

Le Temps Digital Archive

Maud Ehrmann, EPFL DHLAB

Swiss Open Cultural Data Hackathon

2 July 2016

2

- a daily account of the past
- the small and the great history
- people, institutions, places, events, opinions, etc.

A great asset to understand politics, society and culture from the past.

A joint effort



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

LE TEMPS



ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Vincent Barbay (DHLAB), Cyril Bornet (DHLAB), Vincent Buntinx (DHLAB), Alain Clavien (Scientific Committee),
Giovanni Colavizza (DHLAB) , Marie-Christine Doffey (BNS), Frédéric Kaplan (DHLAB), Gael Huerlimann (Le
Temps), Joëlle Kuntz (Scientific Committee), Ivo Marques (Le Temps), Enrico Natale (Scientific Committee),
Yannick Rochat (DHLAB), François Valloton (Scientific Committee), and more...

Outline

1. Context: sources and project
2. Named Entity Recognition
3. Lexical processing & Web Interface
4. Potential uses
5. Future work

Gazette de Lausanne

- Antoine Mieville
- 1st February 1798, *Peuple Vaudois*
- 3 janvier 1804, *Gazette de Lausanne*

N^o. I. Ier. Février 1798.

PEUPLE VAUDOIS. BULLETIN OFFICIEL.

L'assemblée générale provisoire des représentants du peuple Vaudois, ayant décrété l'impression du *Bulletin Officiel*, ce Journal paraîtra régulièrement tous les jours. Il contiendra le tableau exact des travaux & des décrets de l'Assemblée et de ses Comités. Il offrira celui des nouvelles politiques & militaires, qui parviendront des villes et des campagnes, & qui pourront intéresser les amis de la liberté.

Ce Journal qui paraîtra tous les jours, & qui puîfera dans les sources, aura quatre pages in-8o. même format & même caractère que le présent Numéro.

Le prix de la souscription est de L. 5 de Suisse pour 3 mois, 9 L. pour 6 mois, & 16 L. pour l'année entière, payable en souscrivant, lettres & argent franco.

Dès le moment que les Postes auront reçu une organisation régulière, on peut presque promettre de faire parvenir ce Journal *franc de port* dans le Pays-de-Vaud.

On souscrit à Lausanne, chez F. Lacombe, au Café Littéraire.

On souscrit aussi pour l'Etranger.

À PARIS, chez Barreau Libraire aux Louvres.
LYON, - - - Amable Le Roy, Libraire.
GENÈVE, - - G. J. Manget, Libraire.
NEUCHÂTEL, Fauche Borel, Libraire.
FRIBOURG, Eggendorff, Libraire.
BASLE, J. J. Fournelsen, Libraire.
BERNE, Société Typographique.
ZURICH, Orel Gefner Fueseling & Compagnie, Libraires.

Conditions de l'abonnement. — Pour les Cantons de Vaud, Fribourg & Valais, L. 10 de Suisse pour l'année, L. 6 pour six mois, & L. 4 pour trois mois. — Pour les autres cantons de la Suisse, & le Comté de Neuchâtel, L. 12 pour l'année, L. 7 pour six mois, & L. 4. 10s. pour trois mois. On s'abonne dans tous les bureaux des postes.

Chez HENRI VINCENT, (N^o. 1.) Il faut adresser son argent en lettres Imprimeur & Libraire à Lausanne. & l'argent en lettres

GAZETTE DE LAUSANNE.

Mardi 3 Janvier 1804.

(13 Nivose, an 12.)

Avis. — Pour éviter les trop fréquentes méprises faites par divers lecteurs du Bulletin Vaudois, & du Nouvelliste Vaudois, qui, par cette conformité de terminaison, attribuent souvent à l'un ce que l'autre a dit, nous nous sommes déterminés à substituer au titre de notre feuille celui de GAZETTE DE LAUSANNE, sous lequel on la verra désormais paraître.

NOUVELLES ÉTRANGÈRES

ÉTATS-UNIS D'AMÉRIQUE.

De New-York, 13 novembre.

Au tableau des ravages causés par la fièvre jaune, a succédé pour nous un spectacle presque aussi affligeant; c'est celui du retour, sur notre continent, d'un grand nombre de familles françaises de St. Domingue, qui viennent de nouveau nous demander un asyle contre le malheur. Elles paraissent bien persuadées que c'est à la nouvelle guerre rallumée par les Anglais, qu'il faut attribuer la situation actuelle des choses à St. Domingue, où leurs agens ont constamment favorisé la révolte des nègres; mais si les troupes françaises conservent, comme on l'espère, jusqu'à la paix, la clé de cette colonie, elles n'auront plus à lutter contre l'influence du climat qui les a maintenant éprouvées; & celles qui viendront les renforcer pour leur aider à reconquérir cette île, n'auront pas, sans doute, à redouter une épidémie semblable à celle qui y a régné après l'arrivée des Français, & qui, par une fatalité malheureuse, a été la plus cruelle qu'on eût jamais éprouvée dans cette colonie.

TURQUIE.

De Constantinople, le 15 novembre.

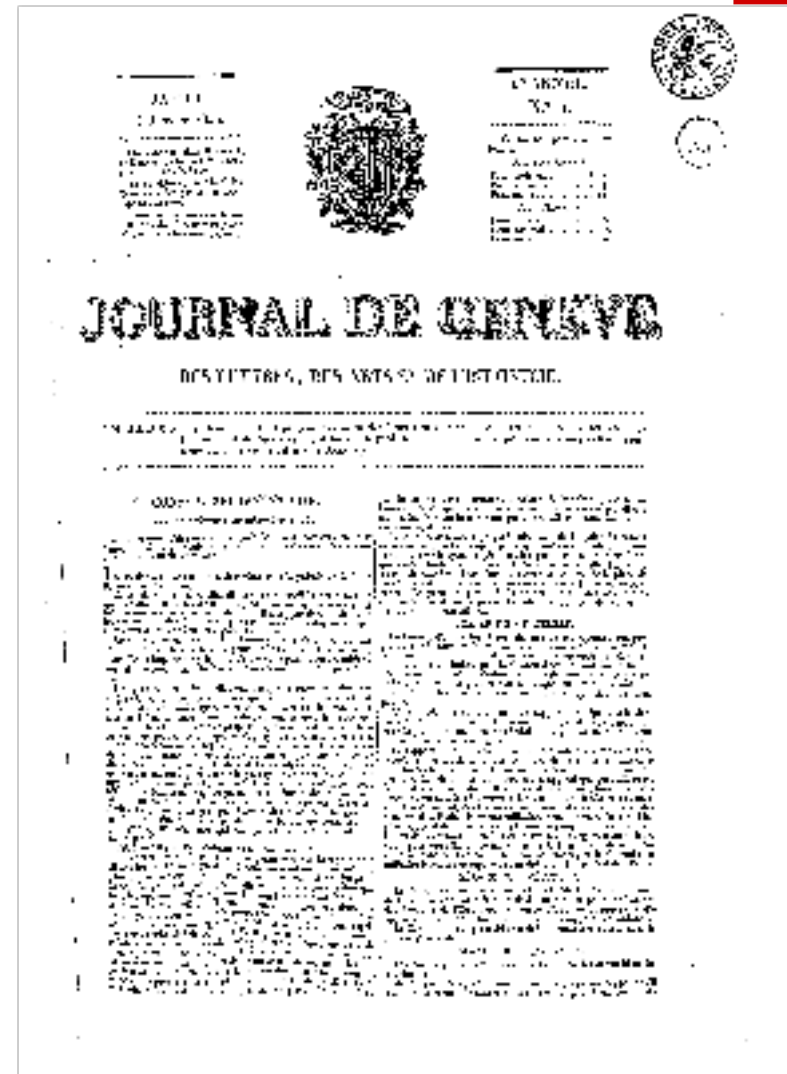
Mr. Drummond, ministre d'Angleterre près la sublime Porte, vient de recevoir de sa cour le rappel qu'il avait demandé peu de tems après son arrivée à Constantinople. Il n'attend qu'un vent favorable pour s'embarquer et retourner dans sa patrie; il prendra la route de Varna, traversera la Pologne, et passera par Berlin. On assure que la retraite de ce ministre n'a aucun motif politique, mais que le mauvais état de sa santé, et son dégoût du séjour de ce pays, et peut-être aussi celui des affaires en général, en sont l'unique cause. M. Stratton, secrétaire de légation, lui succède en qualité de ministre plénipotentiaire. C'est par la même voie qu'on a appris que S. M. Britannique a nommé un consul au Caire et un consul général à Alexandrie. Ce dernier poste a été confié au sieur Morrier, secrétaire particulier de lord Elgin, connu par quelques commissions épineuses dont il a été chargé lors du séjour des français en Egypte.

Les dernières nouvelles de cette contrée n'offrent rien de remarquable. Les négociations continuent entre le pacha du Caire retiré à Alexandrie.

Tome 4.

Journal de Genève

- James Fazy
- 5th January 1826



Fusion in 1991

- 2d September 1991,
Journal de Genève et Gazette de Lausanne

A lundi!

Lundi 2 septembre 1991 est le jour «J» pour votre quotidien: le «Journal de Genève et Gazette de Lausanne» sort dans sa nouvelle formule, en deux cahiers et avec un graphisme modernisé. Nous vous souhaitons d'ores et déjà bonne lecture et attendons avec intérêt vos réactions. A lundi!



TECHNIQUE
LAUSANNE

EDITORIAL

Pari tenu: nous voici!

Par Jasmine Audemars

Pari tenu: le «Journal de Genève et Gazette de Lausanne» se présente à ses lecteurs aujourd'hui, lundi 2 septembre 1991. Ce mariage de deux «conjoints» qui totalisent plus de 360 ans d'histoire paraît la chose la plus naturelle du monde. C'est dire l'évolution des esprits dans ce pays qui a le talent de changer tout en cultivant ses racines. Certes – des amis vaudois nous l'ont amicalement reproché – cette union ne respecte pas, dans notre nouveau titre, le principe de l'égalité. Simplement, nous avons dû trouver un compromis entre les exigences d'un graphisme moderne et celles, impitoyables, d'une concurrence croissante. Finalement, à terme, ce qui importe, c'est le fond. Et nos lecteurs vaudois savent déjà qu'ils conserveront leurs pages et que la rédaction vaudoise, à Lausanne, sous la conduite de Daniel S. Miéville, continuera à illustrer la vie du canton. Parallèlement, notre quotidien aura aussi les yeux ouverts sur une réalité régionale qui est en train de bouleverser les cloisonnements traditionnels. Avec ce premier numéro de notre nouvelle formule, nos lecteurs abonnés ou acheteurs au numéro vont découvrir aussi une présentation graphique modernisée, plus aérée et ordonnée.

Elle est destinée à mieux répondre aux vœux d'un public exigeant: un quotidien de qualité se doit d'avoir un habillage de qualité. De même, la répartition des différentes rubriques en deux cahiers séparés répond à des souhaits souvent exprimés. Surtout, les nombreux amateurs du «Samedi littéraire» retrouveront leur supplément culturel en tête de la deuxième section, chaque week-end.

Sur le plan du contenu, la première page sera désormais plus proche de l'actualité, l'éditorial conservant bien sûr sa place privilégiée. Notre quotidien affirmera ainsi toujours haut et fort sa triple vocation de journal d'information, d'analyse et d'opinion. Un sommaire développé, quant à lui, permettra au lecteur de choisir son itinéraire à l'intérieur des pages. La dernière page du «Journal de Genève et Gazette de Lausanne» quant à elle, offrira, nouveauté, un menu varié destiné au lecteur pressé, à la recherche d'une synthèse des principales informations du jour, ou au lecteur désireux de s'accorder quelques minutes «autrement».

