



A Digital Investigator for Historical Newspapers

Dans le cadre de ce projet, des chercheurs en sciences humaines, des informaticiens et des bibliothécaires travaillent ensemble sur la presse historique conservée dans trois bibliothèques nationales (Autriche, Finlande, France).

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

 www.newseye.eu

 @newseyeeu

Segmentation d'article

✉ roger.labahn@uni-rostock.de

nous avons déjà déterminé les lignes de texte comme une étape préparatoire du traitement pour la reconnaissance automatique de texte, l'on peut décrire ces blocs de texte comme des regroupements de lignes de texte tel que l'illustre cette image. Pour atteindre ce résultat, notre recherche repose principalement sur des algorithmes basés sur l'apprentissage machine, en plus de méthodes de regroupement classiques.



A Digital Investigator for Historical Newspapers

Dans le cadre de ce projet, des chercheurs en sciences humaines, des informaticiens et des bibliothécaires travaillent ensemble sur la presse historique conservée dans trois bibliothèques nationales (Autriche, Finlande, France).

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

 www.newseye.eu

 @newseyeeu

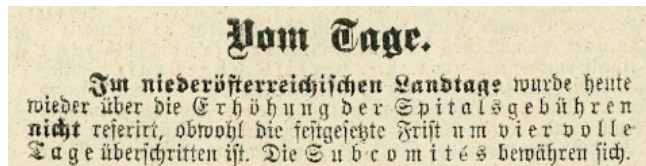


Reconnaissance automatique du texte

Contact :

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.

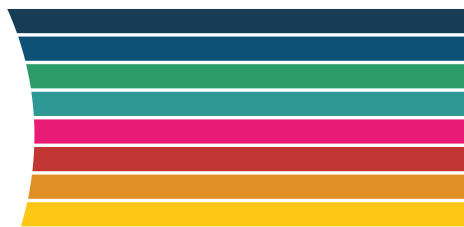
La reconnaissance automatique du texte fournit une transcription pour chaque ligne de texte.

La reconnaissance automatique du texte consiste à transcrire automatiquement les contenus textuels d'une image. C'est une étape fondamentale dans le flux de travail de NewsEye car elle fournit les données textuelles nécessaires pour être ensuite traitées dans des opérations de traitement automatique du langage naturel. Dans notre cas de figure, nous appliquons des modèles de reconnaissance à des images de

lignes de textes segmentées qui sont le produit d'un algorithme d'analyse de la mise en page. Ces modèles impliquent des réseaux de neurones artificiels profonds conçus pour des tâches d'étiquetage de séquences. De plus, les résultats de ces modèles incluent des probabilités appliquées à la reconnaissance de chaque caractère, qui peuvent être utilisées pour chercher dans les transcriptions au moyen de techniques de repérage de mots-clés (*keyword spotting*).



A Digital Investigator for Historical Newspapers



Dans le cadre de ce projet, des chercheurs en sciences humaines, des informaticiens et des bibliothécaires travaillent ensemble sur la presse historique conservée dans trois bibliothèques nationales (Autriche, Finlande, France).

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

 www.newseye.eu

 @newseyeeu

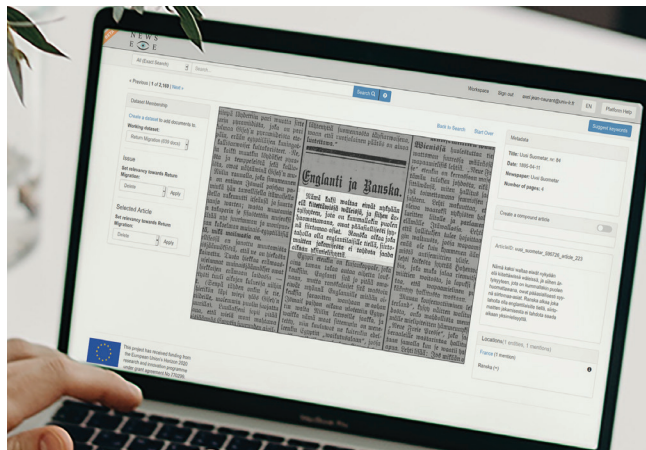


Le démonstrateur NewsEye

Contact :

Antoine Doucet

✉ antoine.doucet@univ-lr.fr



Le démonstrateur NewsEye (<https://platform.newseye.eu>) présente les résultats du projet. Ce site, dont l'accès est gratuit, permet à l'utilisateur d'explorer des collections de presse dans différentes langues. Publiés entre 1850 et 1950, ces journaux ont été mis à disposition

par les trois bibliothèques nationales basées à Vienne, Paris et Helsinki. A l'issue des recherches, l'on peut constituer des corpus personnalisés correspondant à des centres d'intérêt. Ceux-ci peuvent être exportés sous divers formats ou analysés au sein même du démonstrateur.



