

Metadata mining of large collections of historical newspapers

Supervisors: Maud Ehrmann and Matteo Romanello

Bachelor Project from Christian Gasser - student in Electrical and Electronics Engineering

INTRODUCTION

2-1

2-3

3

For my semester project, I created a tool that allows to analyse variation in publication date interval. It can detect outliers and shows the stability of the journal through the years.

4561 issues Le Chroniqueur PUBLICATION DATE INTERVAL ANALYZER x[n] = a1*x[n-1]+a2*x[n-2]. This is the model for the linear prediction. The coefficients al and a2 are choosen in order to difference of the days interval prediction and the real days interval what is the minimum error you want to highlight (in orange on the big Outliers Error threshold Threshold: 0.01 and 0.2 days Error mean 0.4 Day intervall frequencies The tool is composed of a main space for view the intervals, see the outliers and show the selected front-pages. The lateral view contains the outliers controller, the mean error graph and mount Number of occurrence

0.5

2-2

3

the frequency analyzer.

1-1 TOOL

2

2-1

2-2

2-3

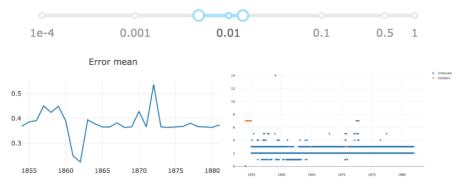
2-4

3

In my programm, rather to choose the coefficient a half for each value I put two coefficients that are computet to minimize the error in order to not detect each Sunday for example...

For my error mean graph and the outlier tool I process the difference between the mean value of the two last intervals and the actual interval. I call that the error. Then I can plot the mean error and choose what point I want to hightlight with a minimum and maximum limit.

