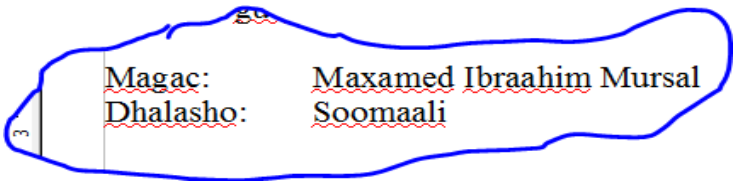
Hingaad-saxaha afka-Soomaaliga (Somali spell checker)

Waxaan ummadda Soomaaliyeed iyo dhammaan mujtamaca qora afka-Soomaaliga ugu bishaareyneynaa in uu soo baxay hingaad-saxahii (Spell checker) ugu horeeyay ee afka-Soomaaliga. Hingaad-saxahaas oo lagu isticmaalikaro barnaamijyada kumbuyuutarada sida, **OpenOffice** iyo Mozilla (sida **Firefox)**. Labada barnaamij iyo hingaad-saxahaba waa barnaamijyo ay xuquuqdooda fasaxantahay (Open Source, Copyright free). Barnaamijyada ay xuquuqdooda fasaxantahay waxaa isticmaalikaro qof walba, mana loo baahna in la bixiyo wax lacag ah.

Hordhac

Waxaa hubaal ah in qofwalba oo afka Soomaaliga wax ku qoro uu la'kulmo xaalad uu kumbuyuutarka yiraahda magacaaga/magacaada ma'saxsana ee makuu saxaa Fiiri sawirka 1aad. Sababta ugu wacana waxaytahay in kumbuyuuutarka uusan aqoonin afka aad wax ku qoreysid (waxaa ka maqan liiska erayadaada). Si baahidaas looga baxo loona helo xal wax ka tara horumarnti afka soomaaliga waxaa lagama maarmaan ah in afkeena loo helo hingaad-saxe. Waana arrinta igu dhalisay in aan bilaabo mashruucan.



Sawir 1: Tusaale erayo af-Soomaali ah oo uusan kumbuyuutarka aqoonin.

Mashruuca hingaad-saxaha wuxuu bilaawaday bishii Juun ee sankii 2007:da, wuxuuna ahaa mid howshiisa lagu qabtay hab tabarucaad ah. Waxaa mashruuca gacan ka geystay Borofeysar Kevin P. Scannell (Saint Louis University) oo ku nool dalka Mareykanka. Kaalmada uu ka geystay Bor. Kevin Scannel waxay ku timid markii aan ka war helay in aan aniga iyo asagaba wada qabnay hammi ah in aan u sameyno hingaad-saxe afafka aan ugu hadlo sida afkeena hooyo. Kevin wuxuu asalkiisa ka soo jeedaa dalka Irland, afkiisa hooyane waa Gaeilge (Irish language), anigana sida uu qof walba u qiyaasankaro wuxuu noqday afka-Soomaaliga (Af maxaatiri). Waxaa jirta in laba luqaddood ama afood ee aan kor ku soo xusnay aysan ahayn afafka adduunka ugu caansan, ayna yihiin afaf aanan in badan ka muuqannin masraxyada qoraalka. Waxaana afkeena (Soomaaliga) u raacay da' yari iyo horumar la'aan ay sii fogeysay burburka dalka ku dhacay. Waxaa jirta in aniga iyo Kevin uusan midkeena ahayn aqoonyahan afeed, ee aan ku taqasusnay cilmiga kumbuyuutarada. Waxaana rabaa in maqaalkaan uu noqdo mid aqoonyahanada Soomaaliyeed ku dhiirageliya, sidii ay ugu faa'ideynlahaayeen ummaddooda baahida u qabta aqoontooda.

Habka loo fuliyey

Mashuuca wuxuu ka bilaawdey markii aan islameeldhignay aniga iyo Kevin in dhowr kun oo eray oo afka-Soomaaliga ah loogu gudbiyo barnaamij-kubuyuutareed (Corpus), si uu tixraac ugu sameeyo bogagga Internetka ee laga helo erayada loo dhiibay. Waxaa ku xigay in si fayllal ahaan ah loo soo dejiyo bogaggaas, lana diiwaangeliyo erayada laga helay. Waxaa dhacday in guud ahaan uu barnaamijka soo helay ilaa 4 200 000 (afar malyal iyo laba boqol oo kun) oo eray. Markii baaritaan

lagu sameeyay soo noqnoqodka eraya (word frequency), waxaa soo baxday in erayga ***iyo*** uu yahay eraygii ugu badnaa inta la soo uruuriyey, waxaana ugu yaraa erayga beeraha oo soo noqnoqday ilaa iyo 150 jeer, fiiri *shaxda 1*.

iyo	ee	soo	ayaa	beeraha
68013	53604	42509	26591	150

Shax 1: Soo noqnoqodka eraydii la soo uruuriyey.

