

Tekoälyllä tehokkuutta otsikointiin – Otsikkokone, uutistoimittajan työkalu

Tekoälyllä tehokkuutta otsikointiin –projektin tavoitteena oli kehittää tietointensiivinen ratkaisu, jonka avulla sisällöntuotantovaiheessa voidaan ennustaa uutisotsikon vaikuttavuutta kohdemediassa. Projektissa kehitettiin *Otsikkokone*, uutistoimittajan työkalu, jonka avulla voidaan sisällöntuotantovaiheessa arvioida otsikkoluonnoksen vaikuttavuutta artikkelin suosioon ja kiinnostavuuteen sekä optimoida uutisotsikon tehokkuus ilman ”klikkiotsikointia”.

Vaikuttava uutisotsikko

Sähköisessä mediassa uutisotsikon merkitys korostuu yhä monipuolisemman mediavirran kilpaillessa lukijan huomiosta. Kilpailusta on tullut globaalia ja uutisten lisäksi lukijoille on tarjolla miljardeittain sähköposteja, twiittejä, blogikirjoituksia, chat-kanavia, SOME-päivityksiä, online-videoita ja muuta multimediasisältöä. Lisäksi kilpailu lukijoista korostuu erityisesti paikallisissa medioissa kilpailuna kansallisten mediasisältöjen kanssa.

Haastattelimme paikallisten uutismedioiden edustajia määritelläksemme heidän kannaltaan *vaikuttavan* uutisotsikon ja selvittääksemme uutisotsikon luontiin sovellettavia kriteerejä erityisesti sähköisessä mediassa. Otsikon tavoitteena haastattelujemme perusteella on parantaa median brändi-imagoa, houkuttella lukijoita viettämään aikaa uutissivustolla sekä vierailemaan sivustolla usein ja sitouttaa lukijat maksamaan uutispalvelusta. Julkaisukanavien erot on tunnistettu arkipäivän työssä. Painetun otsikon suunnittelussa visuaalisuus on tärkeää ja otsikon rajallinen tila pitää ottaa huomioon. Verkkojulkaisussa otsikon on tärkeää olla informatiivinen, houkuttella lukijoita klikkaamaan ja saada lukijat kiinnostumaan uutispalvelusta. Mobiilijakelussa, lukijan ollessa liikkeellä uutisen lukeminen voi rajoittua pelkän otsikon lukemiseen, jolloin itse sisältö luetaan myöhemmin, jos otsikko koetaan kiinnostavaksi. Verkossa otsikkoa eivät sido samat tilarajoitukset kuin painetussa lehdessä ja otsikot ovatkin tyypillisesti pidempiä kuin painetussa lehdessä. Paikallinen uutismedia välttää lukijoiden houkuttelua ”mysteeriotsikoilla” ja ”klikkiotsikoilla”.

Kansainväliset tutkimustulokset liittyen uutisotsikon vaikuttavuuteen tukevat haastattelututkimuksemme löydöksiä. Tutkimusten perusteella otsikon tavoitteeksi on tunnistettu uutisartikkelin sisällöstä kertominen ja lukijoiden kiinnostuksen herättäminen, jotta he haluavat lukea uutisartikkelin. Näistä jälkimmäinen on erityisen merkityksellinen, kun halutaan kasvattaa maksavien lukijoiden osuutta.

Otsikon vaikuttavuuden mittaaminen

Uutisotsikon vaikuttavuuden mittaamiseksi on esitetty erilaisia vaihtoehtoja. Kirjallisuusselvityksemme perusteella vaikuttavuutta mitataan yleisimmin julkaistun uutisen keräämällä klikkimäärällä sekä tykkäysten, jakamisten ja kommenttien lukumäärällä. Näin mittaamalla ei kuitenkaan välttämättä saada kattavaa kuvaa siitä, miten lukijat oikeasti käyttäytyvät uutista lukiessaan. Esimerkiksi on huomattu, ettei uutisen jakamistiheys liity suoraan artikkelin lukemiseen (<https://time.com/12933/what-you-think-you-know-about-the-web-is-wrong/>), vaan lukijat saattavat jakaa sisältöä tutustumatta siihen tarkemmin.

Tekoälyllä tehokkuutta otsikointiin –hankkeessa valitsimme tavoitteeksemme mitata sekä otsikon

keräämien klikkien määrää että artikkelin lukemiseen käytettyä aikaa suhteessa sisällön määrään.

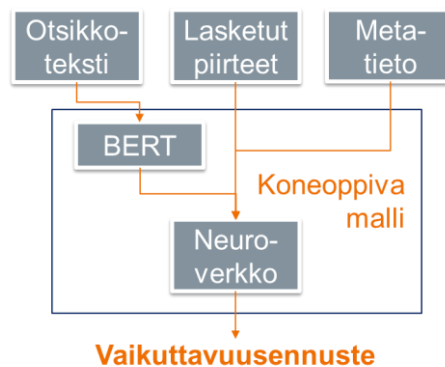
Muodostimme näin metriikan, jonka avulla uutisotsikot voidaan luokitella neljään eri luokkaan tehottomasta kiinnostavaan, houkuttelevaan ja tehokkaaseen (Kuva 1). Uutisotsikolle määriteltyjä tavoitteita ajatellen mahdollisimman suuri osa niistä pitäisi löytyä luokasta ”tehokas”.