$$m_n = \frac{1}{2}x_{n-1} + \frac{1}{2}x_{n-2}$$
$$|e_n| = |m_n - x_n|$$



1-1

2 QUALITATIVE ANALYSE

2-1

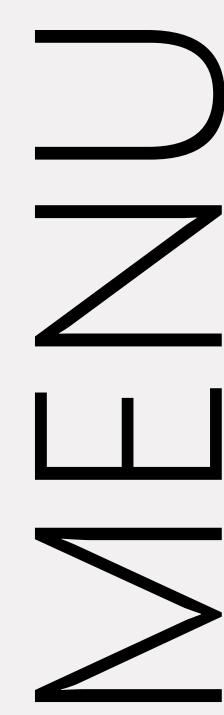
2-2

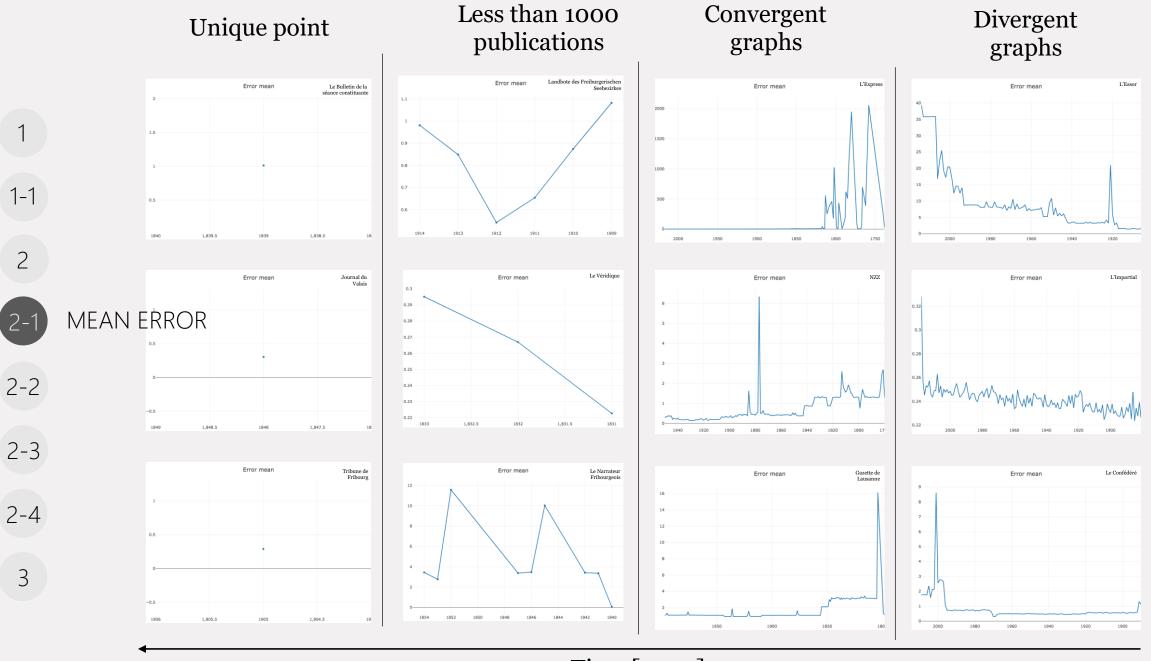
2-3

2-4

3

- 1. Mean error analysis
- 2. Publication interval: analysis example
- 3. History through thumbnail
- 4. Outliers detection





Time [years]

Année	Événement
1830	Invention du télégraphe
1876	Invention du téléphone
1895	Création de l'agence télégraphique suisse, ATS/SDA

Journal	$\Delta t_{telegraph}$	Δt_{phone}	Δt_{ats}	$t_{telegraph}$	t_{phone}	t_{ats}
GDL	19	34	15	1849	1910	1910
JDG	2	22	3	1832	1898	1898
EXP	25	8	-	1855	1884	-
NZZ	5	5	-	1835	1881	-