A l'intérieur du journal, la rubrique culturelle quotidienne et les pages services – cinémas, carnet – prendront un tour résolument inter-cantonal. Chaque vendredi, un agenda culturel romand permettra au lecteur de choisir la meilleure manière possible de passer son temps libre. Et ce n'est là qu'une première étape. La nouvelle formule que nos lecteurs découvrent aujourd'hui va s'affiner et s'enrichir progressivement au fil des semaines. Nous les en informons au fur et à mesure, car nous préférons tenir, plutôt que promettre. D'emblée, toutefois, nous affirmons notre volonté de développer notre position de quotidien romand de qualité. Nous bénéficions d'acquis précieux: des lecteurs et des annonceurs fidèles, une audience reconnue et une équipe de collaborateurs hors pair dans les domaines rédactionnel, technique et commercial. Avec un tel capital et un dynamisme nouveau, nous voulons et nous allons faire encore beaucoup mieux.

Notre nouvelle formule va s'enrichir encore au fil des semaines

Le Temps

- 28th February 1998: last issue of *JDG et GDL*

- 18th March 1998 : first issue of *Le Temps* on (with *Le Nouveau Quotidien* 1991-1998)

7/29/16

Swiss Open Cultural Data Hackathon

JOURNAL DE GENEVE

et Gazette de Lausanne

Fondé en 1826 LE QUOTIDIEN SUISSE D'AUDIENCE INTERNATIONALE

LE DERNIER NÉRO

EPFL
POLYTECHNIQUE
ALE DE LAUSANNE

Samedi 28 février, dimanche 1er mars 1998
No 48 - 2 FS (ou 26) - 10 FF

Au revoir

PAR IGNACE JEANNERAT

Voici venu le jour où, pour la dernière fois, abonnés, lecteurs réguliers ou occasionnels, vos mains d'abord, puis votre regard se posent sur votre quotidien. Le titre de l'Une, l'Éditorial, les fameuses «24 heures en deux minutes», les interviews magiques d'un échange répété à des milliers de reprises.

Ne pas avoir qu'à l'instant notre gorge est nouée serait mentir. On ne refuse pas sans émotion l'héritage, la longue histoire de ce quotidien, au point de cher cette épopée de Racine dans *Bénédict* «et pour j'ai, adieu. Pour l'année de la Se, seigneur, songez en vous-même complet ce n'est peut-être affreux qu'on aime.

Où nous avons beaucoup aimé ce journal, l'esprit de ce journal, les femmes et les hommes qui l'ont fait, qui lui ont donné une chair, des moeurs, mais surtout une âme, un souffle, un sens.

Durant des décennies, le journal a considérablement baigné ce pays sur le plan politique, économique, culturel et social. Il a joué un rôle en vue sur la scène internationale. Il a été le témoin constant et souvent engagé de ce qui se passait dans le monde, à l'abri permanent du débat d'idées. Un exemple? Il a été le moniteur du philhellénisme européen, prenant parti avec passion pour la cause des Grecs. D'autres encore? Ses prises de position lors des massacres des Arméniens ou de l'affaire Dreyfus, le jugement de René Fayet au cœur de la Seconde Guerre mondiale. Et tant d'autres.

Depuis une dizaine d'années, j'ai été heureux et fier de participer à cette destinée, d'apporter ma pierre à l'ouvrage. Et, depuis le printemps 1991, d'en faire partie. C'est une âme, une flamme soumise à des courants d'air devenu parfois tourbillons. Je l'ai fait avec modestie et abnégation, après tant d'autres rédacteurs en chef qui témoignent de cette expérience humaine, intellectuelle et physiquement marquante dans notre cahier spécial.

Si ce journal cesse de paraître, aucun d'entre nous, aucun d'entre vous ne l'oubliera. Non par volonté d'entretenir une mémoire collective, mais simplement parce que ces liens ne se dissolvent pas avec cette disparition. Etendue l'émotion, s'en aller, quelque porteur d'espérance que puisse être l'histoire, serait trahir une sécheresse de cœur qui n'est pas le propre de la rédaction et de l'ensemble du personnel.

Au contraire, dans nos valises et nos cartons, nous avons jéré ce que nous aimons de ce merveilleux quotidien lui-même issu d'une fusion. Appellez-nous à l'élaboration d'un nouveau journal. Le Temps, ou s'inscrit dans la lignée du Journal de Genève et Gazette de Lausanne, nous prenons avec nous une certaine manière de travailler, de dialoguer, de regarder, de témoigner. Nous, corollaire, l'humaine, les des sexes et de l'autoritarisme, médiation de la coexistence, les visions, respect de l'autre, tout cela, nous l'emmenons au Temps qui vous sera offert le 28 mars.

Un mot encore: merci. Merci d'avoir partagé avec nous cette longue histoire.

Une journée pour en finir

Vendredi 27 février 1998: le jour où le Journal de Genève et Gazette de Lausanne a disparu comme il a vécu.

Six heures trente. Quatre ombres se dirigent vers le siège de la Société suisse de radiodiffusion. Pour le rédacteur en chef Ignace Jeannerat, le photographe Daniel Waser et les journalistes Luis Lem et Sylvain Besson, le dernier jour du Journal de Genève et Gazette de Lausanne commence par un bonjour de bonjour à mort. Avant l'aube, dans la nuit qui précède une journée qui s'annonce belle quoique franchement vers le soir.

La petite secouade est attendue pour une heure et demie en direct sur France Culture. «Comment vous sentez-vous? Est-ce dur de vivre la fin d'une institution?» Ces questions courtoises sont reçues comme autant de claque, parce qu'elles renvoient l'équipe à une réalité longtemps repoussée: le dernier jour a commencé. Lundi, le journal ne paraîtra plus.

Même impression mortifère approchant des bâtiments rue de Hesse. Trente mètres de tissu noir ont été suspendus aux balcons du deuxième étage. Et un message d'adieu met l'ensemble du Journal de Genève et Gazette de Lausanne en berne.

Le douloureux jusqu'à l'habillage de la rédaction. Tel journaliste, comme pour ses pages courtes, est en pausaison, arête. À la rubrique sport, un bon vivant est assombri dans un magnifique contact-croquant. Côté de l'équipe en culture, on l'on croise un grand jeune homme vêtu de noir qui va quitter la scène le jour même, après la dernière représentation.

Autant de personnes qui sont passées devant la barre (ou le) chapeau devant l'entrée de la rédaction. Qui ont eu un placement à la vue de Fredy Jüng, 38 ans de maison, l'une des mémoires vives du lieu, qui a dirigé l'équipe technique de nuit et porte malicieusement les cahiers de notes papiers vers la décharge.

Les cahiers, justement, sont

partout. Des piles dans l'entrée, des piles dans l'escalier. Un barrage de plastique rouge qu'il faut contourner pour gagner les étages, où l'on boit sur d'autres dalles de récipients en plastique. Comme une gasterie pour qui gagne peu à peu cet univers habitué à un vert plus à l'air.

Il faut faire vite. Avant la chute du Mur, on avait sans doute sorti du symbole. Aujourd'hui, c'est juste une manière de rappeler aux employés qu'il faut faire vite. Imprimer un dernier journal et voler les lieux. Débrancher les ordinateurs, ramasser les téléphones, débrancher les bibliothèques. Et débrancher. Car Le Temps suspend le vol du journal, et un autre rendez-vous est déjà fixé pour lundi, 9 h 30, à l'aéroport de Genève où tout doit être prêt.

Des heures, briefing et mise en train. Une cérémonie plus frénétique que l'habituel et qui donne raison à Piliotti Rist: l'émotion nous gagne, et c'est bien. Le rédacteur en chef offre son trouble en partage, des hommes et des femmes pleurent discrètement. On y lit aussi l'habituel exemplaire du *Nouveau Quotidien*, paru le matin même. On y pointe du doigt les différences d'approche. Comme chacune des rédactions termine en assistant ses collègues jusqu'au bout, elles sont criantes. Le quotidien mure a poussé sa revendication de subjectivité jusqu'à effacer presque toute actualité de sa dernière édition, pour ne parler que de la rédaction, de ses souvenirs, de son histoire.

Pas question de cela au Journal. Là, on va s'effacer une dernière fois derrière les faits, esquisser, vérifier pour raconter ce qui est passé dans le monde, presque comme si de rien n'était. On sera fidèle jusqu'au bout à la tradition rigoureuse et discrète de l'éditorial.

La tradition, justement, est en

core allongée dans une pièce du troisième étage où repose une collection complète du *Journal de Genève*. Les démanagements en cours touchent les vieilles pages reliées, dérangent la poussière, secouent les ombres. Ils achèvent de réveiller l'âme du lieu, et cette présence, ainsi libérée, s'envole doucement par la fenêtre ouverte.

Un peu plus bas dans la maison, les rendez-vous se succèdent. Photographes. Les photographes, les confrères, sans oublier la valeur des démanagements qui importent une cause après l'autre. Et, au milieu de tout cela, des journalistes, des metteurs en page, des correcteurs qui s'accrochent à leur clavier, à leur écran et à leur téléphone. Pour faire, malgré tout, un dernier journal.

Des gestes dérisoires mais qui doivent montrer qu'on ne quitte pas un travail qu'on a aimé sans lui donner le meilleur. Jusqu'au bout et sans prêter de l'avenir. C'est tellement gratuit, c'est triste et c'est beau.

Du panache. Il y a du panache dans les dernières heures de ce grand avenir qui s'efface doucement dans la nuit. «C'est une histoire d'amour», a dit Ignace Jeannerat le matin. C'est vrai. Elle s'écrit coup brutallement à un moment où la passion entre la grande dame de presse et sa jeune rédaction se perd. Il s'agit de sentiment d'admiration et d'habitude d'ont on fait, après coup, des légendes.

Difficile de l'admettre sur le moment, mais les larmes de ce jour donnent la mesure des souvenirs qui resteront une absence qui fait chaud au cœur, parce qu'elle rappelle ce qu'on peut aussi révéler ce qu'il y a de meilleur en l'homme.

Ignace Jeannerat
et ancien rédacteur du journal

SAMEDI LITTÉRAIRE

Les artistes célèbrent la fin

Écrivains et philosophes, peintres et musiciens fient la métaphore autour du beau mot fin.

Lire en pages 37 à 44

LA DER DES DER

Un cahier pour vous

Un enterrement, ça se fête aussi: le dernier banquet est servi après le Samedi Littéraire.

Lire notre supplément

Fin de partie

Lausanne avait plutôt les traits d'une adolescente prête à croquer la vie. Ses rêves et ses projets étaient portés par une ébullition jeune. C'est qui ont convoqué la camarade superflu pour lacher la vieille dame et champagne.

La Rédaction

Fin de partie

Nous semblerions devoir être corrigés de la vieille dame un peu sérieuse. Pour ceux qui l'ont fait jusqu'à ces dernières années, le Journal de Genève et Gazette de Lausanne, nous disons adieu sans cacher notre émotion à nos lecteurs et à ceux qui n'ont pas pu ou voulu participer à la suite du spectacle. Car c'est aussi grâce à vous et à nous que le Journal de Genève et la Gazette de Lausanne sont devenus la tête haute. Que coulent larmes et champagne.

