir allt tímabilið
- komið inn í TOH

- Fréttablaðið

- 1,5–2 milljónir orða
- af-pdfun slæm frá 2006
- 5 daga tímabil (nærri heil vika)
- dreifist á sjö ár



- **Tímarit**

- Greinum safnað af rafrænum útgáfur tímaritanna

- t.d. Læknablaðið á www.laeknabladid.is/tolublod

- Wordskjöl og pdf-skjöl

- forrit Antons, *Sniðald*, til að taka texta út úr word- og pdf-skjölum og vista á txt.formi



Orðabók Háskólans

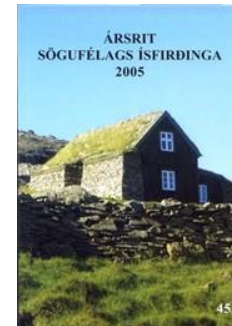
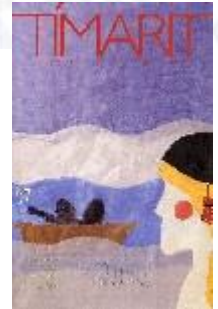
Mörkuð íslensk málheild

- **Ýmis sérblöð Fréttablaðsins**
 - *Allt atvinna, Allt fasteignir, Markaðurinn, Sirkus, Sport*
- **Útgáfufélagið Heimur**
 - *Frjáls verslun, Ský, Veiðimaðurinn, Vísbending*
- **Útgáfufélagið Birtíngur**
 - *Gestgjafinn, Golfblaðið, Hús og híbýli, Ísafold, Leifur, Lifið heil, Mannlíf, Nýtt líf, Sagan öll, Séð og heyrt, Vikan*





- **Tímarit Máls og menningar (TMM)**
 - Menningarlegt efni; tónlist, bókmenntir, kvikmyndir o.s.frv.
- **Sögufélag Ísfirðinga**
 - sagnfræðilegar greinar
- **Eins og fólk er flest**
 - tímarit Alþjóðahússins, málefni nýbúa





- **RHÍ fréttir**
 - Fréttabréf Reiknistofnunar HÍ
- **Íslenskt mál og almenn málfræði**
 - Textar á txt-formi hjá Íslenska málfræðifélaginu
- **Málfregnir**
 - www.ismal.hi.is/Efnisyfirlit.htm
- **Læknablaðið**
 - Safnað af rafrænni útgáfu greina á www.laeknabladid.is
 - Stórar greinar, > 300-400 orð





Orðabók Háskólans

Mörkuð íslensk málheild



Leita

Vísindavefurinn

The screenshot shows the Vísindavefurinn website. At the top, there is a search bar with the text "Leita" and a magnifying glass icon. Below the search bar, there are several logos and banners, including "Forsíða Fyrirtækis", "MERA FYRIR NAMSMENN", and "Landsbankinn". A navigation menu on the left lists various categories: "Vísindavefurinn", "Hvers vegna?", "Vegna þess", "Tenglar", "Um vefinn", "Þjónusta", and "Aðalstyrktaraðili". The main content area is titled "Nokkur svör" and contains several questions and answers, such as "Hvaða stéttir á Íslandi hafa ekki verkfallsrétt?" and "Getur stéttarfélag afsalað sér verkfallsrétti svo að bindandi sé?". There is also a "Senda spurningu" button and a "Senda" button. The bottom of the page shows "Svar dagsins" and "Svör eftir flokkum".

Ýmislegt útgefið

- Vefmiðlar, Vísindavefurinn, vefmiðlar og rafrænar skýrslur
- Flest útgefið á netinu
- Einnig safnaðarblöð Garða- og Bessastaðasóknar



- **Vefmiðlar**
 - héraðsfréttamiðlar
 - *Suðurlugginn* og *Dagur.net* samþykkt
- **Vísindavefurinn**
 - Alþýðufræðiefni (veðurfræði, líffræði o.s.frv.)
 - Komið inn í TOH
- **Vefsetur** fyrirtækja, félaga og stofnana
 - t.d. Barnaheill, Ferðamálastofa, Hjálparstarf kirkjunnar, Sorpa, Vegagerðin, Lýðheilsustöð, Neytendasamtökin
 - Fjölbreytt efni, allt frá fréttum til fróðleiks



- **Dómar**
 - héraðs- og hæstaréttardómar af www.domstolar.is
 - ca. 87.000 orð
- **Lög og reglugerðir**
 - af althingi.is
- **Greinargerðir og skýrslur**
 - frá ráðuneytum og vefsetrum fyrirtækja og stofnana
 - ca. 7–10 pdf frá hverju
 - ólíkt að efni og stærð



Orðabók Háskólans

Mörkuð íslensk málheild

Ýmislegt óútgefið

- Veftímarit og ritgerðir (menntaskóla- og háskólanemendur)
- Leitað eftir leyfum hjá einstaka höfundum



- Veftímarit

- *Vefnir* (Tímarit Félags um 18. aldar fræði)

- efni frá 2003-2008

- *Netla*

- www.netla.khi.hi.is

- *Tímarit um raunvísindi og stærðfræði*

- www.raust.is

- dreifa höfundum yfir fræðasviðin til að fá breidd í orðaforða og –notkun

- *Tölvumál*

- www.sky.is/tolvumal





Ritgerðir nemenda í Verzlunarskólanum

- nemendur leitaðir uppi með netleit
- efni frá árunum 2002–2007
- samþykki fékkst frá 210 höfundum
- stúdentsprófsritgerðir í íslensku, sögu o.fl.
 - ritgerð ca. 500-1000 orð hver
- bréf send í nokkrum hollum (fjölpóstur)
 - sniglapóstur lokaúrræði
- í úrvinnslu eru kynin flokkuð saman



