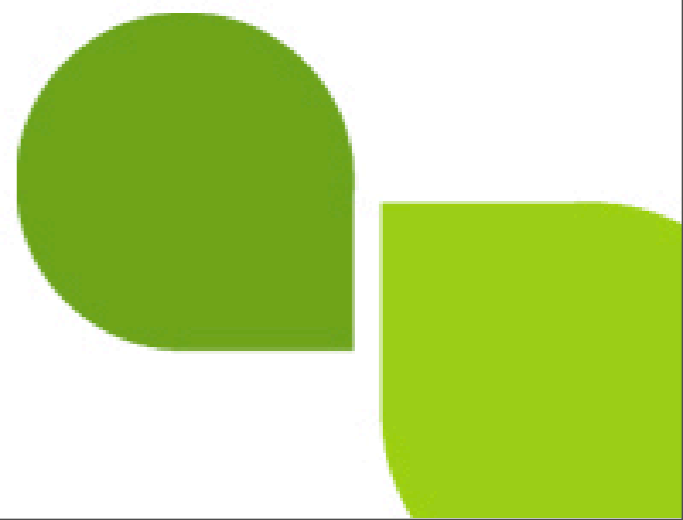




Gögn og gaman

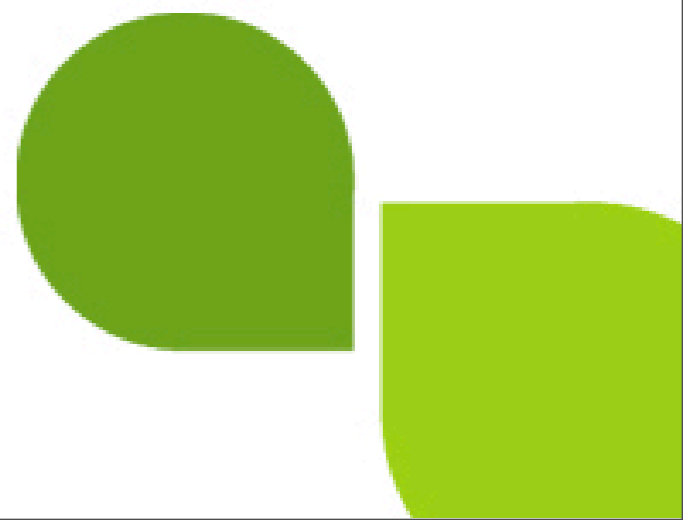
Jarðvegur nýsköpunar í tungutækni





Efnistöð

- Mikilvægi góðra gagna við þróun tungutæknilausna
- Staða og aðgengi að gögnum á Íslandi
- Opnum kisturnar





Tungutækningin og ég

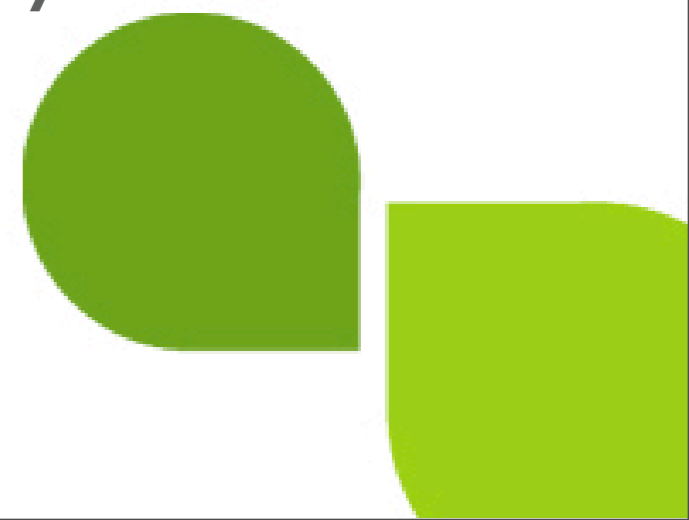
- Skrifaði beygingarforrit í menntaskóla
- Leikjapakki Lon&Don 1997
- Embla og Já.is 2005-2006
- Ýmis verkefni með KB, SH og ER





