



# VIHAPUHEEN HYÖDYNTÄMINEN TEKOÄLYN SEURANNASSA

HOMMAFORUMIN MODERAATIO

*N. N.*

Kirjoitettu  
Oikeusministeriön, Utopia Analyticsin, Helsingin yliopiston ja SUPOn tarkkailussa

Kiitokset: Johan Gadolin (yttriumin löytämisestä)

**HOMMA**  
[www.hommaforum.org](http://www.hommaforum.org)

22. heinäkuuta 2021

# Sisältö

<b>1</b>	<b>Abstrakti</b>	<b>3</b>
<b>2</b>	<b>Hommaforumista</b>	<b>4</b>
<b>3</b>	<b>Muita huomioita</b>	<b>6</b>
<b>4</b>	<b>Ylilaudan ja Hommaforumin aineistosta</b>	<b>7</b>
<b>5</b>	<b>Johtopäätökset</b>	<b>8</b>
<b>6</b>	<b>Liite A</b>	<b>9</b>

TÄMÄ teos on vastine Utopia Analyticsin kokoamaan ja Oikeusministeriön julkaisemaan raporttiin (oik. läpyskä) *Tekoälyn hyödyntäminen vihapuheen seurannassa*. Läpyskässä esitettiin hämmentäviä väitteitä Hommaforumista. Läpyskässä esitelty metodologia ainakin Hommaforumia käsittelevin osin on selkeästi virheellinen. Tässä vastineessa esittelemme merkittävimmät Hommaforumia koskevat virheet sekä esittelemme muutamia yksinkertaisia menetelmiä sellaisten virheiden välttämiseen. Esitämme myös Ylilautaa koskevista kohdista pari huomiota, mutta muihin Läpyskässä mainittuihin keskustelualustoihin emme ota kantaa *robots.txt:tä* lukuunottamatta.

Tätä teosta ei ole tuotettu osana Tiedolla vihaa vastaan -hanketta (Facts against Hate), eikä siten ole saanut rahoitusta Euroopan Unionin perusoikeus, tasa-arvo- ja kansalaisuusohjelmasta (2014–2020). Itseasiassa, tämä teos ei ole saanut tukea mistään. Tiedolla vihaa vastaan -hankkeen kokonaisbudjetti oli 575 846 €, josta Oikeusministeriön rahoitusosuus oli 265 721 €. Utopia Analyticsin tuottamaan raportin kustannus oli 28 000 € (+alv).

Läpyskässä todetaan ”Tämä raportti ei ota kantaa siihen, miten jonkun muun tahon tulisi vihapuhemäärittelmä tehdä, vaan toivottaa jatkokeskustelun aiheesta tervetulleeksi ja pitää jatkotutkimusta tärkeänä.” Siispä päätimme tehdä vastineen, jossa pohditaan läpyskää Hommaforumin näkökulmasta.

## Varoitukset

Tämän vastineen lukeminen voi aiheuttaa voimakasta hikoilua.



## Jokaisella teolla on tarkoitus

**H**OMMAFORUM (jäljempänä myös Homma tai foorumi) perustettiin vuonna 2008 Jussi Halla-ahon blogin, Scriptan, vieraskirjan jatkoksi. Vieraskirjaa oli käytetty jo muutaman vuoden ajan poliittiseen keskusteluun pääosin niistä teemoista joista Halla-aho blogissaan kirjoitteli. Alusta oli kuitenkin hankala laajamittaiselle keskustelulle ja siksi keskustelu siirrettiin Hommalle. Homma on vuosien saatossa kasvanut yli 11 000 jäsenen yhteisöksi, ja viestejä on kertynyt reilusti yli 3 miljoonaa.

Hommaforum on juuristaan huolimatta poliittisesti sitoutumaton. Se näkyy siinä, että kaikkia kirjoittajia moderoidaan samoilla kriteereillä. Hommaforumilla on synnytetty vähintään yksi puolue, MUUTOS 2011 ja foorumia on syytetty toisen puolueen kaappaamisesta. Foorumilla on vuosien saatossa haasteltu tai vaalitentattu useiden eri puolueiden näkyviä hahmoja, esimerkiksi Kai Pöntinen (kok), Sampo Terho (PS, myöh. sin), Anna Mikkola (vas), Sari Essayah (kd), Jyrki Kasvi (vih).