Kuva 1 Otsikon vaikuttavuuden luokittelu

Otsikkokone ja sen toteutus

Otsikkokone-toteutuksen tavoitteena on ennustaa uutisotsikon vaikuttavuutta sisällöntuotantoprosessin tukena. Ennen sisällön julkaisemista toimittaja voi hakea ennusteita eri otsikkovaihtoehdoille ja valita koneen avustamana sopivan otsikon sisällölle. Toteutuksen pohjaksi määrittelimme mallin (Kuva 2.), jossa otsikon vaikuttavuuden ennustamiseen käytetään otsikkotekstiä, otsikosta laskettuja piirteitä (esim. otsikon sisältävien sanojen lukumäärää, erisnimien määrää, sanaluokkajakaumia) ja artikkelin julkaisuun liittyvää metatietoa (esim. julkaisu-aika ja -kanava).



Kuva 2 Otsikkokoneen malli

Otsikkokone käyttää Turun yliopiston NLP-tutkimusryhmän suomen kielelle kouluttamaa versiota BERT (Bidirectional Encoder Representations from Transformers) –neuroverkkomallista otsikkotekstin mallintamiseen. Kyseisen FinBERT-mallin todettiin antavan 6 prosenttiyksikköä tarkemman tuloksen, kun sitä toteutusvaiheessa verrattiin monikieliseen BERT-toteutukseen. FinBERT-mallin ulostulo, lasketut piirteet sekä metatiedot syötetään MLP (Multilayer Perceptron Model) -neuroverkkomalliin, jossa on yksi piilotettu kerros ja yhteensä 256 neuronaa. Malli opetetaan luokittelemaan otsikot neljään lukijoiden kiinnostusta kuvaavaan luokkaan aikaisemmin julkaistun median ja sen online-statistiikan perusteella.

Otsikkokoneen testaaminen toteutettiin yhteistyössä Kaleva Median kanssa. Kaleva Median tarjoaman anonymisoidun datan avulla päästiin opettamaan Otsikkokoneen tekoälymalli todellisella sisällöllä sekä lukijakäyttäytymisellä ja arvioimaan näin kehitetyn työkalun toimivuutta uutistoimittajan tukena. Datasetsi sisälsi uutisportaalin anonymisoidut webanalytiikkatiedot 5,5 kuukauden ajanjaksolta, koostuen noin 17 miljoonasta klikistä liittyen noin 7000 uutisartikkeliin (sisältäen tiedon julkaisuajasta, uutisotsikosta, osastosta, kirjoittajasta, maksullisuudesta, käyttäjästä). Webanalytiikkatiedon lisäksi mallin kouluttamisessa käytettiin itse artikkeliin liittyvää tietoa, kuten sanamäärä, jotta webanalytiikkatiedon pohjalta pystyttiin laskemaan artikkelikohtainen keskimääräinen luku-aika ja lukuosuus.

Otsikkokone uutistoimituksen apuna

Hankkeessa kehitetty Otsikkokone-työkalu kykenee ennustamaan uutisotsikon vaikuttavuuden yhteen neljästä luokasta 58,7% tarkkuudella. Arvioimme myös työkalun tarkkuutta verrattuna asiantuntijoiden tekemään luokitteluun. Viisi asiantuntijaa luokitteli 80 uutisotsikkoa Otsikkokoneen tapaan neljään eri vaikuttavuusluokkaan. Otsikot valittiin satunnaisesti hankkeessa käytetystä testiaineistosta.

Asiantuntijoiden luokittelutarkkuus oli 25,8% ja Otsikkokoneen näille otsikoille 61,3%. Verrattuna asiantuntija-arvioon, Otsikkokone toimii siis merkittävästi paremmalla tarkkuudella. Näin arvioimme, että tuloksista on merkittävää apua toimittajille, kun tulokset otetaan operatiiviseen käyttöön toimituksessa. Operatiivisessa käytössä on tärkeää, että malli opetetaan riittävän usein ajankohtaisella datalla, jotta se pystyy huomioimaan muuttuvia trendejä uutisoinnissa ja ennustamaan tarkasti ajankohtaisia uutisaiheita.

Otsikkokone otetaan testikäyttöön Kaleva Mediassa rajatulla määrällä toimittajia ja testin perusteella arvioidaan tarkemmin tarvittava jatkokehitys Otsikkokoneen prototyypille. Malli on sovitettavissa myös muiden mediatoimijoiden sisältöön ja ohjelmistototeutus tullaan julkaisemaan avoimen lähdekoodin ohjelmistona kevään 2020 aikana.

Lisätiedot:

Sari Järvinen

VTT

sari.jarvinen_vtt.fi (laita -viivan tilalle @-merkki)

p. 040 512 9662