Waxaana xusid mudan markii aan bilaabay sixista erayadaas aadka u tiradabadnaa, in aan ka waayey eraydii aan islahaa waa erayo daruuri ah, sida ***caano, bariis, shaati, saxan*** IWM. Erayga ***dhacay*** wuxuu soo noqnoqday 2600 oo jeer. Arrintaas waxay tusaalo u tahay in maqaallada laga helo Internetka ee ku qoran afka-soomaaliga ay u badanyihiin kuwo wararka ka soo gudbiyo xaaladaha dagaal ee dalka Soomaaliya. Erayadii uu barnaamijkaas soo uruuriyey waxaa ka mid noqday erayo aad ugu soo noqnoqda qoraallada afka-Soomaaliga hadana aanan saxsaneyn, sida ***Mohamed*** ama ***maxamed***, ***Somalia*** ama ***soomaalia***, ***sheikh***, ***Mogadishu*** IWM. Erayada aan kor ku soo sheegnay kuwooda saxan waa: ***Maxamed, Soomaaliya, Sheekh iyo Muqdisho***.

Eraydii aan saxay ayaan markale ku celiyay mudane Kevin si barnaamijka ugu dhiibo, waxaana ka soobaxay erayo cusub. Eraydaas isku darkooda waxay noqdeen ilaa 7000 (todobo kun) oo eray kuna keydsan faylka so_SO.dic (WordList). Waxaa jirta in hingaad-saxaha asal ahaan uu u shaqeeyo sida soo socoto: Markuu arko erayga uu isticmaalaha qoray wuxu ka fiiriyaa liiska erayada uu haysto, hadii uu ka waayo wuxuu ku calaamadiyaa calaamad gaduud ah. Afka Soomaaliga waxaa ddhacda in hal eray oo asla ah uu yeesho erayo badan oo asaga ka farcama. Waxaan bilaabay in aan darso qaab dhismeedka erayadaa una sameeyo shuruuc loo isticmaalo. Barnaamijka ***myspell*** ma ***aspell*** waxay isticmaali karaan faylka shuruucda/sharciyada ku lamaanan so_SO.dic oo lagu magacaabo so_SO.aff (Affix-file). Fiiri *shaxda 2* oo ka muuqata sida faylka sharciyada uu u shaqeeyo.

Asal	**magac**	**aniga**	**adiga**	**asaga**	**ayada**	**anaa**	**idinka**	**ayaga**
nadiifi	nadiifis	nadiifiyay	nadiifisay	nadiifiyay	nadiifisay	nadiifinay	nadiifiseen	nadiifiyeen
sharci	s	yay	say	yay	say	nay	seen	yeen

Shax 2: Sharciyada (affix) iyo falka nadiifi oo loo helay ilaa iyo sideed eray oo cusub.

Waxaa ku filan in aan falyka so_SO.dic ku sheego in erayga *nadiifi* uu qaadankaro shaciga ku jira so_SO.aff sidaasna uu ku yeesho erayga *nadiifi* erayo cusub.

Natiijada iyo Gabagabo

Mashruucan waxaa ka dhashay hingaad-saxe kooban oo u gaar ah afka-Soomaaliga si uu u noqdo mid dhammaystiran wuxuu u baahanyahay in uu helo erayo ka badan kuwa hadda ku jira, lana helo bulsho ka qeybqaadato ururinta iyo sixista erayadaas.

Hingaad-saxahan waa caddadkii ugu horeeyay, waana hingaad-saxe aysan dhankiisa erayada dhameytirneyn. Sababaha aan u soo bandhignay asaga oo aanan dhamestirneyn ayaa waxay ahayd in ay bulshada aragto kana geysato taageerada ay awoodo. Wuxuu ka koobanyahay erayo tiro yar waxaana soo bixidoona caddadyo cusub mustaqbalka haduu Alla idmo.

Wixii dheeraad ah waxaad kala soconkartaa bogga www.opensourcesomalia.org . Waxaana jira fidiyow

(Video) ku saabsan sida loo keydsankaro hingaad-saxaha.

Qore: Inj. Maxamed Ibraahim Mursal