



## A Digital Investigator for Historical Newspapers

Tässä projektissa humanististen tieteiden tutkijat, tietojenkäsittelytieteen tutkijat ja kirjastojen asiantuntijat työskentelevät yhdessä kansalliskirjastojen (Itävalta, Ranska, Suomi) historiallisten sanomalehtien parissa.

 <https://www.univie.ac.at/newseye/>

 [www.newseye.eu](http://www.newseye.eu)

 @newseyeeu

✉ [roger.labahn@uni-rostock.de](mailto:roger.labahn@uni-rostock.de)

Département, mécanicien, tué.  
Lesall, second-maitre, blessé.  
Bémistant, matelot, blessé.

tunnistaminen ja sitä seuraava ryvästys. Tekstialueet voidaan esittää automaattisen tekstintunnistuksen tuottamien tekstirivien ryppäinä, kuten kuvassa, ja artikkelit edelleen niitten ryppäinä. Tätä varten tutkimuksemme keskittyy perinteisten klusterointialgoritmien lisäksi pääosin koneoppimiseen perustuviin algoritmeihin.



## A Digital Investigator for Historical Newspapers

Tässä projektissa humanististen tieteiden tutkijat, tietojenkäsittelytieteen tutkijat ja kirjastojen asiantuntijat työskentelevät yhdessä kansalliskirjastojen (Itävalta, Ranska, Suomi) historiallisten sanomalehtien parissa.

 <https://www.univie.ac.at/newseye/>

 [www.newseye.eu](http://www.newseye.eu)

 @newseyeeu

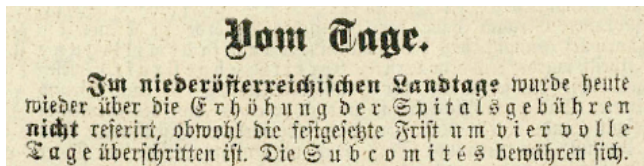


## Automaattinen tekstin- tunnistus

### Lisätietoja:

Roger Labahn

✉ roger.labahn@uni-rostock.de



### Vom Tage.

Im niederösterreichischen Landtage wurde heute wieder über die Erhöhung der Spitalsgebühren nicht referirt, obwohl die festgesetzte Frist um vier volle Tage überschritten ist. Die Subcomités bewähren sich.

*Automaattinen tekstintunnistus tekee transkription jokaisesta tekstirivistä*

Automaattinen tekstintunnistus (Automatic Text Recognition, ATR) muuttaa kuvien sisältämän kirjoituksen tekstimuotoon. Se on keskeinen vaihe NewsEyen työnkulussa, koska siinä muodostetaan tekstimuotoinen data, jota tarvitaan luonnollisen kielen käsittelyprosesseissa.

Käyttämässämme menetelmässä

malleja sovelletaan riveittäin jaettuihin kuviin, jotka rakenneanalyysialgoritmi on muodostanut. Mallit hyödyntävät sekvenssien tunnistamiseen suunniteltuja syväoppivia neuroverkkoja. Lisäksi ATR-mallien tulokset sisältävät merkkikohtaiset tunnistuksen todennäköisyydet, joita voidaan hyödyntää avainsanojen tunnistukseen perustuvissa hauissa.



## A Digital Investigator for Historical Newspapers

Tässä projektissa humanististen tieteiden tutkijat, tietojenkäsittelytieteen tutkijat ja kirjastojen asiantuntijat työskentelevät yhdessä kansalliskirjastojen (Itävalta, Ranska, Suomi) historiallisten sanomalehtien parissa.

 <https://www.univie.ac.at/newseye/>

 [www.newseye.eu](http://www.newseye.eu)

 @newseyeeu



# NewsEye Demonstrator

## Lisätietoja:

Antoine Doucet

✉ antoine.doucet@univ-lr.fr



NewsEye Demonstrator (<https://platform.newseye.eu>) esittelee projektin tuotoksia. Verkossa vapaasti käytettävissä olevalla sivustolla käyttäjät voivat tarkastella erikielisiä sanomalehtikokoelmia, jotka ulottuvat ajallisesti vuodesta 1850 vuoteen 1950. Sanomalehdet on saatu käyttöön projektiin

osallistuvilta Itävallan, Ranskan ja Suomen kansalliskirjastoilta. Tekemiensä hakujen pohjalta käyttäjät voivat muodostaa itselleen kokoelmia heitä kiinnostavista sisällöistä. Näitä kokoelmia on mahdollista tallentaa eri tiedostomuodoissa tai analysoida suoraan Demonstratorissa.