1-1

2

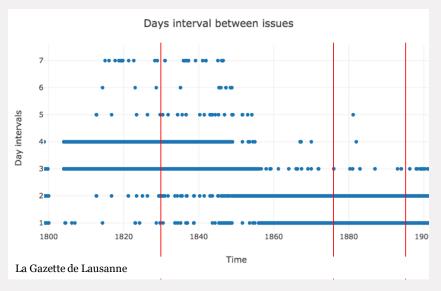
2-

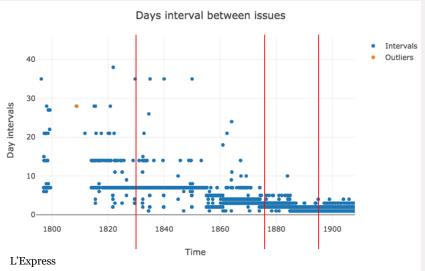
2-2 APPARITION INTERVAL

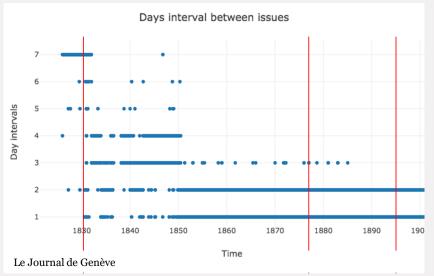
2-3

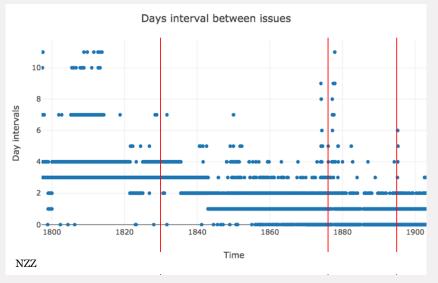
2-4

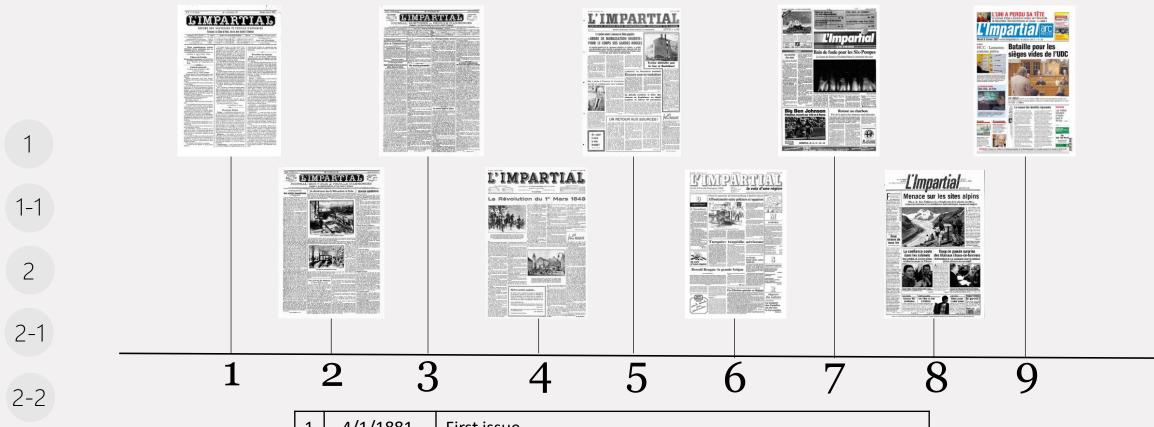
3











2-4

3

1	4/1/1881	First issue
2	29/3/1912	Publication of pictures are frequent
3	4/10/1912	Last time we see the train timetable under the title
4	28/2/1948	New design for the century of Neuchâtel
5	3/4/1967	New design after fusion with «Feuille d'Avis des Montagnes»
6	23/9/1981	New formula
7	31/8/1987	First time color
8	20/3/1987	New design to distinguish «L'Impartial» and «L'Express»
9	6/2/2007	Creation of the branding «Arc Presse»

1-1

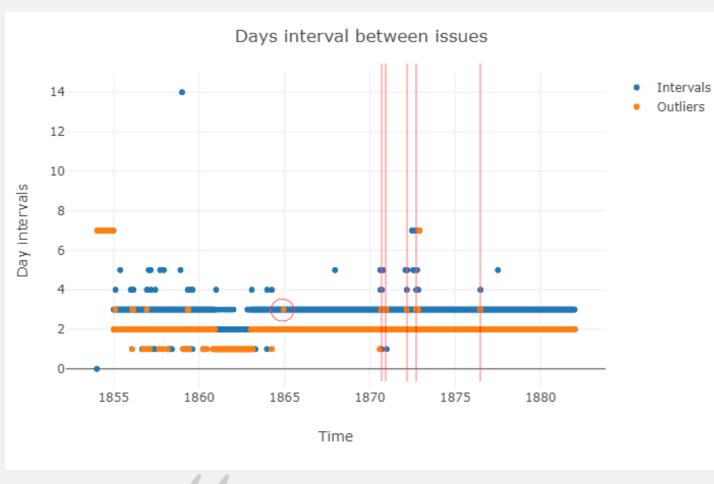
2

2-1

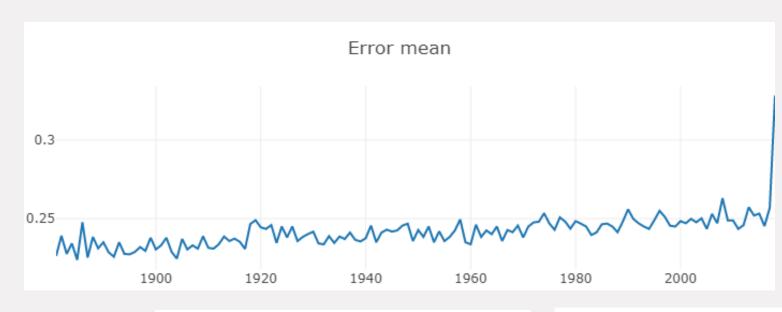
2-2

2-3

2-4 OUTLIERS



Projection of the points on horizontal lines are not interesting. But single points yes!