La Rédaction

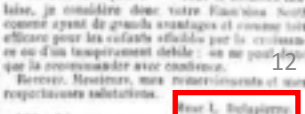
Prix du numéro: 10 francs 70 FF - 12 francs 10 FF - 15 francs 10 FF - 18 francs 10 FF - 21 francs 10 FF - 24 francs 10 FF - 27 francs 10 FF - 30 francs 10 FF - 33 francs 10 FF - 36 francs 10 FF - 39 francs 10 FF - 42 francs 10 FF - 45 francs 10 FF - 48 francs 10 FF - 51 francs 10 FF - 54 francs 10 FF - 57 francs 10 FF - 60 francs 10 FF - 63 francs 10 FF - 66 francs 10 FF - 69 francs 10 FF - 72 francs 10 FF - 75 francs 10 FF - 78 francs 10 FF - 81 francs 10 FF - 84 francs 10 FF - 87 francs 10 FF - 90 francs 10 FF - 93 francs 10 FF - 96 francs 10 FF - 99 francs 10 FF - 102 francs 10 FF - 105 francs 10 FF - 108 francs 10 FF - 111 francs 10 FF - 114 francs 10 FF - 117 francs 10 FF - 120 francs 10 FF - 123 francs 10 FF - 126 francs 10 FF - 129 francs 10 FF - 132 francs 10 FF - 135 francs 10 FF - 138 francs 10 FF - 141 francs 10 FF - 144 francs 10 FF - 147 francs 10 FF - 150 francs 10 FF - 153 francs 10 FF - 156 francs 10 FF - 159 francs 10 FF - 162 francs 10 FF - 165 francs 10 FF - 168 francs 10 FF - 171 francs 10 FF - 174 francs 10 FF - 177 francs 10 FF - 180 francs 10 FF - 183 francs 10 FF - 186 francs 10 FF - 189 francs 10 FF - 192 francs 10 FF - 195 francs 10 FF - 198 francs 10 FF - 201 francs 10 FF - 204 francs 10 FF - 207 francs 10 FF - 210 francs 10 FF - 213 francs 10 FF - 216 francs 10 FF - 219 francs 10 FF - 222 francs 10 FF - 225 francs 10 FF - 228 francs 10 FF - 231 francs 10 FF - 234 francs 10 FF - 237 francs 10 FF - 240 francs 10 FF - 243 francs 10 FF - 246 francs 10 FF - 249 francs 10 FF - 252 francs 10 FF - 255 francs 10 FF - 258 francs 10 FF - 261 francs 10 FF - 264 francs 10 FF - 267 francs 10 FF - 270 francs 10 FF - 273 francs 10 FF - 276 francs 10 FF - 279 francs 10 FF - 282 francs 10 FF - 285 francs 10 FF - 288 francs 10 FF - 291 francs 10 FF - 294 francs 10 FF - 297 francs 10 FF - 300 francs 10 FF - 303 francs 10 FF - 306 francs 10 FF - 309 francs 10 FF - 312 francs 10 FF - 315 francs 10 FF - 318 francs 10 FF - 321 francs 10 FF - 324 francs 10 FF - 327 francs 10 FF - 330 francs 10 FF - 333 francs 10 FF - 336 francs 10 FF - 339 francs 10 FF - 342 francs 10 FF - 345 francs 10 FF - 348 francs 10 FF - 351 francs 10 FF - 354 francs 10 FF - 357 francs 10 FF - 360 francs 10 FF - 363 francs 10 FF - 366 francs 10 FF - 369 francs 10 FF - 372 francs 10 FF - 375 francs 10 FF - 378 francs 10 FF - 381 francs 10 FF - 384 francs 10 FF - 387 francs 10 FF - 390 francs 10 FF - 393 francs 10 FF - 396 francs 10 FF - 399 francs 10 FF - 402 francs 10 FF - 405 francs 10 FF - 408 francs 10 FF - 411 francs 10 FF - 414 francs 10 FF - 417 francs 10 FF - 420 francs 10 FF - 423 francs 10 FF - 426 francs 10 FF - 429 francs 10 FF - 432 francs 10 FF - 435 francs 10 FF - 438 francs 10 FF - 441 francs 10 FF - 444 francs 10 FF - 447 francs 10 FF - 450 francs 10 FF - 453 francs 10 FF - 456 francs 10 FF - 459 francs 10 FF - 462 francs 10 FF - 465 francs 10 FF - 468 francs 10 FF - 471 francs 10 FF - 474 francs 10 FF - 477 francs 10 FF - 480 francs 10 FF - 483 francs 10 FF - 486 francs 10 FF - 489 francs 10 FF - 492 francs 10 FF - 495 francs 10 FF - 498 francs 10 FF - 501 francs 10 FF - 504 francs 10 FF - 507 francs 10 FF - 510 francs 10 FF - 513 francs 10 FF - 516 francs 10 FF - 519 francs 10 FF - 522 francs 10 FF - 525 francs 10 FF - 528 francs 10 FF - 531 francs 10 FF - 534 francs 10 FF - 537 francs 10 FF - 540 francs 10 FF - 543 francs 10 FF - 546 francs 10 FF - 549 francs 10 FF - 552 francs 10 FF - 555 francs 10 FF - 558 francs 10 FF - 561 francs 10 FF - 564 francs 10 FF - 567 francs 10 FF - 570 francs 10 FF - 573 francs 10 FF - 576 francs 10 FF - 579 francs 10 FF - 582 francs 10 FF - 585 francs 10 FF - 588 francs 10 FF - 591 francs 10 FF - 594 francs 10 FF - 597 francs 10 FF - 600 francs 10 FF - 603 francs 10 FF - 606 francs 10 FF - 609 francs 10 FF - 612 francs 10 FF - 615 francs 10 FF - 618 francs 10 FF - 621 francs 10 FF - 624 francs 10 FF - 627 francs 10 FF - 630 francs 10 FF - 633 francs 10 FF - 636 francs 10 FF - 639 francs 10 FF - 642 francs 10 FF - 645 francs 10 FF - 648 francs 10 FF - 651 francs 10 FF - 654 francs 10 FF - 657 francs 10 FF - 660 francs 10 FF - 663 francs 10 FF - 666 francs 10 FF - 669 francs 10 FF - 672 francs 10 FF - 675 francs 10 FF - 678 francs 10 FF - 681 francs 10 FF - 684 francs 10 FF - 687 francs 10 FF - 690 francs 10 FF - 693 francs 10 FF - 696 francs 10 FF - 699 francs 10 FF - 702 francs 10 FF - 705 francs 10 FF - 708 francs 10 FF - 711 francs 10 FF - 714 francs 10 FF - 717 francs 10 FF - 720 francs 10 FF - 723 francs 10 FF - 726 francs 10 FF - 729 francs 10 FF - 732 francs 10 FF - 735 francs 10 FF - 738 francs 10 FF - 741 francs 10 FF - 744 francs 10 FF - 747 francs 10 FF - 750 francs 10 FF - 753 francs 10 FF - 756 francs 10 FF - 759 francs 10 FF - 762 francs 10 FF - 765 francs 10 FF - 768 francs 10 FF - 771 francs 10 FF - 774 francs 10 FF - 777 francs 10 FF - 780 francs 10 FF - 783 francs 10 FF - 786 francs 10 FF - 789 francs 10 FF - 792 francs 10 FF - 795 francs 10 FF - 798 francs 10 FF - 801 francs 10 FF - 804 francs 10 FF - 807 francs 10 FF - 810 francs 10 FF - 813 francs 10 FF - 816 francs 10 FF - 819 francs 10 FF - 822 francs 10 FF - 825 francs 10 FF - 828 francs 10 FF - 831 francs 10 FF - 834 francs 10 FF - 837 francs 10 FF - 840 francs 10 FF - 843 francs 10 FF - 846 francs 10 FF - 849 francs 10 FF - 852 francs 10 FF - 855 francs 10 FF - 858 francs 10 FF - 861 francs 10 FF - 864 francs 10 FF - 867 francs 10 FF - 870 francs 10 FF - 873 francs 10 FF - 876 francs 10 FF - 879 francs 10 FF - 882 francs 10 FF - 885 francs 10 FF - 888 francs 10 FF - 891 francs 10 FF - 894 francs 10 FF - 897 francs 10 FF - 900 francs 10 FF - 903 francs 10 FF - 906 francs 10 FF - 909 francs 10 FF - 912 francs 10 FF - 915 francs 10 FF - 918 francs 10 FF - 921 francs 10 FF - 924 francs 10 FF - 927 francs 10 FF - 930 francs 10 FF - 933 francs 10 FF - 936 francs 10 FF - 939 francs 10 FF - 942 francs 10 FF - 945 francs 10 FF - 948 francs 10 FF - 951 francs 10 FF - 954 francs 10 FF - 957 francs 10 FF - 960 francs 10 FF - 963 francs 10 FF - 966 francs 10 FF - 969 francs 10 FF - 972 francs 10 FF - 975 francs 10 FF - 978 francs 10 FF - 981 francs 10 FF - 984 francs 10 FF - 987 francs 10 FF - 990 francs 10 FF - 993 francs 10 FF - 996 francs 10 FF - 999 francs 10 FF - 1002 francs 10 FF - 1005 francs 10 FF - 1008 francs 10 FF - 1011 francs 10 FF - 1014 francs 10 FF - 1017 francs 10 FF - 1020 francs 10 FF - 1023 francs 10 FF - 1026 francs 10 FF - 1029 francs 10 FF - 1032 francs 10 FF - 1035 francs 10 FF - 1038 francs 10 FF - 1041 francs 10 FF - 1044 francs 10 FF - 1047 francs 10 FF - 1050 francs 10 FF - 1053 francs 10 FF - 1056 francs 10 FF - 1059 francs 10 FF - 1062 francs 10 FF - 1065 francs 10 FF - 1068 francs 10 FF - 1071 francs 10 FF - 1074 francs 10 FF - 1077 francs 10 FF - 1080 francs 10 FF - 1083 francs 10 FF - 1086 francs 10 FF - 1089 francs 10 FF - 1092 francs 10 FF - 1095 francs 10 FF - 1098 francs 10 FF - 1101 francs 10 FF - 1104 francs 10 FF - 1107 francs 10 FF - 1110 francs 10 FF - 1113 francs 10 FF - 1116 francs 10 FF - 1119 francs 10 FF - 1122 francs 10 FF - 1125 francs 10 FF - 1128 francs 10 FF - 1131 francs 10 FF - 1134 francs 10 FF - 1137 francs 10 FF - 1140 francs 10 FF - 1143 francs 10 FF - 1146 francs 10 FF - 1149 francs 10 FF - 1152 francs 10 FF - 1155 francs 10 FF - 1158 francs 10 FF - 1161 francs 10 FF - 1164 francs 10 FF - 1167 francs 10 FF - 1170 francs 10 FF - 1173 francs 10 FF - 1176 francs 10 FF - 1179 francs 10 FF - 1182 francs 10 FF - 1185 francs 10 FF - 1188 francs 10 FF - 1191 francs 10 FF - 1194 francs 10 FF - 1197 francs 10 FF - 1200 francs 10 FF - 1203 francs 10 FF - 1206 francs 10 FF - 1209 francs 10 FF - 1212 francs 10 FF - 1215 francs 10 FF - 1218 francs 10 FF - 1221 francs 10 FF - 1224 francs 10 FF - 1227 francs 10 FF - 1230 francs 10 FF - 1233 francs 10 FF - 1236 francs 10 FF - 1239 francs 10 FF - 1242 francs 10 FF - 1245 francs 10 FF - 1248 francs 10 FF - 1251 francs 10 FF - 1254 francs 10 FF - 1257 francs 10 FF - 1260 francs 10 FF - 1263 francs 10 FF - 1266 francs 10 FF - 1269 francs 10 FF - 1272 francs 10 FF - 1275 francs 10 FF - 1278 francs 10 FF - 1281 francs 10 FF - 1284 francs 10 FF - 1287 francs 10 FF - 1290 francs 10 FF - 1293 francs 10 FF - 1296 francs 10 FF - 1299 francs 10 FF - 1302 francs 10 FF - 1305 francs 10 FF - 1308 francs 10 FF - 1311 francs 10 FF - 1314 francs 10 FF - 1317 francs 10 FF - 1320 francs 10 FF - 1323 francs 10 FF - 1326 francs 10 FF - 1329 francs 10 FF - 1332 francs 10 FF - 1335 francs 10 FF - 1338 francs 10 FF - 1341 francs 10 FF - 1344 francs 10 FF - 1347 francs 10 FF - 1350 francs 10 FF - 1353 francs 10 FF - 1356 francs 10 FF - 1359 francs 10 FF - 1362 francs 10 FF - 1365 francs 10 FF - 1368 francs 10 FF - 1371 francs 10 FF - 1374 francs 10 FF - 1377 francs 10 FF - 1380 francs 10 FF - 1383 francs 10 FF - 1386 francs 10 FF - 1389 francs 10 FF - 1392 francs 10 FF - 1395 francs 10 FF - 1398 francs 10 FF - 1401 francs 10 FF - 1404 francs 10 FF - 1407 francs 10 FF - 1410 francs 10 FF - 1413 francs 10 FF - 1416 francs 10 FF - 1419 francs 10 FF - 1422 francs 10 FF - 1425 francs 10 FF - 1428 francs 10 FF - 1431 francs 10 FF - 1434 francs 10 FF - 1437 francs 10 FF - 1440 francs 10 FF - 1443 francs 10 FF - 1446 francs 10 FF - 1449 francs 10 FF - 1452 francs 10 FF - 1455 francs 10 FF - 1458 francs 10 FF - 1461 francs 10 FF - 1464 francs 10 FF - 1467 francs 10 FF - 1470 francs 10 FF - 1473 francs 10 FF - 1476 francs 10 FF - 1479 francs 10 FF - 1482 francs 10 FF - 1485 francs 10 FF - 1488 francs 10 FF - 1491 francs 10 FF - 1494 francs 10 FF - 1497 francs 10 FF - 1500 francs 10 FF - 1503 francs 10 FF - 1506 francs 10 FF - 1509 francs 10 FF - 1512 francs 10 FF - 1515 francs 10 FF - 1518 francs 10 FF - 1521 francs 10 FF - 1524 francs 10 FF - 1527 francs 10 FF - 1530 francs 10 FF - 1533 francs 10 FF - 1536 francs 10 FF - 1539 francs 10 FF - 1542 francs 10 FF - 1545 francs 10 FF - 1548 francs 10 FF - 1551 francs 10 FF - 1554 francs 10 FF - 1557 francs 10 FF - 1560 francs 10 FF - 1563 francs 10 FF - 1566 francs 10 FF - 1569 francs 10 FF - 1572 francs 10 FF - 1575 francs 10 FF - 1578 francs 10 FF - 1581 francs 10 FF - 1584 francs 10 FF - 1587 francs 10 FF - 1590 francs 10 FF - 1593 francs 10 FF - 1596 francs 10 FF - 1599 francs 10 FF - 1602 francs 10 FF - 1605 francs 10 FF - 1608 francs 10 FF - 1611 francs 10 FF - 1614 francs 10 FF - 1617 francs 10 FF - 1620 francs 10 FF - 1623 francs 10 FF - 1626 francs 10 FF - 1629 francs 10 FF - 1632 francs 10 FF - 1635 francs 10 FF - 1638 francs 10 FF - 1641 francs 10 FF - 1644 francs 10 FF - 1647 francs 10 FF - 1650 francs 10 FF - 1653 francs 10 FF - 1656 francs 10 FF - 1659 francs 10 FF - 1662 francs 10 FF - 1665 francs 10 FF - 1668 francs 10 FF - 1671 francs 10 FF - 1674 francs 10 FF - 1677 francs 10 FF - 1680 francs 10 FF - 1683 francs 10 FF - 1686 francs 10 FF - 1689 francs 10 FF - 1692 francs 10 FF - 1695 francs 10 FF - 1698 francs 10 FF - 1701 francs 10 FF - 1704 francs 10 FF - 1707 francs 10 FF - 1710 francs 10 FF - 1713 francs 10 FF - 1716 francs 10 FF - 1719 francs 10 FF - 1722 francs 10 FF - 1725 francs 10 FF - 1728 francs 10 FF - 1731 francs 10 FF - 1734 francs 10 FF - 1737 francs 10 FF - 1740 francs 10 FF - 1743 francs 10 FF - 1746 francs 10 FF - 1749 francs 10 FF - 1752 francs 10 FF - 1755 francs 10 FF - 1758 francs 10 FF - 1761 francs 10 FF - 1764 francs 10 FF - 1767 francs 10 FF - 1770 francs 10 FF - 1773 francs 10 FF - 1776 francs 10 FF - 1779 francs 10 FF - 1782 francs 10 FF - 1785 francs 10 FF - 1788 francs 10 FF - 1791 francs 10 FF - 1794 francs 10 FF - 1797 francs 10 FF - 1800 francs 10 FF - 1803 francs 10 FF - 1806 francs 10 FF - 1809 francs 10 FF - 1812 francs 10 FF - 1815 francs 10 FF - 1818 francs 10 FF - 1821 francs 10 FF - 1824 francs 10 FF - 1827 francs 10 FF - 1830 francs 10 FF - 1833 francs 10 FF - 1836 francs 10 FF - 1839 francs 10 FF - 1842 francs 10 FF - 1845 francs 10 FF - 1848 francs 10 FF - 1851 francs 10 FF - 1854 francs 10 FF - 1857 francs 10 FF - 1860 francs 10 FF - 1863 francs 10 FF - 1866 francs 10 FF - 1869 francs 10 FF - 1872 francs 10 FF - 1875 francs 10 FF - 1878 francs 10 FF - 1881 francs 10 FF - 1884 francs 10 FF - 1887 francs 10 FF - 1890 francs 10 FF - 1893 francs 10 FF - 1896 francs 10 FF - 1899 francs 10 FF - 1902 francs 10 FF - 1905 francs 10 FF - 1908 francs 10 FF - 1911 francs 10 FF - 1914 francs 10 FF - 1917 francs 10 FF - 1920 francs 10 FF - 1923 francs 10 FF - 1926 francs 10 FF - 1929 francs 10 FF - 1932 francs 10 FF - 1935 francs 10 FF - 1938 francs 10 FF - 1941 francs 10 FF - 1944 francs 10 FF - 1947 francs 10 FF - 1950 francs 10 FF - 1953 francs 10 FF - 1956 francs 10 FF - 1959 francs 10 FF - 1962 francs 10 FF - 1965 francs 10 FF - 1968 francs 10 FF - 1971 francs 10 FF - 1974 francs 10 FF - 1977 francs 10 FF - 1980 francs 10 FF - 1983 francs 10 FF - 1986 francs 10 FF - 1989 francs 10 FF - 1992 francs 10 FF - 1995 francs 10 FF - 1998 francs 10 FF - 2001 francs 10 FF - 2004 francs 10 FF - 2007 francs 10 FF - 2010 francs 10 FF - 2013 francs 10 FF - 2016 francs 10 FF - 2019 francs 10 FF - 2022 francs 10 FF - 2025 francs 10 FF - 2028 francs 10 FF - 2031 francs 10 FF - 2034 francs 10 FF - 2037 francs 10 FF - 2040 francs 10 FF - 2043 francs 10 FF - 2046 francs 10 FF - 2049 francs 10 FF - 2052 francs 10 FF - 2055 francs 10 FF - 2058 francs 10 FF - 2061 francs 10 FF - 2064 francs 10 FF - 2067 francs 10 FF - 2070 francs 10 FF - 2073 francs 10 FF - 2076 francs 10 FF - 2079 francs 10 FF - 2082 francs 10 FF - 2085 francs 10 FF - 2088 francs 10 FF - 2091 francs 10 FF - 2094 francs 10 FF - 2097 francs 10 FF - 2100 francs 10 FF - 2103 francs 10 FF - 2106 francs 10 FF - 2109 francs 10 FF - 2112 francs 10 FF - 2115 francs 10 FF - 2118 francs 10 FF - 2121 francs 10 FF - 2124 francs 10 FF - 2127 francs 10 FF - 2130 francs 10 FF - 2133 francs 10 FF - 2136 francs 10 FF - 2139 francs 10 FF - 2142 francs 10 FF - 2145 francs 10 FF - 2148 francs 10 FF - 2151 francs 10 FF - 2154 francs 10 FF - 2157 francs 10 FF - 2160 francs 10 FF - 2163 francs 10 FF - 2166 francs 10 FF - 2169 francs 10 FF - 2172 francs 10 FF - 2175 francs 10 FF - 2178 francs 10 FF - 2181 francs 10 FF - 2184 francs 10 FF - 2187 francs 10 FF - 2190 francs 10 FF - 2193 francs 10 FF - 2196 francs 10 FF - 2199 francs 10 FF - 2202 francs 10 FF - 2205 francs 10 FF - 2208 francs 10 FF - 2211 francs 10 FF - 2214 francs 10 FF - 2217 francs 10 FF - 2220 francs 10 FF - 2223 francs 10 FF - 2226 francs 10 FF - 2229 francs 10 FF - 2232 francs 10 FF - 2235 francs 10 FF - 2238 francs 10 FF - 2241 francs 10 FF - 2244 francs 10 FF - 2247 francs 10 FF - 2250 francs 10 FF - 2253 francs 10 FF - 2256 francs 10 FF - 2259 francs 10 FF - 2262 francs 10 FF - 2265 francs 10 FF - 2268 francs 10 FF - 2271 francs 10 FF - 2274 francs 10 FF - 2277 francs 10 FF - 2280 francs 10 FF - 2283 francs 10 FF - 2286 francs 10 FF - 2289 francs 10 FF - 2292 francs 10 FF - 2295 francs 10 FF - 2298 francs 10 FF - 2301 francs 10 FF - 2304 francs 10 FF - 2307 francs 10 FF - 2310 francs 10 FF - 2313 francs 10 FF - 2316 francs 10 FF - 2319 francs 10 FF - 2322 francs 10 FF - 2325 francs 10 FF - 2328 francs 10 FF - 2331 francs 10 FF - 2334 francs 10 FF - 2337 francs 10

