




# NEWSEYE

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. Lukuisien semanttisten metatietojen joukosta keskitymme 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 menetelmiä, jotka eivät ole herkkiä tekstintunnistusvirheille ja jotka ovat kielestä riippumattomia.

### Kohtuullinen laatu

Her name is **Clare** she is **Greys' anatomy** her  
25 She has one cat and her name is **George** his favourite movie is  
name is **George** his has one cartoons , her favourite  
sister .She lives in school subject is English  
**England** she has a boy and her favourite sport is  
friend on his name is **Joe** swimming Her favourite  
her birthday is on **1st** favourite hobbies is  
**September** 1988 Her eating and her favourite  
favourite TV series is country is **England**

### Keskittason laatu

Her name is **Clare** she is a she has one cat and her name is **George** his has one  
sister .she lies in **England** she has a boy friend on his name is **Joe** her birthday is  
on **September 1st** Her favourite TV series is **Greys anatomy** her favourite  
movie is cartoons , her favourite sport is swimming Her favourite hobbies is eating  
and her favourite country is **England**

### Huono laatu

Her name is **clare** she is 25. she has one cat and her name is **George** his  
has one  
sister .she lies in **England** she has a boy friend on his name is **Joe** her  
birthday is  
on **September 1st** Her favourite TV series is **Greys anatomy** her favourite  
movie is cartoons , her favourite sport is swimming Her favourite hobbies is  
eating  
and her favourite country is **England**