Háskólaritgerðir

- lokaritgerðir og fagritgerðir
- alls 5 höfundar gefið samþykki og sent ritgerðir
- Skólar og deildir sem leitað var til:
 - **HÍ** (hugvís-, raunvís-, laga- og viðskiptadeildir)
 - **LHÍ** (myndlist, tónlist, leiklist)
 - **HA** (kennaranám)
 - **Bifröst** (viðskiptafræði)
 - **HR** (lagadeild)
- Listi fékkst frá HÍ (hugvís), HA (kennara) og LHÍ



- Til upplestrar
 - ræður, erindi, fréttahandrit
 - upphaflegar hugmyndir: stólræður presta, minningarorð og fréttahandrit
 - viðbætur: ræður og erindi sem fundust við netleit á vefsetrum



- Söfnun frá 2006: Íslenskt mál
 - útvarpserindi starfsfólks OH
 - sett inn í TOH
- Ýmis erindi og ræður
 - höfundar ýmsir formenn félaga og fyrirtækja, t.d. Barnaheill, Ferðamálaráð o.s.frv.
- Fréttahandrit frá RÚV
 - útvarp og sjónvarp
 - fengið frá Fjölmiðlavaktinni
 - www.textavarp.is (fjölbreyttara en fréttatextar)



- Óflokkað
 - bloggtextar og tölvupóstlistar
 - minnsti flokkurinn skv. upphaflegri söfnunaráætlun fyrir MÍM
 - er nú einn af stærstu flokkunum, um 2 milljónir orða



Bloggtextar

- Höfundar eru: frambjóðendur sveitarstjórnarkosninga 2006 (5), prestar (8), alþingismenn (10), aðrir (50).
- Bloggarar leitaðir uppi
 - Nafnlausir bloggarar góð lausn?
 - allur texti á vefnum verndaður af höfundarrétti, þó að hvergi sé hægt að rekja hann til einstaklinga
- Öllum blogg færslum einstakra höfunda steyppt saman eftir efni
 - skipt í þrennt; stjórnámálanna-, presta- og almennt blogg



Tölvupóstur

- Póstlistar frá félagasamtökum, fyrirtækjum og HÍ (alls 23 listar)
 - hi-starf, kynjafræði, tmem-hi, arnastofnun, deildartungur, gandur, deildarmeiningar, hi-nem, Femin.is, Gjordusvovel.is, Hress, Hressó, Hreyfing, Kvinnakórinn, LFC-búðin, Salka, Vegamót, Valur, Þróttur, Plúsferðir, Óperupósturinn.
- Mismikið efni sem barst frá listunum; sumu var aðeins safnað í nokkrar vikur en öðru yfir lengri tíma.
 - ítrekun send eftir samþykki ef orðafjöldi fór yfir 1000 orð
- Póstur á póstlistum í eðli sínu ópersónulegur
- Fékk jákvæð svör frá þeim sem svöruðu
 - oft erfitt fyrir umsjónarmenn listanna að finna gamla pósta hjá sér.



- Flokkun póstlista
 - heiti
 - skilgreining
 - 1) *Almennt*
 - 2) *HÍ*
 - mánuðir
 - ár
 - orðafjöldi
 - eðli pósta
 - 1) *fræði*
 - 2) *auglýsingar*
 - 3) *verslun*
 - 4) *félög*



- Það sem eftir stendur:
 - Sveitarstjórnarmál, Eiðfaxi, Athygli, háskólaritgerðir frá Landbúnaðarháskólanum og Myndir mánaðarins
- Einnig á að safna námsbókum, skólablöðum, notkunarleiðbeiningum og auglýsingartextum.



Samþykkisveitingar höfunda

- Ritstjórar og einstakir höfundar
- Kynningarbréf sent þar sem verkefnið var kynnt og óskað eftir leyfi
 - afrita textann í skjal og skrifa undir með nafni og kennitölu
 - tölvupóstssvar/bréfleiðis/fax
 - Viðhengi í fyrstu →
www.lexis.hi.is/malheild.htm



- Vísindavefur: Efnið sem við fáum hefur verið samþykkt.
- Vefsetur: Forstöðumaður/vefstjóri veittu samþykki sitt.
- Nemendur veittu fúslega leyfi fyrir Verslórítgerðum
 - afrit af ritgerðinni?
- Netla: Sett inn klausa í höfundarskilmála í október 2007



Höfundar athugið!

Þeir sem birta greinar í Netlu heimila að texti greina þeirra verði nýttur til rannsókna, orðabókagerðar og tungutækniverkefna af Orðabók Háskóla Íslands (sem nú er hluti af Stofnun Árna Magnússonar í íslenskum fræðum).

Textanum verður komið fyrir í textasafni stofnunarinnar og hluti af honum nýttur fyrir verkefnið "Markaða íslenska málheild". **Einnig er gert ráð fyrir að nýta megi á sama hátt texta fyrri greina þinna sem birtar hafa verið í Netlu.**

Kynningarefni um Markaða íslenska málheild (MÍM) er aðgengilegt hér (<http://www.lexis.hi.is/malheild.htm> <<http://www.lexis.hi.is/malheild.htm>>).

Ábyrgðaraðilar Netlu hafa þegar veitt verkefninu fulltingi sitt.