Digitization of *Le Temps* archives

- 2008
- Optical Character Recognition and Layout detection (Olive)
- 1M pages, 4M articles

Outline

1. Context: sources and project
- 2. Named Entity Recognition**
3. Lexical processing & Web Interface
4. Potential uses
5. Future work



NE definition

- “elements of interest” generally of type *Person*, *Org* and *Location*
- referential units which underlie text semantics
- proper names & definite descriptions (referential autonomy and unicity)

Origins of NE processing

- 1980's: Text Understanding
 - an over-ambitious project facing technical and theoretical difficulties
- 1990's: Information Extraction
 - focus on elements of interest
 - extraction model defined in advance based on application
 - analysis of only 10-20% of the text
- 1987-1998: Message Understanding Conference (MUC) cycle
 - financed by U.S. DARPA
 - MUC-6 1995, definition of “named entity” subtask

MUC 3 Template example (1991)

<i>19 March - A bomb went off this morning near a power tower in San Salvador leaving a large part of the city without energy, but no casualties have been reported. According to unofficial sources, the bomb - allegedly detonated by urban guerrilla commandos - blew up a power tower in the northwestern part of San Salvador at 0650 (1250 GMT).</i>	
INCIDENT TYPE	bombing
DATE	March 19
LOCATION	El Salvador : San Salvador (city)
PERPETRATOR	urban guerrilla commandos
PHYSICAL TARGET	power tower
HUMAN TARGET	-
EFFECT ON PHYSICAL TARGET	destroyed
EFFECT ON HUMAN TARGET	no injury or death
INSTRUMENT	bomb

Source: Information Extraction: Techniques and Challenges, Ralph Grishman, 1997.

