




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

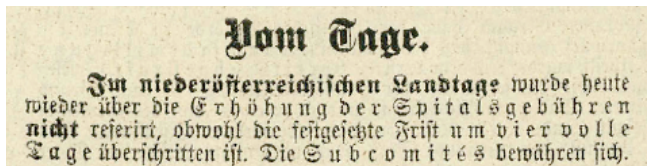
 @newseyeeu

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 tunnituksen todennäköisyydet, joita voidaan hyödyntää avainsanojen tunnistukseen perustuvissa hauissa.