Foorumi on luonteeltaan kansallismielinen ja monet sen jäsenet suhtautuvat kriittisesti esimerkiksi holtittomaan taloudenpitoon ja globalismiin. Mutta foorumilla kirjoitellaan toki paljon muustakin, kuten tieteistä ja taiteista. Vuonna 2020 ja 2021 foorumilla on keskusteltu paljon myös SARS-CoV-2-viruksen aiheuttamasta sairaudesta ja sen lieveilmiöistä. Foorumilla on vuosien saatossa annettu ilmaista tilaa vaalimainoksille useissa eri vaaleissa sekä mainostettu mielenkiintoisia kansalaisaloitteita.

Kaikki eivät rakasta Hommaforumia. Varsinkin poliittisen kentän vasemmalta laidalta on kantautunut hyvinkin kuvailevia mainintoja Hommasta, hommalaisista ja siitä mitä heille pitäisi tehdä. Foorumin tarkoituksena on kuitenkin mahdollistaa keskustelu sellaisista maahanmuuttoon liittyvistä aiheista ja ilmiöistä, joista vallan sylikoiraksi alentunut media tai lainsäädäntö- tai toimeenpanovalta ei pienin poikkeuksin halua keskustella.

Hommaforum ei ole nettijätti, eikä Homma saa tukea mistään, joten palvelinkapasiteettia ei ole rajattomasti käytössä. Poliittisen luonteensa vuoksi Homman toiminta on kärsinyt hajautetuista palvelunestohyökkäyksistä (eng. *DDOS, Distributed Denial of Service*). Hyökkäyksiltä suojautumisen vuoksi foorumilla on käytössä Cloudflaren palveluita, joiden avulla pyrimme tarjoamaan normaaleille käyttäjille parasta mahdollista palvelua. Sen sijaan Hommaforumin

intresseihin ei varsinaisesti kuulu tarjota rajatonta ja reaaliaikaista pääsyä epämääräisille tahoille.

## Forumien säännöistä

### 4§4

4 § LOUKKAAVA, PAHENNUSTA HERÄTTÄVÄ JA SOPI-MATON MATERIAALI

Pahennusta tai sopimattomia mielleyhtymiä herättävän sekä loukkaavan materiaalin julkaisu on kielletty forumin kaikissa osissa. Kaikki sellainen materiaali, joka saattaa antaa satunnaiselle vierailijalle oleellisesti väärän kuvan forumista on kielletty. Tällaista materiaalia saattaa olla, muttei rajoitu totalitarismin ja ääri-ideologioiden ihailu, kannatus tai tukeminen. Kiellettyjä ovat myös yksilöitä tai ryhmiä vastaan suunnatut törkeät ja halventavat loukkaussanat.

Forumia ei tule käyttää ääri-ideologioiden positiivisten puolten tai niiden alkuperäisen positiivisen luonteen esilletuontiin eikä mainittujen ideologioiden isien tahi implementoijien syntymäpäivien juhlintaan.

## Aiempiä tutkimuksia Hommaforumista

### Vänni, Leena (2009)

*Yhteisöllisyyden ilmeneminen verkkokeskustelussa. Tarkastelussa yhteiskuntakriittisen keskustelufoorumin sanasto.* Vaasan yliopisto

Tutkimuksessa käsitellään Homman yhteisöllisyyttä kielitieteellisin menetelmin. Lainaus johtopäätöksistä:

*”Oletin, että yhteisöllisyys keskustelufoorumilla heijastuu esimerkiksi sanaston erikoistumisena ja että keskusteluissa on löydettävissä yhteisölle vastakkaista toiseutta. Nämä molemmat oletukset pitivät paikkansa. Hommaforumin keskustelu on verkon keskustelufoorumeiden mittapuulla laadukasta, sillä keskustelussa käytetään pitkälti kirjakieltä, mielipiteitä usein perustellaan ja mielipiteiden tueksi annetaan usein linkkejä erilaisiin tutkimuksiin tai alkuperäisiin lähteisiin. Linkkien määrä saattaa kertoa jotain enemmän verkon luonteesta keskustelupaikkana. Hommaforumin keskustelun linkkien määrää voisi*

---

vertailla muihin keskustelufoorumeihin ja tutkia, onko linkkien käyttö yhtä yleistä tai peräti yleisempää muissa verkkokeskusteluissa. Keskustelun taso, tyyli ja kieli heijastavat keskustelun aihetta ja keskustelijoita, sillä yhteiskunnallisista asioista keskustelevat ovat usein muun muassa keskimääräistä korkeammin koulutettuja. Jokin muu aihe olisi voinut saada varsin erilaisen yhteisön ympärilleen.”