A Digital Investigator for Historical Newspapers

Dans le cadre de ce projet, des chercheurs en sciences humaines, des informaticiens et des bibliothécaires travaillent ensemble sur la presse historique conservée dans trois bibliothèques nationales (Autriche, Finlande, France).

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

 www.newseye.eu

 @newseyeeu



Enrichissement sémantique de contenus océrés

Contact :

Antoine Doucet

✉ antoine.doucet@univ-lr.fr

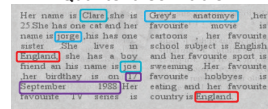
L'enrichissement sémantique de corpus textuels consiste en l'analyse de documents et l'ajout de données sémantiques à leurs contenus. Parmi une palette de données sémantiques, nous nous concentrons sur les entités nommées et, plus précisément, nous visons à reconnaître ces entités dans les documents, à les désambiguïser via une base de connaissances. De plus, les extraits de texte dans lesquels elles figurent leur sont associés.

Cependant, le texte des documents océrés contient de nombreuses erreurs, du fait du vieillissement des documents, de leurs conditions de stockage et/ou de la qualité médiocre du support d'impression originel. Ces erreurs réduisent les performances de tous les traitements ultérieurs en langage naturel (tels que la reconnaissance des noms de personnes, de lieux, d'organisations, et l'analyse d'opinion et de sentiment).

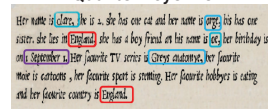
Afin de surmonter ces difficultés, NewsEye conçoit des approches basées sur les techniques les plus récentes (réseaux de neurones artificiels et apprentissage profond) résistantes aux erreurs de la reconnaissance automatique du texte et indépendantes de tout langage. Sur 13 concurrents,

les méthodes développées pendant le projet NewsEye ont atteint la première place dans 50 des 52 classements pour la reconnaissance des noms de personnes, de lieux et d'organisations en anglais, français et allemand au concours CLEF HIPE 2020.

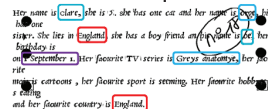
Qualité raisonnable



Qualité moyenne

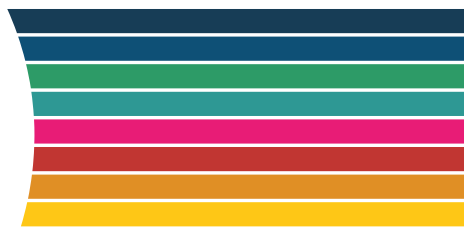


Mauvaise qualité





A Digital Investigator for
Historical Newspapers



Dans le cadre de ce projet, des chercheurs en sciences humaines, des informaticiens et des bibliothécaires travaillent ensemble sur la presse historique conservée dans trois bibliothèques nationales (Autriche, Finlande, France).

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

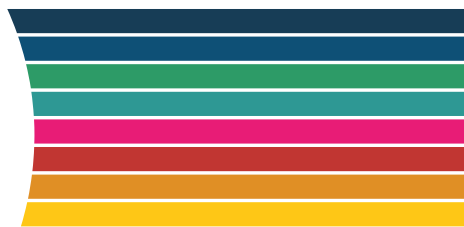
 www.newseye.eu

 @newseyeeu

pantalon) ainsi qu'aux personnalités féminines qui investissent la sphère journalistique. Quelle est la place faite aux femmes dans les journaux ? Quel est le rôle de la presse dans les luttes des femmes ? Les outils numériques (en particulier la modélisation thématique et les outils d'analyse de fréquence) sont particulièrement précieux dans ce cas d'étude pour révéler les tendances et les éléments clés.



A Digital Investigator for
Historical Newspapers



Dans le cadre de ce projet, des chercheurs en sciences humaines, des informaticiens et des bibliothécaires travaillent ensemble sur la presse historique conservée dans trois bibliothèques nationales (Autriche, Finlande, France).

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

 www.newseye.eu

 @newseyeeu

Étude de cas : médias et journalisme

L'interaction entre
la rédaction et les
lecteurs

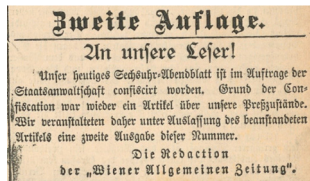
Contact :

Eva Pfanzelter

✉ Eva.Pfanzelter@uibk.ac.at

An die künftigen Abonnenten.