## A Digital Investigator for Historical Newspapers

Tässä projektissa humanististen tieteiden tutkijat, tietojenkäsittelytieteen tutkijat ja kirjastojen asiantuntijat työskentelevät yhdessä kansalliskirjastojen (Itävalta, Ranska, Suomi) historiallisten sanomalehtien parissa.

 <https://www.univie.ac.at/newseye/>

 [www.newseye.eu](http://www.newseye.eu)

 @newseyeeu



# Semanttinen tekstin rikastaminen

Automaattisesti  
tunnistetun tekstin

## Lisätietoja:

Antoine Doucet

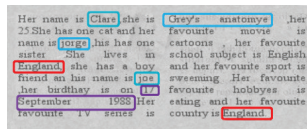
✉ antoine.doucet@univ-lr.fr

Semanttinen tekstin rikastaminen tarkoittaa asiakirjojen analysointia ja semanttisten metatietojen lisäämistä niiden sisältöön. Lukuisten semanttisten metatietojen joukosta keskityimme nimettyihin kohteisiin. Tarkemmin sanottuna pyrimme tunnistamaan nimetyt kohteet asiakirjoissa ja liittämään ne yksiselitteisesti tietämuskantaan. Kohteisiin liitetään myös tieto, miten niihin tekstissä asennoidutaan.

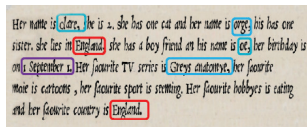
Koneellisesti tunnistetussa tekstissä voi olla runsaasti virheitä. Tämä johtuu ajan ja huonojen säilytysolosuhteiden vaikutuksesta asiakirjojen kuntoon ja/tai alkuperäisten painomateriaalien heikosta laadusta. Virheet heikentävät kaikenlaisten luonnollisen kielen rikastamisen tapojen tuottamaa laatua.

NewsEye suunnittelee uusimpiin tekniikoihin (neuroverkkoihin ja syväoppimiseen) perustuvia tekstintunnistusvirheille ja jotka ovat kielestä riippumattomia.

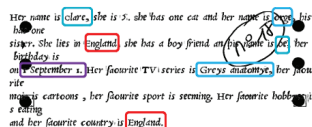
## Kohtuullinen laatu



## Keskittason laatu



## Huono laatu







## A Digital Investigator for Historical Newspapers

Tässä projektissa humanististen tieteiden tutkijat, tietojenkäsittelytieteen tutkijat ja kirjastojen asiantuntijat työskentelevät yhdessä kansalliskirjastojen (Itävalta, Ranska, Suomi) historiallisten sanomalehtien parissa.

 <https://www.univie.ac.at/newseye/>

 [www.newseye.eu](http://www.newseye.eu)

 @newseyeeu

käyttää housuja) sekä naisista, jotka tunkeutuvat journalismin alueelle: naispuolisista kirjoittajista/ journalisteista. Millainen paikka heille annetaan sanomalehdissä? Millainen rooli sanomalehdillä on naisten kamppailuissa? Digitaaliset työkalut (etenkin aihemallinnus ja frekvenssi työkalut) ovat erittäin hyödyllisiä trendien ja avainhenkilöiden paljastamisessa.



## A Digital Investigator for Historical Newspapers

Tässä projektissa humanististen tieteiden tutkijat, tietojenkäsittelytieteen tutkijat ja kirjastojen asiantuntijat työskentelevät yhdessä kansalliskirjastojen (Itävalta, Ranska, Suomi) historiallisten sanomalehtien parissa.

 <https://www.univie.ac.at/newseye/>

 [www.newseye.eu](http://www.newseye.eu)

 @newseyeeu

# NEWS E E

## Median & journalismin tapaustutkimus

Toimituksen ja  
lukijoiden  
vuorovaikutus

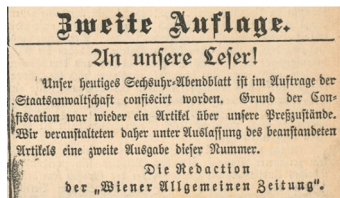
### Lisätietoja:

Eva Pfanzelter

✉ Eva.Pfanzelter@uibk.ac.at

## An die künftigen Abonnenten.

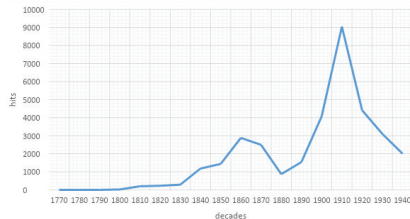
Haft du nie ein Schiff vom Stapel laufen sehen, Lieber Leser? Es ist ein interessanter Anblick, und gar manche Betrachtungen und Gedanken, heiz



Wiener Allgemeiner Zeitung, 13. marraskuuta, 1880, sivu 1, iltapäivä

Sanomalehdet muodostuivat ns. pitkän 1800-luvun aikana massamediaiksi jonkalaista ei ennen ollut olemassa. Tämä muokkasi julkisuutta, jossa määrittely kansakunnasta oli tärkeässä roolissa. Melko yleinen käsitys siitä, että uutisia välitettiin

yksisuuntaisesti toimitukselta lukijoille, ei pidä täysin paikkaansa. Lehdissä oli käytössä monia keinoja kaksisuuntaiselle kommunikaatiolle. Tämä tapaustutkimus keskittyy tähän vuorovaikutukseen.