2-1

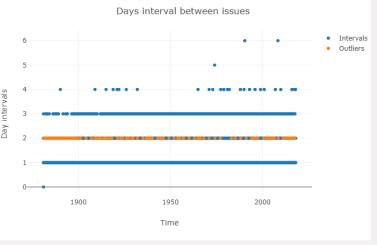
2

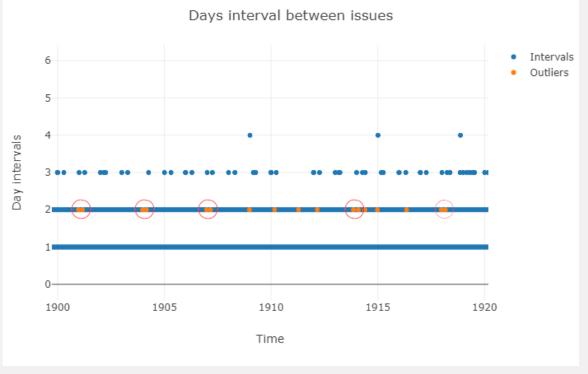
2-2

2-3

3

2-4 OUTLIERS





LES JOURNAUX REUNIS des Montagnes Heuchâteloises

A nos	lectours	Motes d'un Passant		
philipses of priori of rishs the revent share in a part of rishs the revent share in a part of a	differed for mediagnitic receives in print search and metally disc control or call variables inscribed in mediagnitic measurable. Search and search and force machinered device. It is made to express that advantum or an experiment on a compan- dative advantum or an experiment on a compan- dative advantum or an experiment of the search and advantum or an experiment of the search and advantum or an experiment of the search and an experiment of protection and search and an experimental processing of the party values are as providentation layer.	Ed leçon d'a	schilles except 602 resp. Or pay prome ton. On list contine.)	
	Aches sectates ns l'assurance-vieillesse	Le commun.— Not. Aug. pelo, he was not de cel- me as par sun pelot. Co. de sin cito de 200. Maistiment non antre basciere. Militario- tura la Regionale que dema cervigas personi- dans la rua. Che so compete de principal- den la rua. Per la constitución de principal- ne de la recompetencia.	Le binqueren Obserbinden? Le rétaculire Obserbinden? Le decoire Obserbinden? Le decoire 46879? Le réport Blue le vair des laire de progrès réportante? Il fabre de birs ceres le réport occident? Il les faire qui de la réport occident ? les faire qui de	
La Chesa-do-France, le les départes. Les autrisses de l'intrative Grédule de par-	Expansional de remaio dicionale con circles sociales. La possible condition pass en Chel aut dem è reside un diciona acciona, est d'illes d'u- legé en lateral de clara acciona, est d'illes d'u- legé en lateral de clara ac lang le lateral	In Sugramo dir sale red and successive Van and :	depart d'uniter fave le mainer Crédeded e Cit, hanclere discounte en mon grants per fraperation, soign marbil à Wess, A la per claime lague? Firegibe.	
bedikte ser in meinin pellingen en fongeringen. In ett sent Thomstrockisk deltage die petitic. I finned de mein deutstene unknotsitiksen – fitte merenteliniet 2 halbe demander – de Bart erstenderen penander tri de karn scholichisen Mattaliana Aure, ist auf hands ein orderte ser la bronze für de pelleministe, als in ein einstelle Ser de Barten metitat, de harr sloss, mein ein Ser de Barten metitat, de harr sloss, mein ein Ser die Barten metitat, de harr sloss, mein ein Ser die Barten metitat, de harr sloss, mein ein ein der sloss de s	200 m Handle de vicin au just je tar 1 de auser sekter einemendig eine ein eileben au. Gester wer in finner proteitien, it nach ause der trei de janeckt dichteriens verschafte, der abertielte gebindelten, er nach auser der abertielte gebindelten, er nach auser fallen werden, Ander Steller Gesteren. All die meldenset diesbor die batten de na- teile meldenset diesbor die batten de na- teile.	La tragédie grecque	requirity does From that you go problems from a does in our be to real proportions one to be part of a transplatfor authority of the to become your your business or as or from gold converts to people favour bandle a proof	
Les devineurs du 5 les Gerballes Frances part cer de qui cel arracter, arts aband la estimat frances; les holes : « Nove terret- les, peur l'aussisse-violteux, cleve de tant- des peur l'aussisse-violteux, cleve de tant- taix de recharde un violité du accesseux fatural :	Anne der mitten die dendelbere versiende (Fa- rone et die mitte an mitt I ETP a fallent pirt- der anne mittelle partie de see fallentiep, et dans refers dat met mittelle (diesel, ette a part- tel in impeal time!)	On the part point and althoral houseast allow	Conspells not on sell, exists at him purior it real area is neared, or other, if one ser is	
We in other of other protects, one error billion on or a passion of viscolicital infliction of the passion of viscolicital infliction to be only the other protection of the passion of th	has, her performant für mitte die beim ihren liegen – der nach auf die dies deutschieft. Vielle omwennt his menn dem deutschie in publi- chiefferen ein in hoheren sonstennissen zum deutschieferen im in hoheren der deutschieferen im Somm er melden, jesse ein serbentlik. Zinge vom beim der melden jesse ein serbentlik. Zinge vom beim die nichtlichen, melderen Zit zeuspeller für deutschie die demokrative die gegen gestellt, aus den melden die Arbeiter die den sein sonste den Jahren sordien in demokrative vom sonste deuts- liche der demokrative der der der der Jahren sordien in demokrative vom sonste deuts- geliebt unt der der der der der der der der der der der der der der der der der	on darge in organism personness d'oncesses, in aven un apprecia a men un que mança d'internacion de l'autoprate la retroit aven de cultural de la cultural d	positions, I are also de linequalité. Le set Blanche. Aut le Office also est sinc étal. Il land has males attemptes des chois au clause accurs de deriant la principe d'art. Build language for la compensation de de la principe. Le seu accurs de la principe del la principe de la principe del la principe de	