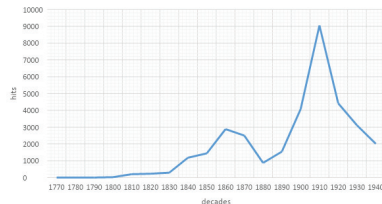
Hast du nie ein Schiff vom Stapel laufen sehen, lieber Leser? Es ist ein interessanter Anblick, und gar manche Betrachtungen und Gedanken, heiz



Wiener Allgemeine Zeitung, 13 novembre 1880, page 1, édition du soir

Pendant le « long » 19ème siècle, la presse est devenue le média de masse par excellence. Dans le même temps, un nouvel espace public dans lequel les négociations autour de l'État-nation moderne ont joué un rôle important, a vu le jour. Cependant, la croyance populaire selon

laquelle la diffusion d'informations et d'opinions dans la presse est un processus linéaire et à sens unique, où l'information n'est transmise que des éditeurs aux lecteurs, n'est pas entièrement vraie. Différents types d'interactions et de communications existaient. Cette étude de cas se concentre par conséquent sur ces interactions.



Fréquence absolue de l'expression "an die Redaktion" ("pour les éditeurs") par décennie



A Digital Investigator for Historical Newspapers

Dans le cadre de ce projet, des chercheurs en sciences humaines, des informaticiens et des bibliothécaires travaillent ensemble sur la presse historique conservée dans trois bibliothèques nationales (Autriche, Finlande, France).

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

 www.newseye.eu

 @newseyeeu

NEWS E E

Etude de cas sur le nationalisme

Le langage du nationalisme 1850-1950

Contact :

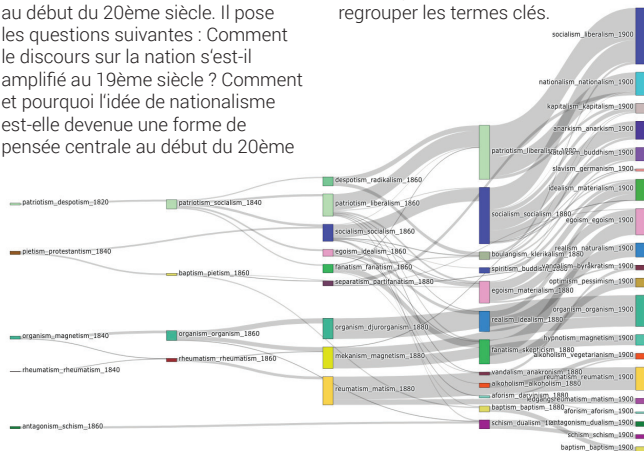
Jani Marjanen

✉ jani.marjanen@helsinki.fi

Finska Språket, betraktadt såsom Nationalspråk.

Ce cas d'étude consacré au nationalisme met l'accent sur le changement linguistique du concept de nation à la fin du 19ème siècle et au début du 20ème siècle. Il pose les questions suivantes : Comment le discours sur la nation s'est-il amplifié au 19ème siècle ? Comment et pourquoi l'idée de nationalisme est-elle devenue une forme de pensée centrale au début du 20ème

siècle ? Une lecture qualitative des sources est combinée avec une analyse de fréquence et des modèles de plongement lexical destinés à regrouper les termes clés.



Clusters 1820-1900



A Digital Investigator for Historical Newspapers

Dans le cadre de ce projet, des chercheurs en sciences humaines, des informaticiens et des bibliothécaires travaillent ensemble sur la presse historique conservée dans trois bibliothèques nationales (Autriche, Finlande, France).

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

 www.newseye.eu

 @newseyeeu

Etude de cas sur les migrations

Migrations de retour
1850-1950

Contact :

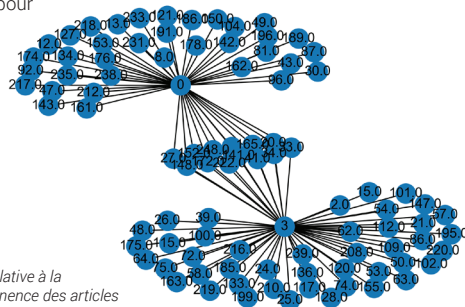
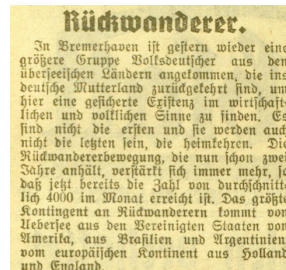
Sarah Oberbichler

✉ sarah.oberbichler@uibk.ac.at

Sie sind da, sie sind da!

Ce cas d'étude sur les migrations tente de mettre en lumière ce phénomène migratoire vers l'Europe entre 1850 et 1950. Pour ce faire, les questions suivantes sont posées : Comment et dans quel contexte les quotidiens autrichiens ont-ils évoqué les rapatriés, et comment le discours a-t-il évolué au fil du temps ?

Le cas d'étude utilise des méthodes numériques comme la modélisation thématique, le plongement lexical pour constituer une sous-collection pour des analyses qualitatives ultérieures (analyse du discours et des arguments).



Visualisation thématique relative à la
pertinence ou à la non pertinence des articles