### **Ylä-Anttila, Tuukka (2017)**

*The Populist Toolkit: Finnish Populism in Action 2007–2016* Helsingin yliopisto

Ylä-Anttila on tutkinut ns. populismin työkalupakkia. Aineistona hän on käyttänyt Perussuomalaisten julkaisija, vaalikonevastauksia, MV-lehteä ja Hommaforumia. Tutkimusmetodeina on käytetty kvalitatiivista tekstianalyysiä. Tässä Koneen säätöön rahoittamassa tutkimuksessa Ylä-Anttila sitoo Perussuomalaiset vahvasti populismiin, *vastatietoon* ja *vasta-asiantuntijuuksiin*. Mikäli Ylä-Anttila on Hommaforumilla vihapuhetta havainnut, sitä hän ei tutkimuksessaan mainitse.

### **Malkki, Leena; Sallamaa, Daniel; Saarinen, Juha; Eerola, Sami (2021)**

*Ekstremistinen puhe verkossa ja uutismediassa* Valtioneuvoston kanslia

Paperin ensimmäinen maininta Hommaforumista kuuluu näin: ”Äärioikeistolaista ja maahanmuuttovastaista keskustelua koskeva aineisto on kerätty Hommaforumilta, Ylilaudalta ja Pohjoismaisen Vastarintaliikkeen Suomen-osaston verkkosivuilta. Hommaforum on Suomen vanhin edelleen toimiva maahanmuuttovastainen verkkofoorumi ja sillä on keskeinen rooli äärioikeistolaisten ja maahanmuuttovastaisten verkkokeskusteluiden alustana.”

Mielestäni rinnastus Pohjoismaiseen Vastarintaliikeseen ei ole kovin onnistunut lähtökohta paperille. On siinä kuitenkin joitakin oikeaan osuneita havainnoja:

”Väkivaltaa oikeuttavaa ja siihen kannustavaa aineistoa on poistettu suhteellisen aktiivisesti verkosta viime vuosien aikana. Tutkittuja alustoja moderoidaan myös tällaisen puheen osalta vaihtelevissa määrin. Siinä missä moderointi on varsin aktiivista esimerkiksi Hommaforumilla, Ylilaudalla se vaikuttaisi olevan selvästi sallivampaa. Myös Ylilau-

dalta kuitenkin poistetaan toistuvasti aineistoa muun muassa väkivallan ihannoinnin vuoksi.”

Paperi kuitenkin menee metsään ihan reilusti joissakin muissa kohdin: ”Hommaforumilla on vastajihadististen näkemysten lisäksi havaittavissa myös runsaasti etnonationalistista puhetta ja argumentointia. Keskustelupalstalla syksyllä 2019 tehdyn kyselyn mukaan noin puolet sen käyttäjistä kannattaakin aatesuuntaa.” Tässä yhteydessä voidaan todeta, että ääniä on (kirjoituspäivään) mennessä kertynyt 380, eli reilut 3 % foorumin käyttäjistä on äänestänyt, eikä tulos ole tilastotieteellisesti pätevä.

### **Keränen, Pyry (2020)**

*”Huutista v\*tun pelle” – Poliittinen keskustelu ja käytönhallinta suomalaisilla verkkofoorumeilla* Tampereen yliopisto

Tässä opinnäytetyössä on vertailtu kirjoitustyynejä ja käyttäjähallintaa Suomi24.fi:ssä, Ylilaudalla ja Hommaforumilla. Työ on mielenkiintoinen, joskaan sen sisällöllä ei ole paljoa yhtymäkohtia läpyskään. Muutamia huomioita kuitenkin. Hommaforumin moderointi on vertailun foorumeista selkeimmin havaittavissa. Mikäli Keränen on Hommalla vihapuhetta havainnut, sitä hän ei opinnäytetyössään mainitse.