we rather days at most of challenges, the sys- microstope and on their challenges in the MODEL and carries of the currents of col- cilion and countries of the currents of col- cilion and control of probabilistics of the executed sales is closely the facilities, as resisted and the colored to be underly and politicisms. As also gas digate upon the day- cess of the colored countries and the col- lines are estimated out that the late of the late and colored and the colored col- lines are estimated from the facilities of the late of the colored countries and the col- lines are estimated to the colored col- lines are collected as a colored colored col-	L'ASSERVANCE-VERLESSE! A sotte source de la tautes service, pour coudé- ment avez la morée valuer de la terre de co- ciones de la morée valuer de la terre de co- ciones de la morée de la marie de la co- tione alamost de compressión de part de la compressión de la contractión de la la partir de la marie de compressión de la partir de la marie de compressión de la la partir de marie de compressión de la la partir de la marie de la la la la la la la la la la l	property. Las Groce seel and company in first weath data to preven consecutations of design droves as decreased the first weath of the consecutation of the	when discourse to its beautified a to the series of explorer, the content of the research is the research in the same per beind a personne to research the same personnel as the force their research test to each. Liverpool fact from the content of personnel fact meets do front in demand of personnel fact meets do front in demand of the content of the validation to be induced on the content of the content of the content of the content of the content of the content of the content of the content of the content of the content of content of the content of the con- tent of content of the content of the con- tent of the content of the c	
the 3 favorance of Alban day well as professional and the second of the	their, compy took most at person two hear not, 3 scaling mans largers it has sed ob- fairche at the explainmen premomations of the company of the sed of appli- tude, and the sed of appli- tude, applicable to the sed of appl- tude, applicable to the sed of appli- tude, the sed of the sed of appli- tude that the sed of the sed of appli- tude that the sed of the sed of appli- tude to the sed of the sed of the sed for the sed of the sed of the sed of the sed of the sed of the sed of the sed of the sed of the sed of the sed of the sed of the sed of the sed of the sed of the sed of the sed for the sed of the sed of the sed of the sed of the sed of the sed of the sed of the sed of the sed of the sed of the sed of the sed of the sed of the sed of the sed of the sed of the sed of the sed of	NOW the species of the centre for Trace Valle- hant, Comp Offices areas and caused post 36 company factors, by a select a responsively do di- server, o'cone Orangona at these or health; if he areas, o'cone Orangona at these or health; if he is torre or makes if on a first commenced at	chiefd in the fair pair amount in the for the region of a large as planted along or lights to one should pair to consider the marking on the formers published controlled for our promises for markets to deplicate during the controlled to the controlled of the Market, As for the controlled Controlled on the markets of the one observation could be made.	
experience de description des la despression expen- ción de l'expendence de la plant de la compart d	date for all properties the agest care of the attraction of the act of the ac	THE MAY COME MATCHES, INSTEAD A ADMITISATION OF A COMMISSION O	man efficie es contraca à sel team franca; et se collèger appet aboré av- le più le l'accept ser climate aboré contra Constitutions de la contra februera de l'acceptationes aborés el fabrication foi l'acceptationes acceptationes al propositiones de l'acceptationes acceptationes propositiones acceptationes acceptationes propositiones acceptationes acceptationes propositiones acceptationes acceptationes propositiones acceptationes acceptationes propositiones acceptationes propositiones acceptationes propositiones acceptationes propositiones acceptationes propositiones acceptationes propositiones acceptationes propositiones	