NE tasks

- **recognition**: detecting named entities in texts with correct boundaries
- **classification**: categorizing NE according to a predefined typology
- **disambiguation**: linking NE mentions to a unique reference
- **relation detection**: discovering and specifying relations between NE

NE tasks

Le ministre des Affaires étrangères M. Burkhalter a esquissé les priorités du Conseil fédéral pour nouer de nouvelles relations bilatérales avec le Royaume-Uni. (Le Temps 25.06.2016)

NE tasks

*Le ministre des Affaires étrangères **M. Burkhalter** a esquissé les priorités du **Conseil fédéral** pour nouer de nouvelles relations bilatérales avec le **Royaume-Uni**. (Le Temps 25.06.2016)*

NE tasks

Le ministre des Affaires étrangères M. Burkhalter a esquisé les priorités

PERSON

du Conseil fédéral pour nouer de nouvelles relations bilatérales avec le

ORGANIZATION

Royaume-Uni. (Le Temps 25.06.2016)

LOCATION

NE tasks

FUNCTION **TITLE** **LASTNAME**
*Le ministre des Affaires étrangères **M. Burkhalter** a esquisé les priorités*
PERSON
*du **Conseil fédéral** pour nouer de nouvelles relations bilatérales avec le*
ORGANIZATION
***Royaume-Uni.** (Le Temps 25.06.2016)*
LOCATION → GEO-POLITICAL ORG

NE tasks

http://dbpedia.org/page/Didier_Burkhalter
↗
FUNCTION TITLE LASTNAME
Le ministre des Affaires étrangères **M. Burkhalter** a esquisé les priorités
PERSON
 du **Conseil fédéral** pour nouer de nouvelles relations bilatérales avec le
ORGANIZATION
Royaume-Uni. (Le Temps 25.06.2016)
LOCATION → GEO-POLITICAL ORG

Applications

Internal

- POS tagging and chunking
- Syntactic & Dependency Analysis
- Co-reference
- Translation
- Word Sense Disambiguation

External

- Media Monitoring (“5W”)
- (Cross-lingual) document clustering
- Opinion Mining
- Question-Answering
- Anonymization
- Summarization

Applications

Seat and Porsche has fewer registrations in July.

Internal

- POS tagging and chunking
- Syntactic & Dependency Analysis
- Coreference
- Translation
- Word Sense Disambiguation

External

- Media Monitoring (“5W”)
- (Cross-lingual) document clustering
- Opinion Mining
- Question-Answering
- Anonymisation
- Summarization

Applications

Jack London was an american writer. → London
London is a busy city. → Londres

Internal

- POS tagging and chunking
- Syntactic & Dependency Analysis
- Coreference
- Translation
- Word Sense Disambiguation

External

- Media Monitoring (“5W”)
- (Cross-lingual) document clustering
- Opinion Mining
- Question-Answering
- Anonymisation
- Summarization

Applications

Internal

- POS tagging and chunking
- Syntactic & Dependency Analysis
- Coreference
- Translation
- Word Sense Disambiguation

External

- Media Monitoring (“5W”)
- (Cross-lingual) document clustering
- Opinion Mining
- Question-Answering
- Anonymisation
- Summarization

Performances

Good when:

- language: English
- domain: news
- typology: simple

NE processing on historical texts

Challenges

- noisy inputs
- language of earlier stages
- lack of appropriate linguistics resources

NER on *Le Temps*: strategy

Off-the-shelf Entity Linking tool?

Convenient but:

- based on Knowledge Bases, do not *guess* new names
- KBs derived from Wikipedia, contain VIPs
- usually not always available for French
- accessible via web-services, difficult to parse millions of files

NER implementation

Rule-based, offline system

- ExPRESS, extraction pattern engine
- definition of rules to detect specific phrases
- usage of lexical resources
- light-weight: does not require morphological analysis

Rule example

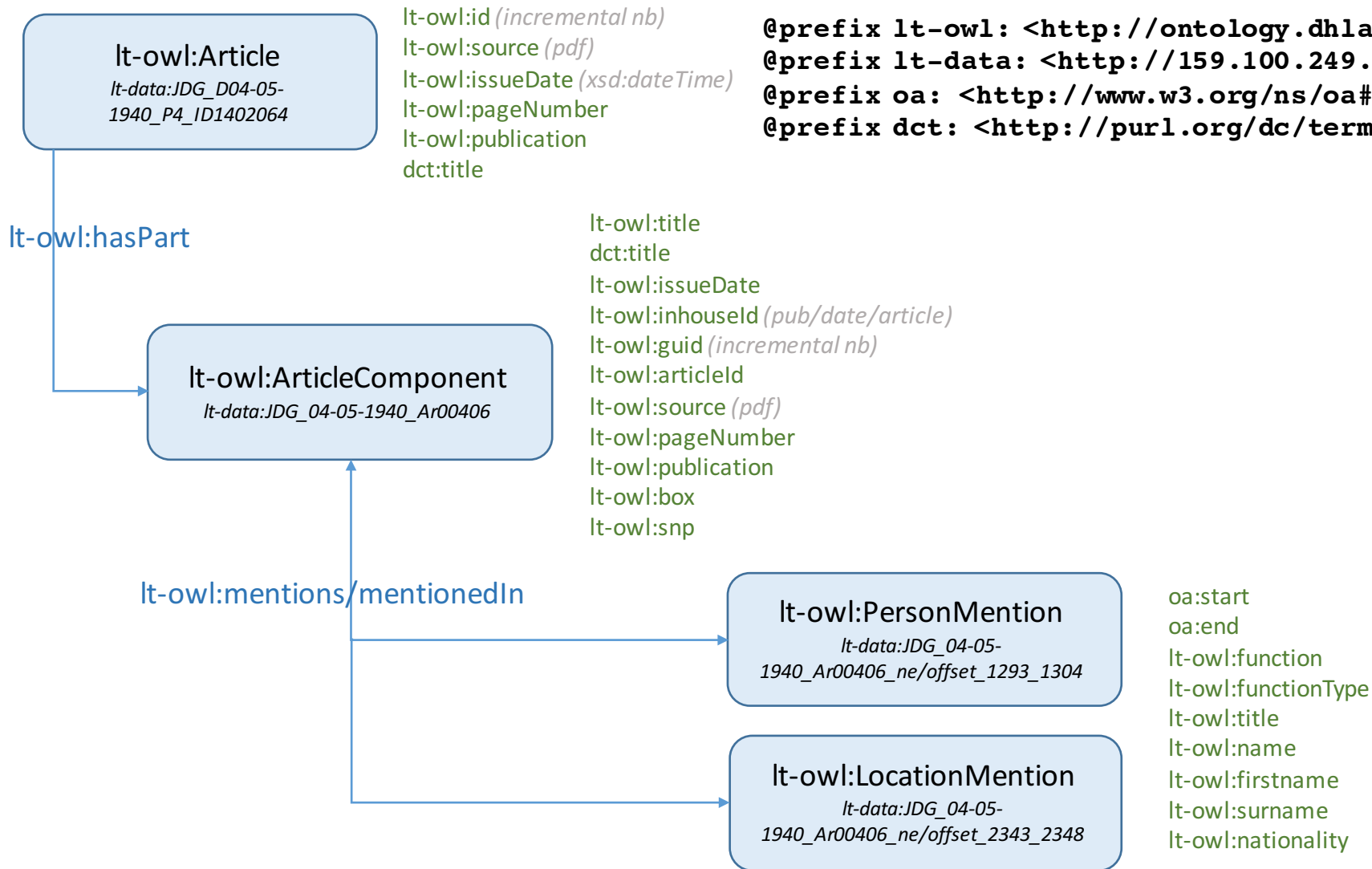
- Pattern: firstname<1,2> lastname infix UC (ex: Javier Gomez de Liano)

```
myRule :>
  (patternmatch & [GTYPE: "first_name", SURFACE:#first_name1]<1,2>
   patternmatch & [GTYPE: "last_name", SURFACE:#first_name2]
   patternmatch & [GTYPE: "name_infix", SURFACE: #infix]
   token & [TYPE: "first_capital_word", SURFACE:#last_name]):name
-> name : person_name &
[NAME:#full_name, TYPE:"Person", FIRST_NAME:#full_first_name, GLAST_NAME:#last_name,
RULENAME:"person1_basic_infix_14_stw",RULECONFIDENCE:"low"]
```

134 rules (language independent)

NER parsing

- parallelized on a cluster's node (256GB, 48 cores)
- results stored in JSON and converted into RDF
- NEs accessible via a SPARQL endpoint



Diachronic evaluation

Comparison of 4 systems over 7 time-series

- 1804, 1826, 1841, 1881, 1921, 1961, 1981
- ExPRESS
- mXS (machine learning – tuned for speech transcriptions)
- Dandelion
- Alchemy

Evaluation dataset

Manual annotation following *Quaero* guidelines

	# words		# pers		# loc		# entities	
	GDL	JDG	GDL	JDG	GDL	JDG	GDL	JDG
1804	33,769	-	425	-	966	-	1,391	-
1826	33,353	14,074	409	184	843	149	1,252	333
1841	40,784	5,558	551	68	1,117	55	1,668	123
1881	55,751	12,360	937	227	844	271	1,781	498
1921	20,117	3,587	377	47	561	134	938	181
1961	23,302	7,842	524	111	549	145	1,073	256
1981	17,759	3,521	257	77	324	54	581	131
TOTAL	299,159	64,370	3,480	714	5,204	808	8,684	1,522

Table 1: Data set statistics.