Leikjapakkin

- Stærsti vandinn að finna gott orðasafn
- 100þús orðmynda skrá á Háskólanetinu
- Vissulega takmörkuð, en hafði margt til að bera
- Hentugt form, aðgengileg, uppruninn skýr og ókeypis





Embla

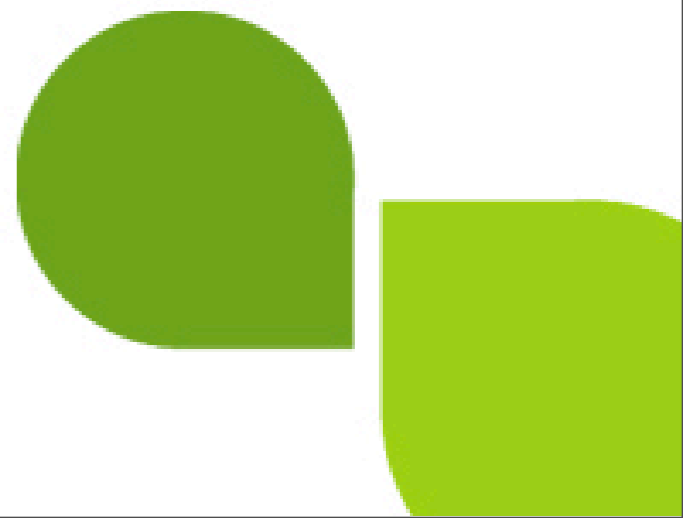
- Fyrsta leitarvélin í sögunni sem tók tillit til íslenskra beygingarmynda
- Beygingarlýsingin ómetanleg
- 2,7 milljónir ólíkra orðmynda
- Góð gögn:
 - Umfangsmikil, á hentugu formi, áreiðanleg, stöðugt viðhald, hófstíllt gjaldtaka, ágætlega aðgengileg





Leitarstarfsemi Já

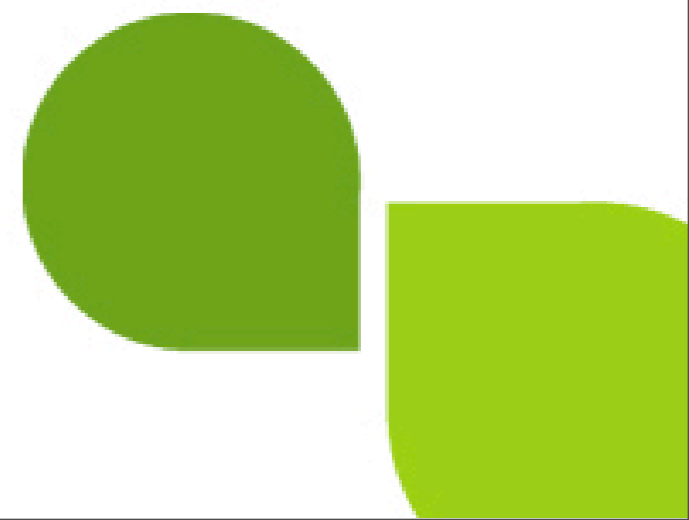
- Embla og Já.is nota einnig
 - Tíðnitöflur orðmynda (Morgunblaðið)
 - Nafnaskrár þekkra einstaklinga (Edda)
 - Bókatitla (Edda)
 - Skammstafanaskrár (heimasmíðað)
 - Tækni hugtök (mbl.is)
 - o.m.fl