**T**IETOSUOJAVALTUUTETTU Reijo Aarnio kirjoitti ansiokkaasti [blogiinsa](#) näin:

”Tietosuoja-asetuksessa säädetään automatisoituista päätöksentekojärjestelmistä ja profiloinnista eli sellaisesta henkilötietojen automaattisesta käsittelystä, jossa arvioidaan ihmisen henkilökohtaisia ominaisuuksia. Ne ovat käsitteinä kumpikin omia asioitaan, vaikka ne menevät osin päällekkäin; usein automatisoitu päätös pitää sisällään profiloinnin. Korostan kuitenkin, että tietosuoja-asetusta sovelletaan myös silloin, vaikka profiloinnin myötä ei tehtäisi asetuksessa tarkoitettuja automatisoituja päätöksiä. Profilointi on yhtä kaikki henkilötietojen käsittelyä, jonka tulee tapahtua asetuksen edellyttämällä tavalla ja perusteilla sekä aina rekisteröidyn oikeudet huomioiden.

...

Tietosuoja-asetus GDPR tuo itse asiassa aika paljon syötettä tekoälyyn liittyvään keskusteluun. Asetuksessa on sisäänrakennettuna henkilötietojen minimointiperiaate, läpinäkyvyys sekä rekisteröidyn oikeus haastaa automaattisessa päätöksenteossa tekoälyn tekemä päätös ja olla joutumatta tällaisen päätöksenteon tai profiloinnin kohteeksi. Tekoälyä käyttävän rekisterinpitäjän on haettava tähän pääsääntöön oikeuttava poikkeus tekoälyn käytölle tietosuoja-asetuksen 22 artiklan 2 kohdasta. Sen tulee sitoutua hyväksyttävään käyttötarkoitukseen, suojata järjestelmänsä ja kyetä osoittamaan noudattavansa tietosuojalainsäädäntöä. Erityisten tietoryhmien eli aiemmin arkaluonteisina tunnettujen henkilötietojen käsittelyä rajoitetaan luonnollisesti vieläkin tarkemmin.”

Me foorumin ylläpidon jäsenet suhtaudumme GDPR:ään ja tietoturvaan vakavasti. Suhtaudumme myös sananvapauteen hyvin vakavasti. Emme katso kovin hyvällä kehitystä, jossa poliittisen foorumin jäseniin kohdistetaan automaattisen tietojenkäsittelyn menetelmiä kuten profilointia ja päätöksentekoa. Siksi esimerkiksi anonymiyden aste on täysin foorumin käyttäjän omassa hallinnassa. Osin näistä syistä suhtaudumme Utopia Analyticsin menetelmiin melko skeptisesti.

## Median kritiikiton suhtautuminen

Ainakin seuraavat mediat ovat uutisoineet läpyskystä, ilman että niille on tullut mieleenkään kyseenalaistaa siinä esitettyjä hurjia väitteitä tai haasta-

tella Homman ylläpitoa: [MTV](#): ”Hommaforumilla vihaviestejä julkaistiin noin 1600 kappaletta. Se on vajaat yhdeksän prosenttia sivuston kaikesta sisällöstä.” [Salon seudun sanomat](#): ”Tämän lisäksi osa aineistosta jouduttiin keräämään käsin, sillä Hommaforum ja Ylilauta olivat kieltäneet hakurobottien toiminnan sivuillaan.

– Eräs kiinnostava huomio on, että tietyt sivustot, joita pidetään merkittävänä vihapuheen osalta, ovat kieltäneet automaattisen datan keräyksen, sanoo toimitusjohtaja Mari-Sanna Paukkeri selvityksen toteuttaneesta Utopia Analyticsistä.”

Emme toki halua riistää toimituksilta päätäntävaltaa artikkeleiden tai uutisten sisällöstä, mutta haluaisimme muistuttaa JSN:n ohjeiden kohdista 8, 10, 12 ja 20.

Läpyskästä on päätyntä virheellinen tieto myös Wikipedia-artikkeleihin [Vihapuhe](#) ja [Homma](#).

”Oikeusministeriön raportin mukaan 97% suomenkielisen Internetin vihapuheesta esiintyy erilaisilla keskustelupalstoilla[14]. Eniten vihapuhetta julkaiseva keskustelupalsta on Ylilauta, jossa tuotetaan 96% netin kaikesta vihapuheesta[14]. Selvästi pienempiä määriä vihapuhetta tuotettiin [suomi24.fi:ssä](#), [Hommaforumissa](#) ja [vauva.fi:ssä](#)[14]. Hommaforum kuitenkin pitää kärkipaikkaa vihapuheen pitoisuudessa, sillä sen viesteistä 8,7% oli vihapuhetta[14]. Toisella sijalla oli Ylilauta, jossa vihapuhetta oli 5,9% viesteistä[14].”