	Dandelion			Alchemy			Rule-based			mXS		
	P	R	F	P	R	F	P	R	F	P	R	F
1804	0.48 0.18 / 0.55	0.48 0.11 / 0.64	0.48 0.14 / 0.59	0.53 0.45 / 0.58	0.28 0.28 / 0.29	0.37 0.34 / 0.38	0.76 0.53 / 0.81	0.34 0.12 / 0.44	0.47 0.2 / 0.57	0.47 0.22 / 0.64	0.28 0.18 / 0.32	0.35 0.2 / 0.43
1826	0.48 0.19 / 0.55	0.47 0.11 / 0.65	0.48 0.14 / 0.59	0.49 0.4 / 0.56	0.28 0.31 / 0.27	0.36 0.35 / 0.36	0.74 0.44 / 0.8	0.37 0.12 / 0.48	0.49 0.19 / 0.6	0.45 0.22 / 0.66	0.31 0.22 / 0.36	0.37 0.22 / 0.47
1841	0.61 0.24 / 0.69	0.53 0.12 / 0.74	0.57 0.16 / 0.72	0.56 0.6 / 0.54	0.3 0.33 / 0.28	0.39 0.43 / 0.37	0.87 0.69 / 0.89	0.38 0.12 / 0.51	0.53 0.2 / 0.65	0.52 0.25 / 0.69	0.29 0.17 / 0.35	0.38 0.2 / 0.47
1881	0.43 0.26 / 0.48	0.37 0.09 / 0.68	0.4 0.13 / 0.56	0.61 0.72 / 0.51	0.37 0.4 / 0.34	0.46 0.52 / 0.4	0.69 0.66 / 0.7	0.32 0.14 / 0.53	0.44 0.23 / 0.6	0.47 0.38 / 0.56	0.32 0.26 / 0.37	0.38 0.31 / 0.45
1921	0.59 0.38 / 0.64	0.5 0.14 / 0.75	0.55 0.2 / 0.69	0.6 0.72 / 0.51	0.33 0.42 / 0.28	0.43 0.53 / 0.36	0.81 0.7 / 0.85	0.39 0.23 / 0.5	0.53 0.35 / 0.63	0.47 0.32 / 0.62	0.28 0.23 / 0.3	0.35 0.27 / 0.41
1961	0.49 0.39 / 0.53	0.46 0.22 / 0.69	0.48 0.28 / 0.6	0.67 0.72 / 0.58	0.39 0.51 / 0.27	0.49 0.6 / 0.37	0.73 0.66 / 0.79	0.29 0.25 / 0.34	0.42 0.37 / 0.47	0.52 0.41 / 0.66	0.29 0.27 / 0.3	0.37 0.33 / 0.42
1981	0.41 0.3 / 0.47	0.51 0.3 / 0.68	0.46 0.3 / 0.56	0.6 0.76 / 0.46	0.42 0.56 / 0.31	0.5 0.65 / 0.37	0.68 0.72 / 0.65	0.41 0.41 / 0.4	0.51 0.52 / 0.5	0.43 0.31 / 0.6	0.35 0.32 / 0.37	0.38 0.31 / 0.45
All years	0.5 0.27 / 0.56	0.47 0.14 / 0.69	0.48 0.18 / 0.62	0.58 0.63 / 0.54	0.33 0.39 / 0.29	0.42 0.48 / 0.38	0.76 0.64 / 0.8	0.35 0.18 / 0.47	0.48 0.28 / 0.59	0.48 0.31 / 0.63	0.3 0.24 / 0.34	0.37 0.27 / 0.45

Table 2: Precision, Recall and F-measure of the 4 systems on GDL corpus. Global figures for all entities are reported (bigger font), as well as figures per entity type (smaller font, with person/location).

	Dandelion			Alchemy			Rule-based			mXS		
	P	R	F	P	R	F	P	R	F	P	R	F
1804	0.48 0.18 / 0.55	0.48 0.11 / 0.64	0.48 0.14 / 0.59	0.53 0.45 / 0.58	0.28 0.28 / 0.29	0.37 0.34 / 0.38	0.76 0.53 / 0.81	0.34 0.12 / 0.44	0.47 0.2 / 0.57	0.47 0.22 / 0.64	0.28 0.18 / 0.32	0.35 0.2 / 0.43
1826	0.48 0.19 / 0.55	0.47 0.11 / 0.65	0.48 0.14 / 0.59	0.49 0.4 / 0.56	0.28 0.31 / 0.27	0.36 0.35 / 0.36	0.74 0.44 / 0.8	0.37 0.12 / 0.48	0.49 0.19 / 0.6	0.45 0.22 / 0.66	0.31 0.22 / 0.36	0.37 0.22 / 0.47
1841	0.61 0.24 / 0.69	0.53 0.12 / 0.74	0.57 0.16 / 0.72	0.56 0.6 / 0.54	0.3 0.33 / 0.28	0.39 0.43 / 0.37	0.87 0.69 / 0.89	0.38 0.12 / 0.51	0.53 0.2 / 0.65	0.52 0.25 / 0.69	0.29 0.17 / 0.35	0.38 0.2 / 0.47
1881	0.43 0.26 / 0.48	0.37 0.09 / 0.68	0.4 0.13 / 0.56	0.61 0.72 / 0.51	0.37 0.4 / 0.34	0.46 0.52 / 0.4	0.69 0.66 / 0.7	0.32 0.14 / 0.53	0.44 0.23 / 0.6	0.47 0.38 / 0.56	0.32 0.26 / 0.37	0.38 0.31 / 0.45
1921	0.59 0.38 / 0.64	0.5 0.14 / 0.75	0.55 0.2 / 0.69	0.6 0.72 / 0.51	0.33 0.42 / 0.28	0.43 0.53 / 0.36	0.81 0.7 / 0.85	0.39 0.23 / 0.5	0.53 0.35 / 0.63	0.47 0.32 / 0.62	0.28 0.23 / 0.3	0.35 0.27 / 0.41
1961	0.49 0.39 / 0.53	0.46 0.22 / 0.69	0.48 0.28 / 0.6	0.67 0.72 / 0.58	0.39 0.51 / 0.27	0.49 0.6 / 0.37	0.73 0.66 / 0.79	0.29 0.25 / 0.34	0.42 0.37 / 0.47	0.52 0.41 / 0.66	0.29 0.27 / 0.3	0.37 0.33 / 0.42
1981	0.41 0.3 / 0.47	0.51 0.3 / 0.68	0.46 0.3 / 0.56	0.6 0.76 / 0.46	0.42 0.56 / 0.31	0.5 0.65 / 0.37	0.68 0.72 / 0.65	0.41 0.41 / 0.4	0.51 0.52 / 0.5	0.43 0.31 / 0.6	0.35 0.32 / 0.37	0.38 0.31 / 0.45
All years	0.5 0.27 / 0.56	0.47 0.14 / 0.69	0.48 0.18 / 0.62	0.58 0.63 / 0.54	0.33 0.39 / 0.29	0.42 0.48 / 0.38	0.76 0.64 / 0.8	0.35 0.18 / 0.47	0.48 0.28 / 0.59	0.48 0.31 / 0.63	0.3 0.24 / 0.34	0.37 0.27 / 0.45

Table 2: Precision, Recall and F-measure of the 4 systems on GDL corpus. Global figures for all entities are reported (bigger font), as well as figures per entity type (smaller font, with person/location).

	Dandelion			Alchemy			Rule-based			mXS		
	P	R	F	P	R	F	P	R	F	P	R	F
1804	0.48 0.18 / 0.55	0.48 0.11 / 0.64	0.48 0.14 / 0.59	0.53 0.45 / 0.58	0.28 0.28 / 0.29	0.37 0.34 / 0.38	0.76 0.53 / 0.81	0.34 0.12 / 0.44	0.47 0.2 / 0.57	0.47 0.22 / 0.64	0.28 0.18 / 0.32	0.35 0.2 / 0.43
1826	0.48 0.19 / 0.55	0.47 0.11 / 0.65	0.48 0.14 / 0.59	0.49 0.4 / 0.56	0.28 0.31 / 0.27	0.36 0.35 / 0.36	0.74 0.44 / 0.8	0.37 0.12 / 0.48	0.49 0.19 / 0.6	0.45 0.22 / 0.65	0.31 0.22 / 0.36	0.37 0.22 / 0.47
1841	0.61 0.24 / 0.69	0.53 0.12 / 0.74	0.57 0.16 / 0.72	0.56 0.6 / 0.54	0.3 0.33 / 0.28	0.39 0.43 / 0.37	0.87 0.69 / 0.89	0.38 0.12 / 0.51	0.53 0.2 / 0.65	0.52 0.25 / 0.69	0.29 0.17 / 0.35	0.38 0.2 / 0.47
1881	0.43 0.26 / 0.48	0.37 0.09 / 0.68	0.4 0.13 / 0.56	0.61 0.72 / 0.51	0.37 0.4 / 0.34	0.46 0.52 / 0.4	0.69 0.66 / 0.7	0.32 0.14 / 0.53	0.44 0.23 / 0.6	0.47 0.38 / 0.55	0.32 0.26 / 0.37	0.38 0.31 / 0.45
1921	0.59 0.38 / 0.64	0.5 0.14 / 0.75	0.55 0.2 / 0.69	0.6 0.72 / 0.51	0.33 0.42 / 0.28	0.43 0.53 / 0.36	0.81 0.7 / 0.85	0.39 0.23 / 0.5	0.53 0.35 / 0.63	0.47 0.32 / 0.62	0.28 0.23 / 0.3	0.35 0.27 / 0.41
1961	0.49 0.39 / 0.53	0.46 0.22 / 0.69	0.48 0.28 / 0.6	0.67 0.72 / 0.58	0.39 0.51 / 0.27	0.49 0.6 / 0.37	0.73 0.66 / 0.79	0.29 0.25 / 0.34	0.42 0.37 / 0.47	0.52 0.41 / 0.65	0.29 0.27 / 0.3	0.37 0.33 / 0.42
1981	0.41 0.3 / 0.47	0.51 0.3 / 0.68	0.46 0.3 / 0.56	0.6 0.76 / 0.46	0.42 0.56 / 0.31	0.5 0.65 / 0.37	0.68 0.72 / 0.65	0.41 0.41 / 0.4	0.51 0.52 / 0.5	0.43 0.31 / 0.6	0.35 0.32 / 0.37	0.38 0.31 / 0.45
All years	0.5 0.27 / 0.56	0.47 0.14 / 0.69	0.48 0.18 / 0.62	0.58 0.63 / 0.54	0.33 0.39 / 0.29	0.42 0.48 / 0.38	0.76 0.64 / 0.8	0.35 0.18 / 0.47	0.48 0.28 / 0.59	0.48 0.31 / 0.63	0.3 0.24 / 0.34	0.37 0.27 / 0.45

Table 2: Precision, Recall and F-measure of the 4 systems on GDL corpus. Global figures for all entities are reported (bigger font), as well as figures per entity type (smaller font, with person/location).

Outline

1. Context: sources and project
2. Named Entity Recognition
- 3. Lexical processing & Web Interface**
4. Potential uses
5. Future work

Lexical processing

Lexical indexing

- Apache Solr

N-grams

- contiguous sequence of n items from a given sequence (char, word)
- word 1/2/3 grams with the relative frequency per year

Web interface

Main principles

- full text, high performance search engine
- articles shown in their full publication context
- browsing search results via n-gram viewer



Rechercher dans les archives de la Gazette de Lausanne et du Journal de Genève



Ex: Ramuz, Confédération helvétique, tailleur, gdl:22.07.1969

[Aide](#)

200 ans d'histoire en ligne

Le Temps est heureux de vous proposer les collections numérisées des trois prestigieux quotidiens dont il est l'héritier : le Journal de Genève, la Gazette de Lausanne et Le Nouveau Quotidien.

