Kanditutkielma

Tietojenkäsittelytieteen kandiohjelma

NLP-hyökkäysten käyttökohteet

Akira Taguchi

21.2.2022

Matemaattis-luonnontieteellinen tiedekunta

Helsingin yliopisto

Yhteystiedot

PL 68 (Pietari Kalmin katu 5)
00014 Helsingin yliopisto

Sähkopostiosoite: info@cs.helsinki.fi
URL: http://www.cs.helsinki.fi/

HELSINGIN YLIOPISTO – HELSINGFORS UNIVERSITET – UNIVERSITY OF HELSINKI

Tiedekunta — Fakultet — Faculty	Koulutusohjelma — Utbildningsprogram — Study programme
Matemaattis-luonnontieteellinen tiedekunta	Tietojenkäsittelytieteen kandiohjelma

Tekijä — Författare — Author

Akira Taguchi

Työn nimi — Arbetets titel — Title

NLP-hyökkäysten käyttökohteet

Ohjaajat — Handledare — Supervisors

FT Nikolaj Tatti

Työn laji — Arbetets art — Level	Aika — Datum — Month and year	Sivumäärä — Sidoantal — Number of pages
Kanditutkielma	21.2.2022	6 sivua, 1 liitesivua

Tiivistelmä — Referat — Abstract

Write your abstract here.

In addition, make sure that all the entries in this form are completed.

Finally, specify 1–3 ACM Computing Classification System (CCS) topics, as per https://dl.acm.org/ccs. Each topic is specified with one path, as shown in the example below, and elements of the path separated with an arrow. Emphasis of each element individually can be indicated by the use of bold face for high importance or italics for intermediate level.

ACM Computing Classification System (CCS)
General and reference → Document types → Surveys and overviews
Applied computing → Document management and text processing → Document management → Text editing

Avainsanat — Nyckelord — Keywords

algorithms, data structures

Säilytyspaikka — Förvaringsställe — Where deposited

Helsingin yliopiston kirjasto

Muita tietoja — övriga uppgifter — Additional information

Sisältö

1 Sisällys

Ohjelmistojen hyökkäysrajapinta-ala kasvaa jatkuvasti. Osa haavoittuvaisuuksista korjataan heti havainnoinnin jälkeen, osa mitigoidaan ja osan vaikutusalue on manifestoituu vasta tulevaisuudessa . Luonnollisen kielen prosessointi (eng. Natural Language Processing, NLP) on osoittautunut hyväksi hyökkäysrajapinnaksi tätä teknologiaa hyödyntäviä osapuolia vastaan (Boucher et al., 2021). NLP-järjestelmät on tehty tulkitsemaan ihmisen luonnollista kieltä. Tämän kielen konekääntäminen aloitettiin jo vuonna 1949.

Tässä tutkielmassa tarkastellaan NLP-hyökkäysten käyttökohteita. Tähän kuuluu oleellisen historian esittely, hyökkäystaksonomia sekä puolustusmetodit. On tärkeää ymmärtää luonnollisen kielen prosessoinnin tarkoitus, jotta voidaan syventyä hyökkäysiä mahdollistaviin ongelmiin sekä näiden ratkaisemiseen.

2 Historia

3 Hyökkäystaksonomia

4 Puolustusmetodit

5 Yhteenveto

Lähteet

Boucher, N., Shumailov, I., Anderson, R. ja Papernot, N. (2021). *Bad Characters: Imperceptible NLP Attacks*. arXiv: 2106.09898 [cs.CL].

Liite A Sample Appendix

You can add one or more appendices to your thesis.