## Jokainen teko on arvoitus

"Raportissa analysoidaan sosiaalisen median kommentteja julkisesta internetistä. Suurempi käytetyistä aineistoista on ostettu ulkopuoliselta palveluntarjoajalta, Mohawk Analyticsiltä, joka hakurobottien avulla seuraa suomenkielistä julkista sosiaalista mediaa, ja tallentaa viestejä suurimpien sivustojen osalta jopa muutamassa sekunnissa viestin ilmestyttyä sivustolle. Tätä aineistoa kutsutaan tässä raportissa nimellä "sosiaalisen median aineisto". Toinen, pienempi aineisto on kerätty käsin Ylilauta- ja Hommaforum-sivustoilta, joilta hakurobottien vierailut on estetty. Tätä aineistoa kutsutaan nimeltä "Ylilauta ja Hommaforum -aineisto"."

Ylilautaan emme ota kantaa, koska se ei ole sidoksissa Hommaforumiin millään tavoin. Hommaforumin osalta totean, että käytössä on *Cloudflaren* tarjoama palvelunestohyökkäysten (DDoS) torjunta, SMF:n lisäosa *Bad Behavior* (laajennoksen tarkoitus on hankaloittaa spammereiden toimintaa) ja *robots.txt* (tekstimuotoinen tiedosto, jossa on listattu sellaisia palvelimen resursseja joita bottien ei toivota indeksoivan).

"Erillisessä Ylilauta ja Hommaforum -aineistossa Ylilaudan vihapuheeksi tunnistetuissa viesteissä olivat erittäin yleisiä sanat vammainen ja huora, sekä lisäksi homo ja nekeri. Tämän lisäksi aineistossa korostui nimettyjen henkilöiden terveydentilan yksilöivä arvailu. Hommaforumin viesteissä korostui sana **nekeri**, joka tässä pienessä aineistossa esiintyi 15:ssä (43 %) vihapuheeksi tunnistetuista viesteistä."

Pyysimme sekä Oikeusministeriöltä, Utopia Analyticsin henkilöstöltä että tutkimukseen sekaantuneelta tutkijalta heidän foorumilta löytämiä vihaviestejä. Oikeusministeriön Aaltosen mukaan heillä ei ole aineistoa. Tutkija ei akateemisesta ylhäisyydestään ja ylemmydentunnossaan vaivautunut vastaamaan mitään. Utopia Analyticsin Paukkeri antoi linkin kolmeen (3) Hommaforumilla olevaan vihaviestiin.

"Hommaforumilta kerättiin 343 viestiä kategorioista Kylänraitti, Mylly ja Tupa. Ylilauta linkkaa pääsivuillaan tilastoihinsa, joista tarkasteluajavälillä syys-lokakuussa 2020 on keskimäärin noin 79 000 viestiä päivässä. Hommaforumilla seurattiin sivuston sadan uusimman viestin julkaisunopeutta ja päädyttiin keskiarvoon noin 300 viestiä päivässä tammi-kuussa 2021. Syys-lokakuussa 2020 sivuston julkaisumäärät ovat voineet olla erilaisia, mutta suuruusluokka lienee oikeansuuntainen."

Hommaforumin kategorioista Kylänraitissa,



Kuva 4.1: Kuvankaappaus etusivulta. Punainen nuoli lisäty selvennökeksi

Vuosittainen yhteenveto	Uusia aiheita	Uusia viestejä
2021	409	70 846
Helinäkuu 2021	33	5 031
Kesäkuu 2021	60	10 834
Toukokuu 2021	72	10 130
Huhtikuu 2021	59	9 674
Maaliskuu 2021	56	11 458
Helmikuu 2021	50	10 527
Tammikuu 2021	79	13 192
2020	675	120 043
Joulukuu 2020	49	10 096
Marraskuu 2020	65	11 504
Lokakuu 2020	65	10 042
Syyskuu 2020	49	8 962
2020-09-01	0	273
2020-09-02	0	227
2020-09-03	1	286
2020-09-04	1	256
2020-09-05	0	316
2020-09-06	3	266

Kuva 4.2: Kuvankaappaus tilastoista. Monien muiden tietojen lisäksi päiväkohtaiset luvut saa helposti.