Mikilvægi gagna

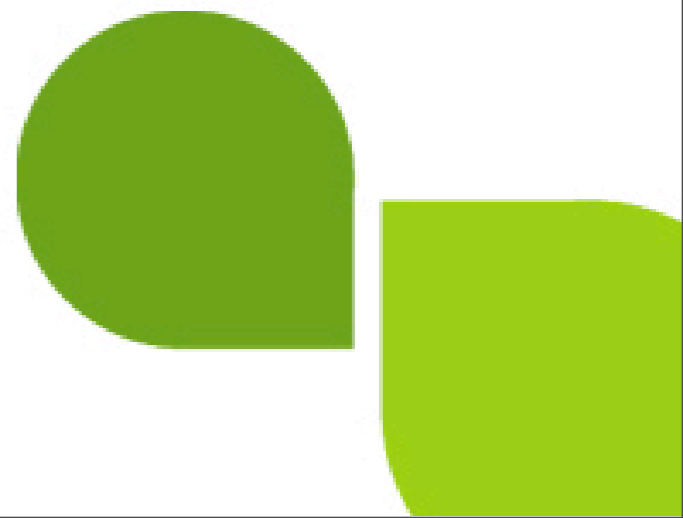
- Ekkert þessarra verkefna hefði orðið að veruleika án góðra gagna
- Unnið af litlum fyrirtækjum með lítil fjárráð en mikinn áhuga
- Leystu úr læðingi mikil verðmæti





Staða gagna

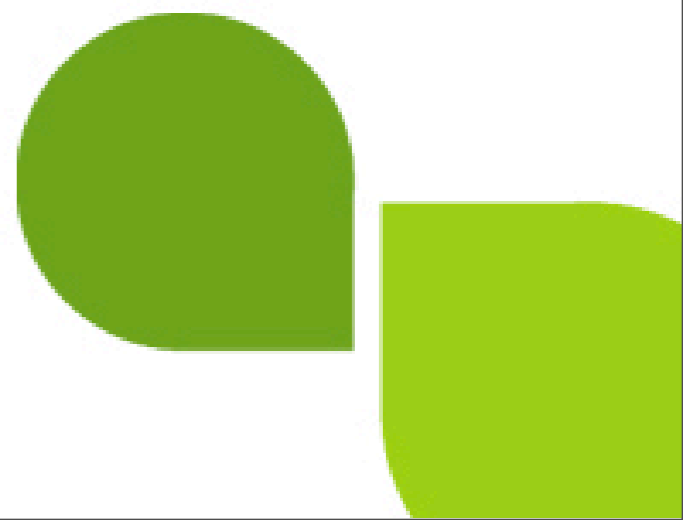
- Gríðarlega mikið til af “gagnlegum gögnum” hjá opinberum aðilum
- Tungutækni: Orðabókin (og aðrir hlutar SÁMÍF), Minningarsjóður ÞÞ, ýmis ríkisstyrkt verkefni, RÚV, o.fl.
- Önnur svið: Hagstofan, Landmælingar, Seðlabankinn, Alþingi, Háskólarnir, Þjóðskjalasafnið, Veðurstofan, RÚV, o.s.frv.





Staða gagna (2)

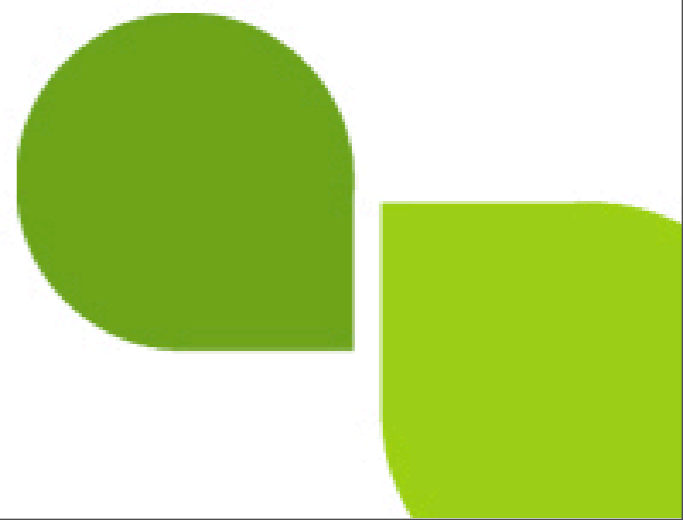
- Oft á tíðum mjög óaðgengileg
 - Ekki til á stafrænu formi
 - Leyfismál óljós
 - Erfitt að nálgast þau og finna
 - “Ormar á gulli”
 - Gjaldtaka





Jarðvegur nýsköpunar

- Góð gögn eru grundvöllur allrar nýsköpunar í tungutækni
- Mikið af þessum gögnum er til, en ýmsar hindranir fyrir því að nálgast þau