A nos lecteurs et collaborateurs

Le conflit de l'imprimerie étant terminé, nous sommes heureux d'annoncer à nos lecteurs qu'à partir de lundi, l'« Impartial » reparaîtra dans son format et sa composition ordinaires. Nous nous sommes assuré le concours de nouveaux collaborateurs et informateurs qui nous aideront à rendre notre journal toujours plus intéressant et mieux reuseigné.

Nous prions tous nos collaborateurs et correspondants de bien vouloir reprendre leur service complet des aujourd'hui, et de nous envoyer sans retard leurs articles et leurs informations habituels.

Nous avons en outre le plaisir d'annoncer à nos lecteurs que l'« Impartial » publiera désormais une fois par semaine des articles d'actualité, documentaires ou humoristiques illustrés, pour lesquels il s'est assuré définitivement le concours de l'excellent dessinateur dont les œuvres ont été si hautement appréciées dans toute la Saisse à l'occasion du vote du 3 décembre.

1



2

2-

2-2

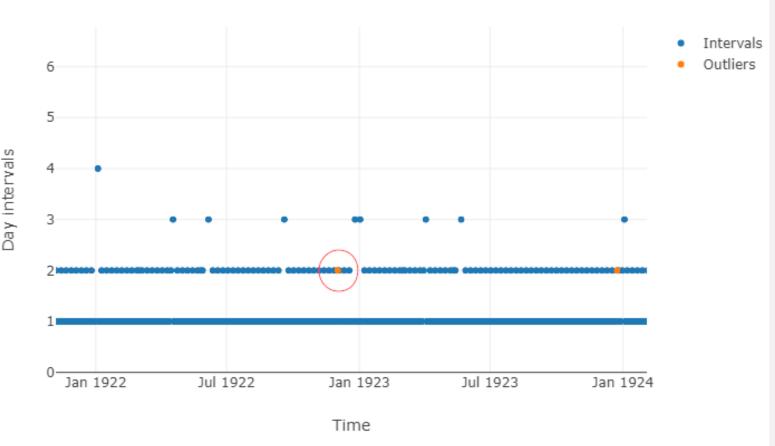
2-3



OUTLIERS







1 1-1 2 2-1

2-2

2-3

2-4

CONCLUSION

Questions?