

Metodologinen kehittäminen median ja viestinnän tutkimuksessa

Katariina Mäkinen katariina.makinen@tuni.fi ja Pihla Toivanen pihla.toivanen@helsinki.fi

Esitykset:

1. Tekstin erilaisten ominaispiirteiden vertailu automaattisessa tekstiluokittelussa

Ümit Bedretdin umit.bedretdin@helsinki.fi, Helsingin yliopisto, Digitaalisten ihmistieteiden osasto

Tämä työ esittelee ohjattuun koneoppimiseen perustuvan tekstiluokittelijan kehitysprosessin mediatutkimuksen näkökulmasta. Lähetymistapa mahdollistaa mediatutkijan asiantuntijatiedon valjastamisen laaja-alaiseen laskennalliseen analyysiin ja suurten aineistojen käsittelyyn.

Työssä vertailemme erilaisia tekstin ominaisuuksia, joita hyödyntämällä opetamme automaattisen luokittelijan tunnistamaan tekstien kehystystapoja ja aihepiirejä. Kehitystyössä käyttämämme aineistot on annotoitu osana kahta erilaista mediatutkimusprojektia. Näistä ensimmäisessä tutkitaan tapoja, joilla vastamedia uudelleenkäytää valtamedian artikkeleita. Siinä on aineistona 37 185 MV-lehden artikkelia, joista on eristetty kolme erilaista kehystystaktiikkaa (Toivanen et al. 2021), jotka luokittelijan on määärä tunnistaa tekstillä automaattisesti. Toisessa projektissa keskiössä on valtamedioissa käyty alkoholipoliikkaa koskeva keskustelu, jota varten kerättiin 33 902 artikkelin aineisto Ylen, Iltalehden ja STT:n uutisista (Käynnissä oleva Valla virrat -tutkimusprojekti). Luokittelijan tehtävään on tunnistaa, sisältääkö artikkeli keskustelua alkoholipoliikasta.

Erlaisten kehysanalyysien automatisointi on mahdollista, mutta luokittelijoiden tarkkuudessa on kehityksen varaa, eivätkä ne ole tarpeeksi tarkkoja korkeaa tarkkuutta vaativiin tehtäviin.

Tarkoituksesta on selvittää, mitkä tekstin ominaisuudet johtavat parhaaseen luokittelutarkkuuteen ja kuinka valita niistä kulloiseenkin tehtävään sopivimmat.

Vertailemme luokittelupiirteinä virketason kontekstuaalista tietoa, artikkelin muotoiluun liittyviä ominaisuuksia, kuten lihavointeja ja html-koodia ja aihemallin avulla tuotettuja artikkeli kohtaisia aihejakaumia. Alustavat kokeet pelkästään kontekstuaalista tietoa hyödyntävällä luokittelijalla olivat lupaavia, mutta niidenkään tarkkuus ei yltänyt tarvittavalle tasolle. Oli siis tarpeen selvittää, paraneeko luokittelijan suorituskyky yhdistelemällä eri piirteitä. Hypoteesi on uskottava, sillä esimerkiksi BERT-pohjaiset upotukset koodaavat muutaman virkkeen pituisen sekvenssin lingvististä ja jakaumallista informaatiota, kun taas aihemalli sisältää laajempaa rakenteellista informaatiota. Nämä piirteet täydentäisivät toisiaan artikkeliteason luokitustehtävässä.

Yhdistelemällä kontekstuaalista informaatiota aihemallinnukseen on saavutettu merkittäviä parannuksia erilaisissa tekstinluokittelutesteissä (Peinelt et al. 2020). Näytämme, että yhdistämällä kontekstuaiset piirteet aihemallin informaatioon päästään kokeissamme vain pieniin parannuksiin. Tästä huolimatta näemme vielä potentiaalisia kehityskohteita, joilla voitaisiin päästä parempaan luokittelutarkkuuteen.

2. Estimating and measuring East African journalists' digital skills to understand the causes and impacts of the digital divide

Felippe Constancio fconstan@ulapland.fi, University of Lapland

Relation to the working group:

The research explores an alternative methodological practice, as it uses computational methods to collect large samples of data, combining qualitative and quantitative approaches.

The proposed research aims to investigate the digital divide among English-speaking journalists in some of the East African countries (preferably Kenya, Tanzania, and Uganda) analysing data from a lifelong learning platform built cooperatively by the researcher and local journalists. Therefore, the data collection includes the first phase of qualitative data generated, through the critical communicative methodology (See: Goméz et al, 2006) approach among research and researched public, followed quantitative approach to harvest data from the platform itself.