[EN SAVOIR PLUS](#)

Articles

Images

Publicités

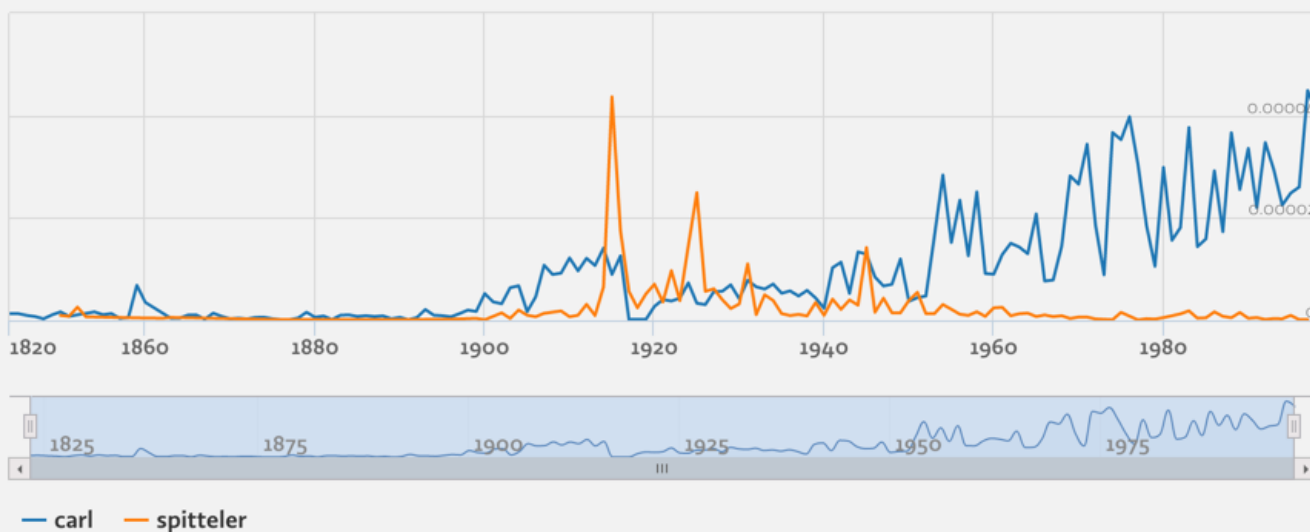
Sources (2) ▾

Trier par: Pertinence ▾

Il y a **43** résultat(s) - [Afficher le graphique](#) ^

Zoom 10a 100a Tous

De 1820 à 1998

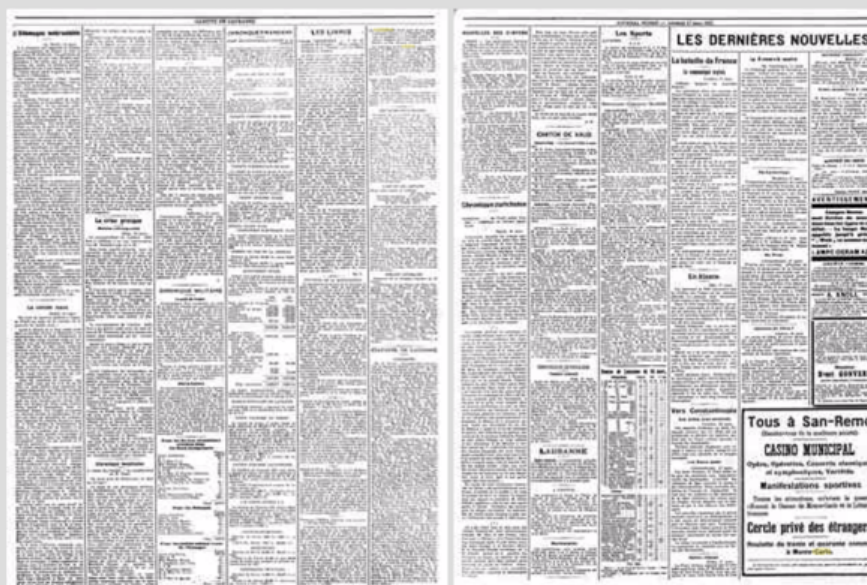


Highcharts.com

SOMMAIRE :
CAUSERIE. — La Promenade. — J. Cd.
UN POÈTE SUISSE. **CARL SPITTELER.** — I. Kaiser.
CHRONIQUE SCIENTIFIQUE. — Les Rayons N
de Blondlot. — R.

— Vous parlez d'or, repartit M. Durand, et pour ma part je trouve admirable cette conquête moderne du suffrage universel qui permet de trancher les questions le plus simplement du monde. Vous êtes neuf, nous sommes dix : comment douter que l'opinion des dix ne soit la meilleure, et bonne par conséquent à imposer à la généralité ? Il n'y a plus d'aristocrates et tout le monde est

nel
ils
co
Ur



ERE

LES LIVRES

L. A. G.
URICH
xième)
fonds
500 fr.;
1 fr. 04.
quitté
er avril
.. à Zu-

E

le pre-
mie. Ce
ert par

OIS

gnie du
général

« © »

omande

Valentin GRANDJEAN. — *A la découverte de l'amour*. — Hors commerce. Editions de la « Violette », 16, rue Etienne-Dumont, Genève.

Ce livre, de lecture très attachante, est une étude psychologique et fait abstraction presque complète de toute intrigue. Ce sont quatre aventures amoureuses juxtaposées à la façon d'un livret « à tiroir », comme celui des *Contes d'Hoffmann*. Chacune expose un cas particulier, et met en scène un type de femme bien distinct. Sans doute n'épuise-t-il pas la matière. Il existe encore d'autres types de femmes, et sans prétendre que la perfection soit de ce monde, on peut admettre qu'Edmond Laurence n'a pas été très chanceux. Le héros est lui-même un être un peu à part, sans personnalité accusée, trop passif, sans

Carl SPITTELER. *Notre point de vue suisse*. Conférence donnée le 14 décembre à Zurich sous les auspices de la Nouvelle société helvétique. Traduite par Catherine Guillard. — Rascher et Cie, éditeurs, à Zurich.

La conférence de Carl Spitteler qui a si fort défrayé la chronique et que la *Bibliothèque universelle* a publiée intégralement dans l'excellente traduction française de Mlle Catherine Guillard, vient de paraître en librairie sous forme d'une élégante plaquette de 28 pages, éditée par la maison Rascher et Cie dans sa série de publications sur la vie et l'art suisses. C'est un document précieux à posséder et à conserver.

André ALLAZ. *L'helvétisme, péril national*. En vente dans les kiosques et bibliothèques de gares et à l'Agence bibliographique, Pérolles, 57, Fribourg.

Dans cette brochure l'auteur s'attache à mettre en lumière certaines tendances de nos jeunes écrivains nationalistes et montre le danger de ces tendances pour l'avenir de la Suisse.

REVUE DES DEUX MONDES

maire de la livraison du 1er mars :
empereur Guillaume, par M. le baron
s. — Le mystère des béatitudes, troisième
partie, par Colette Yver. — Un témoi-
alsacien sur l'Alsace-Lorraine, par M.

NOUVEL

TESSIN. —
de Lugano on
tion du 28 ma
té d'action au
vent arrêter ensemble les comités radical
conservateur.

— Le comité cantonal conservateur a délégué une commission pour délibérer en commun avec le parti radical sur le programme économique à arrêter pour la législature à venir.

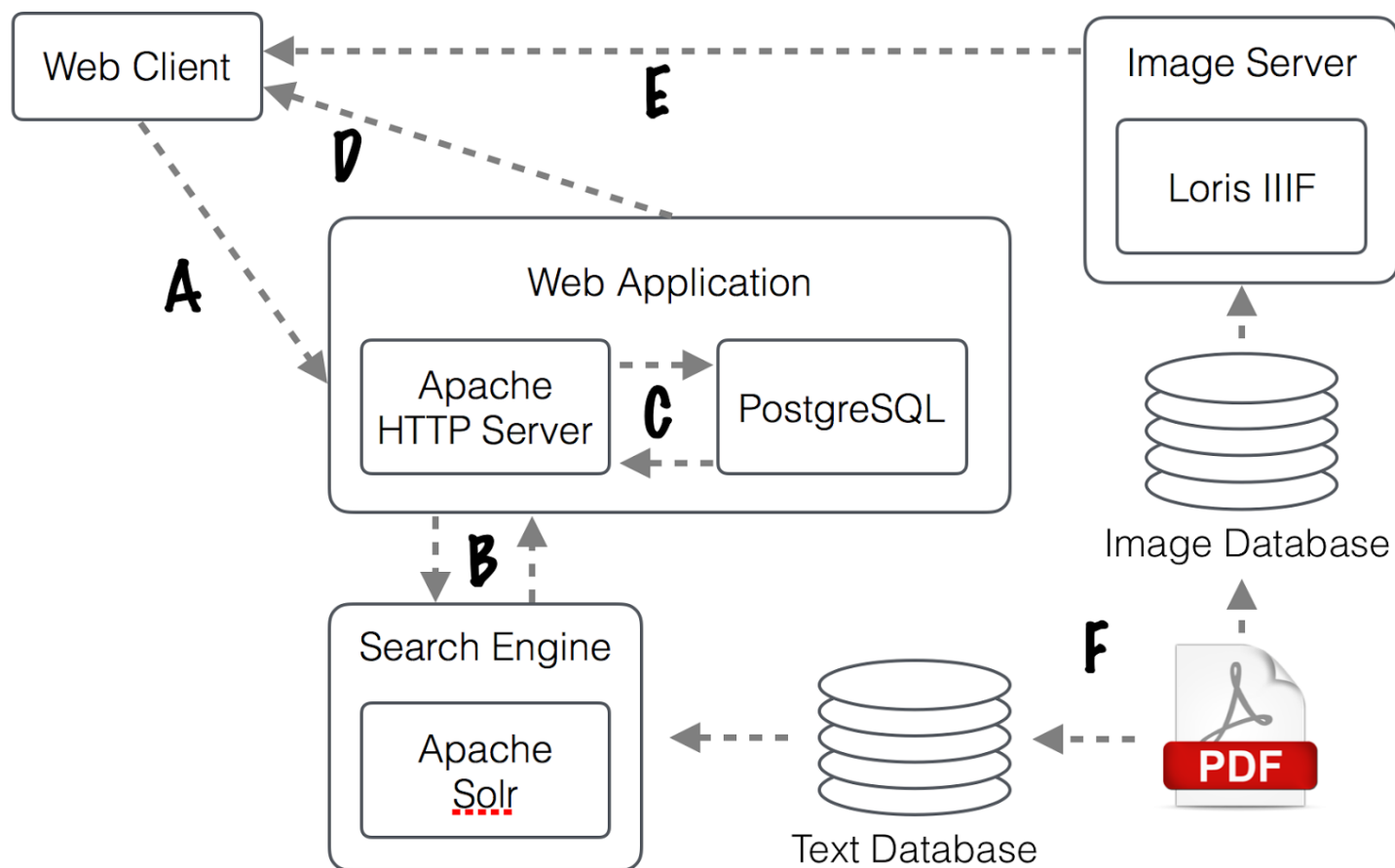
Les délégués libéraux sont MM. Garban, Nerini, conseiller national, Borella, Rossi, Mazzini, conseiller d'Etat. Les délégués conservateurs sont MM. Riva, conseiller national, Cattori et Martinoli, conseiller d'Etat.

— Le Conseil d'Etat publie le texte du décret destituant le professeur Paul Bardazzi Italien, de ses fonctions à l'Ecole technique de Locarno. Il a résolu en outre de poursuivre le journal socialiste la *Libera Stampa* pour son attitude dans la question de la destitution du prof. Bardazzi.

FRIBOURG. — La municipalité de Bulle a interdit aux revendeurs l'accès du marché avant 11 h. 30 du matin, de façon à éviter l'accaparement des denrées alimentaires.

— La *Liberté* annonce que mercredi de nombreux religieux de la Congrégation du D

Architecture



auteur : Cyril Bornet

Potential uses

- study of linguistic evolution (e.g. word senses)
- contextualization support for journalists (and citizens)
- study of local, national and international history
- study of editorial practices
- importance and role of images and/or advertisements
- ...