Jarðvegur nýsköpunar (2)

- Nýsköpun er í eðli sínu áhættusöm
- Allar hindranir draga úr nýsköpun, líka hjá stórum fyrirtækjum
- Stór hluti nýsköpunar fer fram af áhuga frekar en vísindalega útreiknaðri hagnaðarvon





Opnum kisturnar

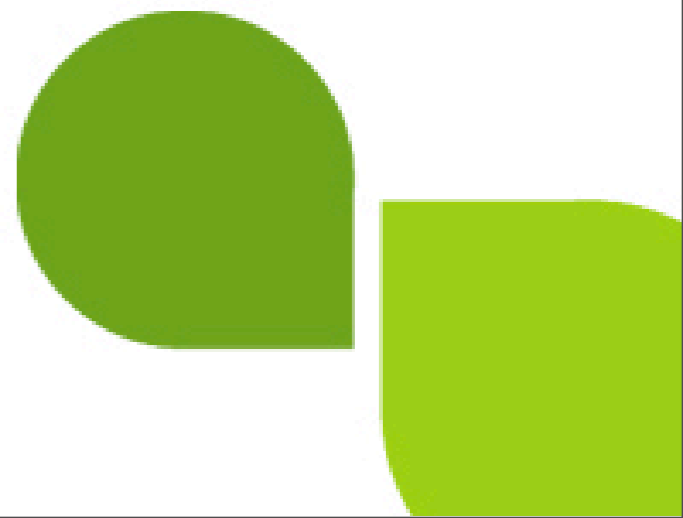
- Þjóðhagslegur ávinningur af því að tryggja opið aðgengi að opinberum gögnum er margfaldur á við mögulegar leyfistekjur og kostnað
- Ekki síst í tungutækni!
 - Næg vandamál samt





Opnum kisturnar

- Bretland: Office of Public Sector Information
 - Partur af National Archives
 - Opin Gögn
- Nýleg skýrsla: I ma punda á ári í glötuðum þjóðartekjum
 - Samsvarar 700 m.kr. á ári á Íslandi





Opin gögn skilgreind

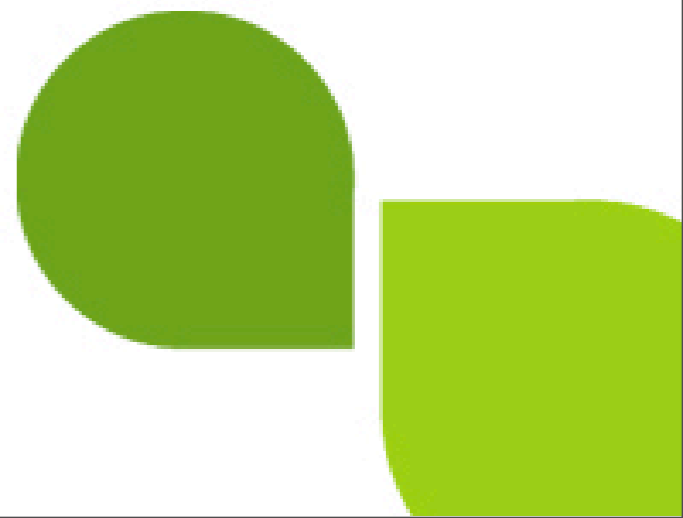
- I I atriði:
- Aðgangur (þ.m.t. verð), dreifingarréttur, notkunarréttur, tæknilegt hlutleysi, eignun, heilleiki, hlutleysi m.t.t. eðlis starfsemi, hlutleysi m.t.t. notenda, framvirkni leyfis í afleiddum verkum, óháð öðrum verkum, takmarki ekki dreifingu afleiddra verka
- <http://opendefinition.org/1.0>





Sköpum skilning

- Aukið fjármagn í verkefni viðkomandi stofnanna:
 - “Opnun” fyrirliggjandi gagna
 - Meiri skilningur fyrir nýjum verkefnum
- Aukin nýsköpun
- Nauðsynlegt til að tungutækni hér standist samanburð





Gögn og gaman

Jarðvegur nýsköpunar í tungutækni

hjalli@ja.is