Abstract:

East Africa region, in Sub-Saharan Africa, has a promising perspective for its technological resources and Internet diffusion along the next decades (Alper et al, 2019). Given the foreseen increase of access to the Internet, journalism tends to transform among East African countries, framing the modern journalism's paradigm in which audience participation and web search engineering demand specific professional abilities (See: Broersma, 2013; See: Bruns, 2018; See: Trilling, Tolochko & Burscher, 2017). The understanding of the local journalists' digital skills allows capturing signs of the most important elements in the journalists' professional development while new media practices come into view, and also capture prematurely the scenarios to result. Thus, the unmapped causes and unknown consequences of the digital divide among East African journalists need to be clarified to support the understanding of how the local society builds its foundations as an information society (Lupač, 2018, p. 176-177; Tsatsou, 2011, p. 323; Wessels, 2013, p. 17).

Preliminary insights:

- Methodological approaches may tend to demand more computing skills from the researchers, such as familiarity with computing languages, low coding, frameworks and libraries, data analytics;
- Datafication of methodological approaches in media studies may lead to new ethical issues.

References:

Alper C. E., Miktus M., & Clements, B. J. (2019). Bridging the mobile digital divide in sub-Saharan Africa: Costing under demographic change and urbanization. International Monetary Fund Working

Paper (in progress). Retrieved from:

<https://www.imf.org/en/Publications/WP/Issues/2019/11/15/Bridging-the-Mobile-Digital-Divide-in-Sub-Saharan-Africa-Costing-under-Demographic-Change-48793>

Broersma M. (2013). A Refractured Paradigm. Journalism, Hoaxes and the Challenge of Trust.

Chris Peters, Marcel Broersma (Eds). Rethinking Journalism. Trust and Participation in a Transformed News Landscape (pp.28-44). Routledge. Retrieved from:

https://www.researchgate.net/publication/312539393_A_Refrauctured_Paradigm_Journalism_Hoaxes_and_the_Challenge_of_Trust

Bruns A. (2018). Gatewatching and news curation. Journalism, social media, and the public sphere. Vol: 113. Steve Jones (Ed.). Series: Digital formations. New York: Peter Lang Publishing

Gómez J. , Latorre A. , Sánchez M. & Flecha R. (2006). Metodología Comunicativa Crítica. 1st edition. Barcelona, Spain: El Roure Editorial

Lupač P. (2018). Beyond the digital divide: Contextualizing the information society. 1st edition. Wagon Lane, UK: Emerald Publishing Limited

Trilling D., Tolochko P., & Burscher, B. (2017). From newsworthiness to shareworthiness: How to predict news sharing based on article characteristics. *Journalism & Mass Communication Quarterly*, Sage Journals, Vol: 94(1), (pp. 38-60). DOI: <https://doi.org/10.1177/1077699016654682>

Tsatsou, P. (2011). Digital divides revisited: what is new about divides and their research? *Media, Culture & Society*, Vol: 33 (2) (pp. 317-331). Publisher: Sage. DOI: DOI: 10.1177/0163443710393865

Wessels B. (2013). The reproduction and reconfiguration of inequality - Differentiation and class, status and power in the dynamics of digital divides. Edited by Massimo R. & Glenn W. M. (Eds). *The Digital Divide. The internet and social inequality in international perspective* (pp. 17-28). Publisher: Routledge

3. Automatic classification in qualitative research

Pihla Toivanen pihla.toivanen@helsinki.fi (HY), Antti Kanner (HY), Katariina Mäkinen (TAU)

In this presentation, we focus on methodological development with a specific case example. Our aim is to consider what kind of methodological and epistemological questions appear when qualitative research is combined with computational methods.

We introduce an automatic classification case study that combines qualitative and quantitative research. In the Flows of Power project, we study media as an arena and agent of power. Related to a case study on alcohol policy media discussion, the computational classification task was to

classify Finnish news articles into three categories of alcohol policy discussion. We recruited students who classified 1600 articles so that 1500 articles got three classifications by different students and 100 articles got ten classifications. As a typical end result of news annotation projects, the level of disagreement in the annotated data was quite high.

Our case study brings up conditions of knowledge production within social scientific research methodology, especially in relation to classifications and categories. Forming categories and doing classification are always interpretative processes in qualitative research. When classifications are done computationally, the question of how to justify them acquires a new context. How is it possible to advocate for the reliance on automatic classifications within qualitative research, and to develop sustainable and robust ways for doing them? We seek to answer these questions by reflecting our case study in relation to psychological and psycholinguistic research on classification as a cognitive process.