"Toimitukselle"-fraasin ("an die Redaktion") absoluuttinen frekvenssi vuosikymmentä kohden



## A Digital Investigator for Historical Newspapers

Tässä projektissa humanististen tieteiden tutkijat, tietojenkäsittelytieteen tutkijat ja kirjastojen asiantuntijat työskentelevät yhdessä kansalliskirjastojen (Itävalta, Ranska, Suomi) historiallisten sanomalehtien parissa.

 <https://www.univie.ac.at/newseye/>

 [www.newseye.eu](http://www.newseye.eu)

 @newseyeeu

# NEWS E E

## Nationalismin tapaustutkimus

Nationalismin kieli  
1850-1950

### Lisätietoja:

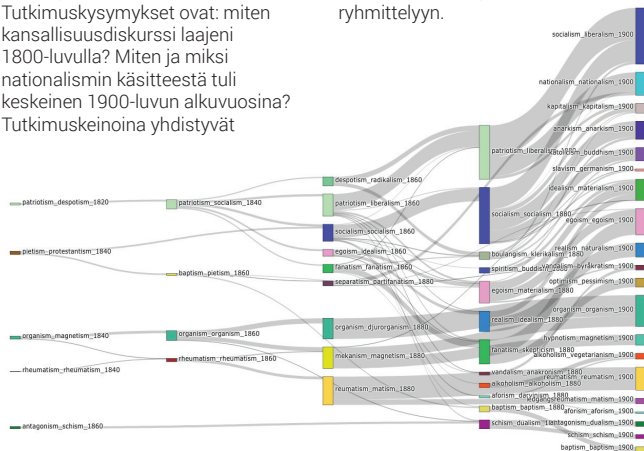
Jani Marjanen

✉ jani.marjanen@helsinki.fi

*Finska Språket, betraktadt såsom Nationalspråk.*

Tämä tapaustutkimus keskittyy kansakuntaa koskevan kielen muutokseen 1800-luvun lopulla ja 1900-luvun alussa. Tutkimuskysymykset ovat: miten kansallisuuskysymys laajeni 1800-luvulla? Miten ja miksi nationalismiin käsitteestä tuli keskeinen 1900-luvun alkuvuosina? Tutkimuskeinoina yhdistyvät

lähteiden laadullinen lukeminen ja kvantitatiivinen frekvenssianalyysi. Lisäksi tutkimuksessa käytetään sanaopetusmalleja avaintermien ryhmittelyyn.



Klusterit 1820 – 1900



## A Digital Investigator for Historical Newspapers

Tässä projektissa humanististen tieteiden tutkijat, tietojenkäsittelytieteen tutkijat ja kirjastojen asiantuntijat työskentelevät yhdessä kansalliskirjastojen (Itävalta, Ranska, Suomi) historiallisten sanomalehtien parissa.

 <https://www.univie.ac.at/newseye/>

 [www.newseye.eu](http://www.newseye.eu)

 @newseyeeu

# NEWS E E

## Siirtolaisuuden tapaustutkimus

Paluumuutto  
1850-1950

### Lisätietoja:

Sarah Oberbichler

✉ [sarah.oberbichler@uibk.ac.at](mailto:sarah.oberbichler@uibk.ac.at)

**Sie sind da, sie sind da!**

Siirtolaisuuden tapaustutkimus pyrkii paljastamaan Eurooppaan suuntautuneen paluumuuton prosesseja vuosina 1850 – 1950. Tutkimuksessa kysytään: kuinka ja missä asiayhteydessä itävaltalaiset päivälehdet tiedottivat paluumuuttajista, ja miten tiedottaminen muuttui ajan kuluessa?

Tutkimuksessa käytetään digitaalisia menetelmiä kuten aihemallinnusta ja sanaopetusmalleja. Näiden avulla luodaan osakokoelma tarkempaa laadullista analyysiä (diskurssi- ja argumentaatioanalyysiä) varten.

*Aiheiden visualisointi: relevantit ja epärelevantit artikkelit*