Myllyssä ja Tuvassa kirjoitettiin tutkimusjakson aikana 9928 viestiä. Läpyskässä mainittu 300 viestiä päivässä on oikeansuuntainen arvio, mutta ilmeisesti *Lisää tilastoja* -linkin löytäminen oli tutkijoille liian haastavaa.

**F**OORUMIN päätarkoituksena on poliittinen keskustelu kansallismielisestä näkökulmista. Foorumilla on sallittua keskustella muistakin aiheista, ja sitä on myös käytetty hyväksi. Foorumilla on keskusteltu tieteistä, historiasta, tekniikasta, uskonnoista. Foorumilla on synnytetty vähintään yksi puolue, ja väitetyksi myös kaapattu yksi.

Foorumin tarkoituksena ei ole tarjota ilmaiseksi sisältöä kaikenmaailman virkavyösteille ja norsunluutornissa eläville hahattelijoilta. Palvelimen verkoyhteydet, sähkö ja muut tarpeelliset resurssit, ml. palvelunestohyökkäysten torjunta, eivät ole ilmaisia. Näitesti pyydettyä olisiimme saattaneet tarjota anonymisoidun tietokantadumpin tutkittavaksi.

Käytämme tästä *tutkimuksesta* sanaa läpyskä, koska se ei täytä tieteellisille papereille asetettavia laatuvaatimuksia. Toivottavasti se ei täytä Oikeusministeriön tilaamien papereiden laatuvaatimuksia.

Läpyskässä ja Paukerin haastattelussa todetaan, että Hommaforumilla ja Ylilaudalla on kielletty hakurobottien toiminta. Hommaforumilla on käytössä viestialusta SMF:n laajennos *Bad Behavior*. Sen ei pitäisi estää tutkijoita millään tavoin, koska tutkijat eivät (toivottavasti) ole kirjoittaneet foorumille mitään (eli manipuloineet foorumin sisältöä). Foorumilla käytössä oleva *robots.txt* ei oleellisesti poikkea esimerkiksi *Twitterin* tai *Facebookin* vastaavasta. Muutenkin *robots.txt*:n sisältö on deskriptiivistä, eli se kuvailee palvelimen ylläpitäjän tahtoa, käytännössä se ei estä pahantahtaisen robotin toimintaa mitenkään. Hommaforumilla on käytössä myös Cloudflaren tarjoama palvelu palvelinestohyökkäysten rajoittamiseen. Kyseinen palvelu on voinut rajoittaa tutkijoiden reaaliaikaista pääsyä foorumin sisältöön, mutta kuten sanottua sekin on vain rajoite, ei este. Liitteessä A on yksinkertainen esimerkki hakurobotista, jolla voi hakea yksittäisestä säikeestä dataa ilman että palvelin tukehtuu.

Läpyskässä käytettyjä menetelmiä Hommaforumin ja Ylilaudan osalta ei ole kuvattu kovin selkeästi. Hommaforumilta tutkijat ovat poimineet käsin 343 viestiä. ”Hommaforumilta kerättiin 343 viestiä kategoriosta *Kylänraitti, Mylly ja Tupa*.” (Läpyskä, s 22) Vihapuhettakin on tunnistettu. ”Hommaforumilla vihapuheeksi tunnistettiin 30 viestiä (8.7 %) aineiston 343 viestistä.” (Läpyskä, s 30). ”Hommaforumin viesteissä korostui sana **neekeri**, joka tässä pienessä aineistossa esiintyi 15:ssä (43 %) vihapuheeksi tunnistetuista viesteistä.” (Läpyskä, s 41). Reaalimaailmassa 15 viestiä

30:stä on 50 %.

Läpyskän tekijöillä matematiikka vaikuttaa muutenkin olevan vähän hakusessa. Tutkimme tilannetta hieman tarkemmin. Tarkasteluajanjaksolla (syys- ja lokakuu 2020) mainituissa kategorioissa julkaistiin yhteensä **9928** viestiä, joista **64** sisältää sanan **neekeri** missä tahansa muodossa tai yhdyssanan osana. Mainittakoon, että se esiintyi myös yhdyssanassa **neekerinsuukko**. Jos oletetaan, että **neekeri** esiintyy 50 %:ssa vihaviesteistä, niin vihavieheistä on läpyskässä mainittujen 1600 viestin sijasta 128, eli läpyskässä ja lehtijutuissa mainitun 8,7 %:n sijasta oikea lukema olisi 1,3 %.