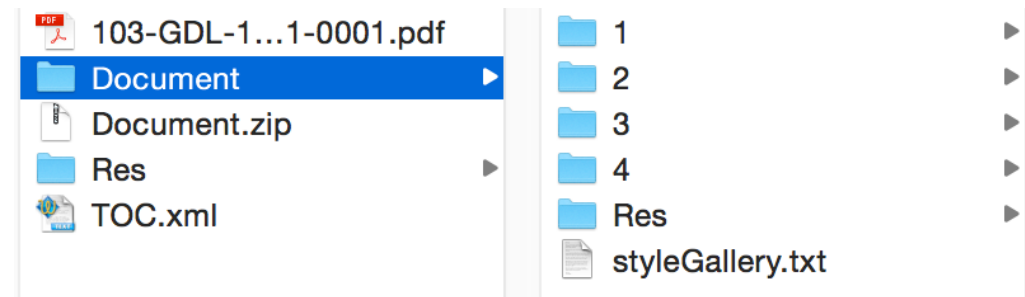
Future work

- NER refinement
- NE disambiguation
- Multilayer linguistic annotation: POS, Dependencies, WSD
- Article classification
- Corpus release

Thank you

Back up

XML database



OCR output for each issue:

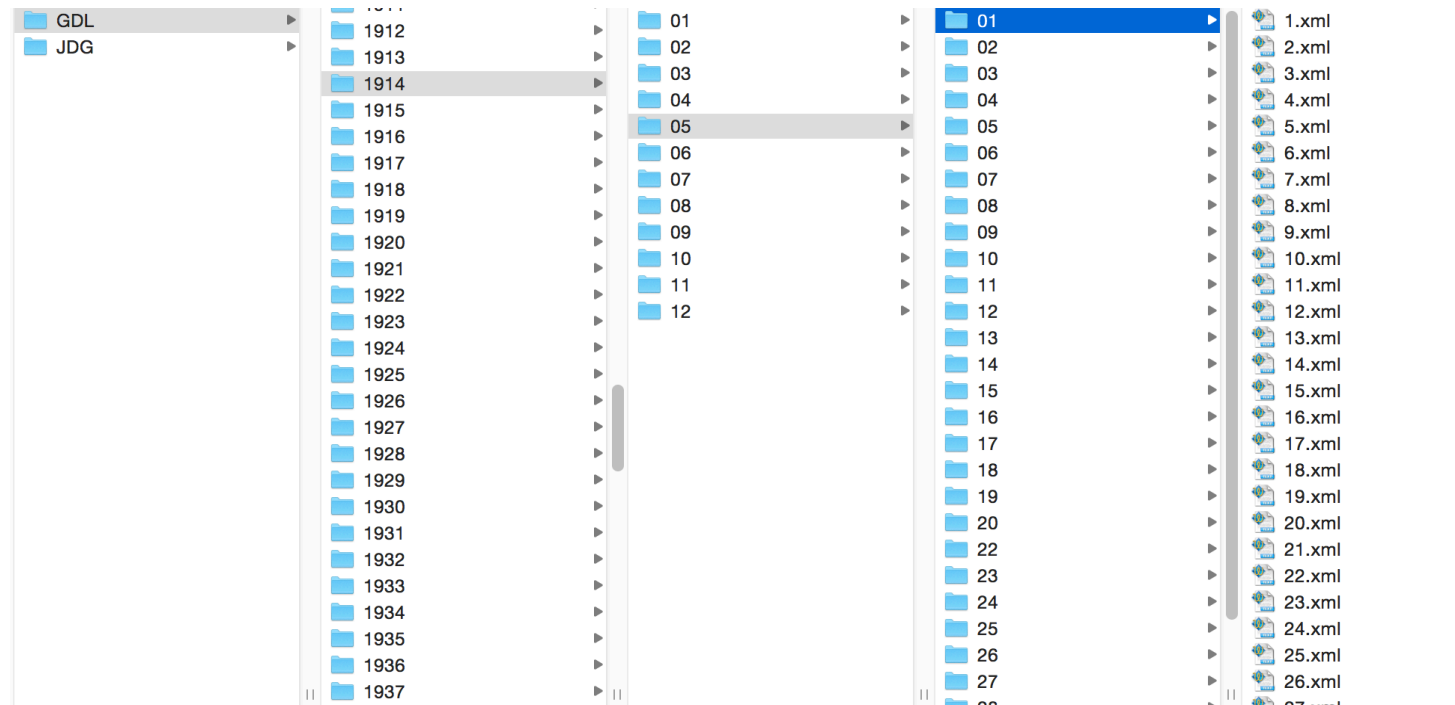
- PDF for the whole issue
- PDFs per page
- PDFs per primitives (pieces of images)
- XML per piece of text, one-word per box
- Master file to link everything

quite cumbersome...

Text-only XML files

- articles re-built from XML database
- text re-built from tokens

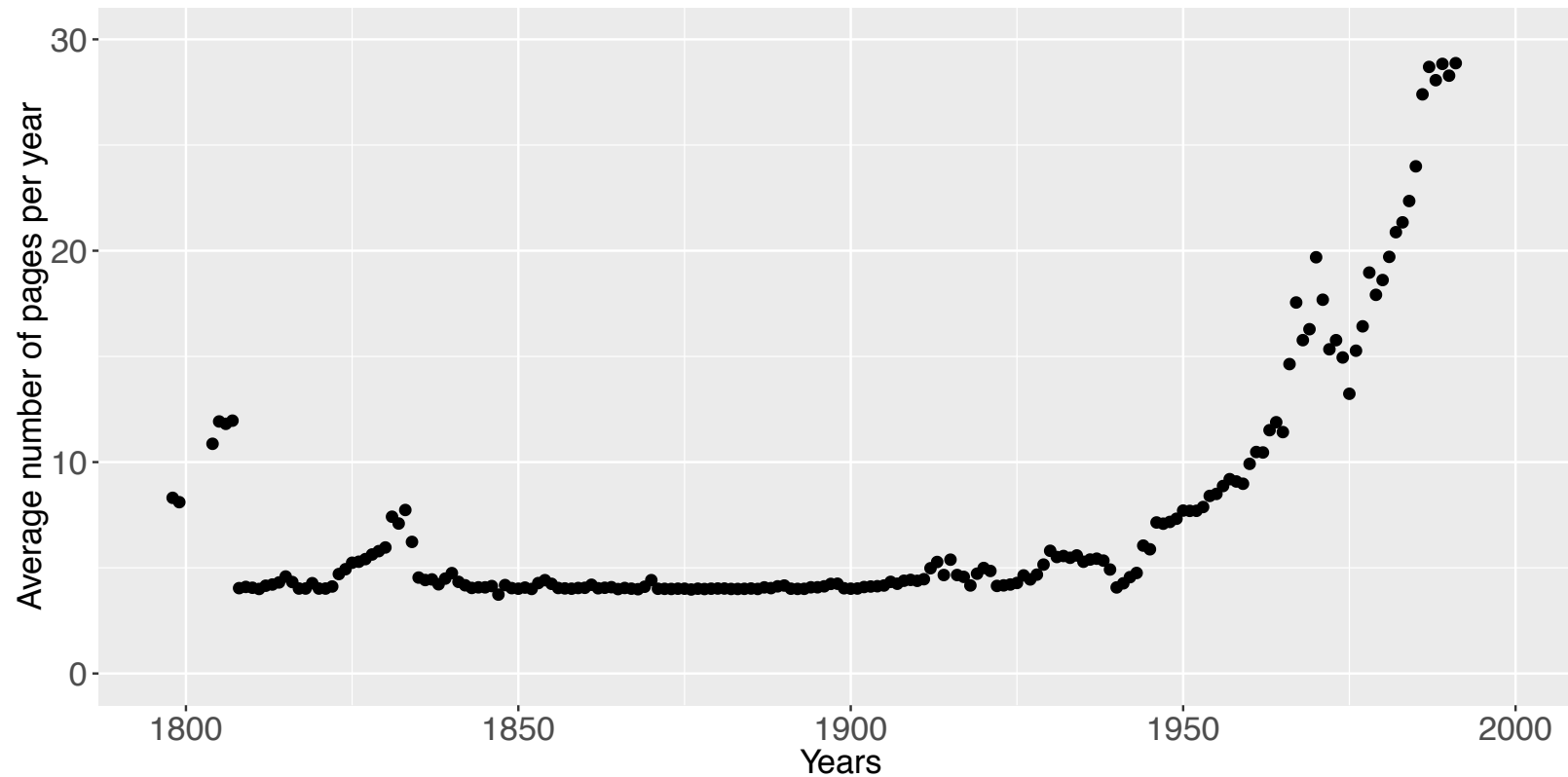
... much easier !



```
<?xml version="1.0" encoding="UTF-8"?><article><entity><meta><box>2038 1967 2438 2422</box><snp>Ar00104S.png</snp><id>Ar00104</id><language>French</language><entity_type>Article</entity_type><page_no>1</page_no><name>L'Orient</name><publication>GDL</publication><issue_date>01/05/1914</issue_date><word_count>110</word_count><total_chars_count>488</total_chars_count><suspicious_chars_count>0</suspicious_chars_count><updated_word_count>76</updated_word_count><updated_char_count>484</updated_char_count></meta><links><source>103-GDL-1914-05-01-0001.pdf</source><first_id>Ar00100</first_id><last_id>Ar00108</last_id><next_id>Ar00105</next_id><prev_id>Ar00103</prev_id><continuation_to>Ar00200</continuation_to></links><full_text>L'Orient. En Macédoine grecque. Salonique, 30 avril. Trois vapeurs remplis de réfugiés de la Thrace viennent d'arriver à Salonique amenant en tout 4400 personnes. Dans l'espace de neuf jours, du 18 au 27 avril, 19,600 réfugiés de la Thrace sont arrivés sur 14 vapeurs. Il y a donc actuellement 24,000 réfugiés. Selon un rapport officiel, au moment de l'occupation de Salonique par les troupes grecques il y avait 10,000 réfugiés musulmans pour lesquels le gouvernement hellénique a dépensé 71,709 fr. Après la guerre gréco-bulgare le nombre des nouveaux réfugiés musulmans à</full_text></entity><entity><meta><box>94 152 468 370</box><snp></snp><id>Ar00200</id><language>French</language><entity_type>Article</entity_type><page_no>2</page_no><name>Untitled Article</name><publication>GDL</publication><issue_date>01/05/1914</issue_date><word_count>60</word_count><total_chars_count>279</total_chars_count><suspicious_chars_count>0</suspicious_chars_count><updated_word_count>45</updated_word_count><updated_char_count>278</updated_char_count></meta><links><source>103-GDL-1914-05-01-0001.pdf</source><first_id>Ar00200</first_id><last_id>Ad00470</last_id><next_id>Ar00201</next_id><prev_id></prev_id><continuation_from>Ar00104</continuation_from></links><full_text>Salonique et Cavala s'éleva à plus de 75,000 jusqu'à la fin de décembre. Ils furent tous l'objet de la sollicitude du gouvernement hellénique. Le gouverneur général Sophoulis est arrivé aujourd'hui et est entré aussitôt en fonctions. M. Venizelos est en route pour la Macédoine. Son absence d'Athènes durera une huitaine de jours.</full_text></entity></article>
```

- <article> can contain several <entity>
- meta:
 - box
 - snp
 - id
 - language
 - entity_type
 - page_no
 - name
 - publication
 - issue date
 - word/chars counts
- links
- full_text

GDL pages per year



JDG pages per year