Läpyskän mainitsemassa 343 viestissä siis 15:ssä oli **neekeri**. Olettakaamme viestien otanta satunnaisotannaksi, koska muutakaan ei ole kerrottu. Todennäköisyys, että sana esiintyy satunnaisesti valitussa viestissä on  $64/9928 \cdot 100\% = 0,645\%$ . Kun viestimasasta poimitaan 343 viestiä, odotusarvo sanan sisältämien viestien määrälle:  $343 \cdot 0,645\% = 2$ . Lasketaan todennäköisyys 15 osumalle 343:n viestistä, käytetään binomijakaumaa:

$$\binom{343}{15} \cdot 0,00645^{15} \cdot 0,99355^{328} = 9,993 \cdot 10^{-9}$$

Tämä tarkoittaa sitä, että koejärjestely pitäisi toistaa *sata miljoonaa* kertaa, jotta satunnaisotannalla päästäisiin yhden kerran samaan tulokseen kuin Utopia Analytics on päätenyt. Vaikuttaakin vahvasti siltä, että Hommaforumin tapauksessa tutkimusmenetelmänä on käytetty *Harrison–Stetson -menetelmää*. Kysymme nyt *tutkimuksen* (oik. läpyskän) tilaajalta eli Oikeusministeriöltä, että saitteko verorahoille haluamaanne vastinetta?



```
#!/usr/bin/perl -w
use strict;
use HTTP::Tiny;

my $base = "https://hommaforum.org/index.php/topic,112861.";
for (my $i=0; $i <= 960; $i += 30) {
    my $res = HTTP::Tiny->new->get("$base$i.html");
    print join(" ", map {/Tarkastele profiilia \S+ ([^"]+)/ ? $1 : ()}
        split /\n/, $res->{content}), "\n" if ($res->{success});
    sleep(int(rand(5)) + 10);
}
```

## Linkkien listaus

<a href="http://urn.fi/URN:ISBN:978-952-259-893-6">http://urn.fi/URN:ISBN:978-952-259-893-6</a>	s. 3
<a href="https://osuva.uwasa.fi/handle/10024/6588">https://osuva.uwasa.fi/handle/10024/6588</a>	s. 4
<a href="https://helda.helsinki.fi/handle/10138/232291">https://helda.helsinki.fi/handle/10138/232291</a>	s. 5
<a href="http://urn.fi/URN:ISBN:978-952-383-201-5">http://urn.fi/URN:ISBN:978-952-383-201-5</a>	s. 5
<a href="https://trepo.tuni.fi/bitstream/handle/10024/121324/Ker%C3%A4nenPyry.pdf">https://trepo.tuni.fi/bitstream/handle/10024/121324/Ker%C3%A4nenPyry.pdf</a>	s. 5
<a href="https://tietosuoja.fi/-/lex-tekoaly-">https://tietosuoja.fi/-/lex-tekoaly-</a>	s. 6
<a href="https://www.mtvuutiset.fi/artikkeli/raportti-lahes-kaikki-julkisen-sosiaalisen-median-vihapuheesta-loytyy-ylilaudalta/8123198">https://www.mtvuutiset.fi/artikkeli/raportti-lahes-kaikki-julkisen-sosiaalisen-median-vihapuheesta-loytyy-ylilaudalta/8123198</a>	s. 6
<a href="https://www.sss.fi/2021/04/raportti-lahes-kaikki-julkisen-sosiaalisen-median-vihapuheesta-loytyy-ylilaudalta/">https://www.sss.fi/2021/04/raportti-lahes-kaikki-julkisen-sosiaalisen-median-vihapuheesta-loytyy-ylilaudalta/</a>	s. 6
<a href="https://fi.wikipedia.org/wiki/Vihapuhe">https://fi.wikipedia.org/wiki/Vihapuhe</a>	s. 6
<a href="https://fi.wikipedia.org/wiki/Homma">https://fi.wikipedia.org/wiki/Homma</a>	s. 6
<a href="https://hommaforum.org/robots.txt">https://hommaforum.org/robots.txt</a>	s. 8
<a href="https://twitter.com/robots.txt">https://twitter.com/robots.txt</a>	s. 8
<a href="https://facebook.com/robots.txt">https://facebook.com/robots.txt</a>	s. 8