



HAL
open science

LectAuRep : Données d'archives en français des XIXe et XXe siècles

Hugo Scheithauer

► To cite this version:

Hugo Scheithauer. LectAuRep : Données d'archives en français des XIXe et XXe siècles. Transkribus / eScriptorium : Transcrire, annoter et éditer numériquement des documents d'archives, Bibliothèque nationale de France, May 2022, Paris, France. hal-03666884

HAL Id: hal-03666884

<https://inria.hal.science/hal-03666884v1>

Submitted on 12 May 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

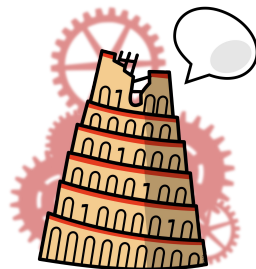
LectAuRep

Données d'archives en français des XIXe et XXe siècles

Transkribus / eScriptorium : Transcrire, annoter et éditer numériquement des documents
d'archives (09/05/22)

Hugo Scheithauer (Inria)

Inria



LectAuRep

Inria - Archives nationales (Département du Minutier Central)

- Faciliter et massifier la consultation des répertoires de notaires conservés au DMC grâce aux technologies de reconnaissance d'écriture manuscrite.
- Offrir aux usagers une lecture enrichie des répertoires et permettre la fouille de données dans le corpus.

Voir Alix Chagué, Rostaing Aurélia. LECTAUREP: Paris Notary Record Books Automated Reading. Fantastic Futures 2021 / Futures Fantastiques 2021, AI4LAM; BnF; Université Paris Saclay, Dec 2021, Paris, France.

N ^{os} DU RÉPERTOIRE	DATES DES ACTES	NATURE ET ESPÈCE DES ACTES :		NOMS, PRÉNOMS ET DOMICILES DES PARTIES INDICATIONS, SITUATIONS ET PRIX DES BIENS	RELATION DE l'Enregistrement.		
		EN BREVETS	EN MINUTES		DATES	DROITS	
An 1911 , mois d' Août							
733	18	Mainlevée	(Subst. M ^e Moreau)	Sebourg (par Maria Rigaudat N ^o d'Amable) d'Asnières, routes, des bois arrières les mains de l'ancien	4	"	"
734	18	(Subst. M ^e Moreau)	Quittance	Sebourg à la d. dame par Demoucheux, d'Asnières, de 11106 ^f 73	"	"	"
735	18	(- 2 ^e -)	Dépôt de décharge	Dubuffet demandeur par Paul Lucien, sa femme Bayonne, à M. M. Lepur et Brochelain	"	"	"
736	18	(Subst. M ^e L. Baudrier)	Dépôt de testament	Devillers de Marie Joseph f ^e divorcée, à Paris, rue Viollet la Rue 5	"	"	"
737	19	Procuration		Guignan (par Hermann) et Gabrielle Françoise Pauline Faure, ép ^{ux} à Paris. Bd ^e Magenta, 118, se acquies par licitation un immeuble à Bordeaux, rue Lagrange, 99	19	3	75
738	19		Obligation	Sterckeman (par Jules Charles Bertin Lucien) et Lucie Aimée Secusble, ép ^{ux} à Paris, Bd ^e Lames, 31 ^{er} et Adolphe Weil, rue Beaurepaire 27, de 22000 ^f .	22	350	"
739	19		- 2 ^e -	- 2 ^e - (Charles mêmes) à Gabrielle Zoé Marie Henriette Bimelle, V ^e de Lion Marcellin Albert Petit, à Paris, faubourg St-Denis, 210, de 30000 ^f .	22	375	"
740	19	(Subst. M ^e Moreau)	Décharge de legs	Varcille (par Marie) f ^e divorcée, 33 Bd ^e de la Seine, à la Noue Louis Jean Alfred Renaud	"	"	"

Extrait d'une page d'un répertoire de notaire.

Transcrire les répertoires de notaires

N ^{os} DU RÉPERTOIRE	DATES DES ACTES	NATURE ET ESPÈCE DES ACTES :		NOMS, PRÉNOMS ET DOMICILES DES PARTIES INDICATIONS, SITUATIONS ET PRIX DES BIENS	RELATION DE l'Enregistrement.		
		EN BREVETS	EN MINUTES		DATES	DROITS	
An 1911 , mois d' Août							
733	18	Mainlevée	(Subst. M. Moreau)	Lebon (par Maria Papaudat l'éd. Amable) d'Asnières, pour autres, de saise. arrières les uns de l'Autre	4	"	"
734	18	(Subst. M. Moreau) Quittance		Lebourg (à la d. dame) par Beaumont & Co, d'Asnières, de 4406 ^{fr} 73	"	"	"
735	18	(2 ^e) Décl. de décharge		Dubuffet demandeur par Paul Lucien, s. f. d. à Bayonne, à M. M. Lefeur et Bachelain	"	"	"
736	18	(Subst. M. L. Baudrier) Décl. de testament		Devillers (de l'Élie Joseph) f. divorce, à Paris, rue Vollet la Duc 5	"	"	"
737	19	Procuration		Guignan (par Hermann) et Gabrielle Françoise Pauline Faux, ép. à Paris. Bd Magenta, 148, pr. acquisition par licitation un immeuble à Bordeaux, rue Lagrange, 99	19	3	75
738	19	Obligation		Streckeman (par Jules Charles Bertin Lucien) et Lucie Amélie Levasseur, ép. à Paris, Bd James, 21 ^{er} à Adolphe Weil, rue Beaurepaire 27, de 28000 ^{fr} ..	22	350	"
739	19	- 2 ^e -		- 2 ^e - (par les mêmes) à Gabrielle Zoé Flora Henriette Brunelle, V ^e de Lion	22	375	"
740	19	(Subst. M. Moreau) Décharge de legs		Marcellin Albert Petit, à Paris, faubourg St Denis, 210, de 30000 ^{fr} ..	"	"	"
741	19	Inventaire		Vareille (par Marie) f. divorce, 32 Bd de la Chapelle, à la H ^o Louis Jean Alfred Renault	"	"	"
				Garnier (après décès de M. l'abbé Henri Joseph f. B ^{te}) pr. vic. de l'Église St Genevieve des Grandes Carrières, 2 ^e à Paris, rue Champoumet 174, d'après le not. Richal le 20 juillet 1911	21	7	50



S ET DOMICILES DES PARTIES

SITUATIONS ET PRIX DES BIENS

RELATION
de
l'Enregistrement.

Nos	DATES	Noms	NATURE ET ESPÈCE	SITUATIONS ET PRIX DES BIENS	RELATION de l'Enregistrement.	
					DATES	DROITS
292	15	P.V. o.u.v. Col.			22	-
293	15	Inventaire			22	35 -
294	16	Vente			23	1747 04
295	17	Certif. sign.			23	35 -
296	17	Promission			23	35 -

An 1939, mois de Mars

Bajul au décès de Désiré Alexandre Marais à Romaniville
 get de mes Oserais arrivé le 10⁸ 1938

Machinau décès de Dieudonné Joseph Adolphe à Paris 6 rue
 des Batignolles y arrivé le 22. 12-1938 ep. de Henriette Bugéine
 Moutigny

Perrin et Anderson à Jules Menny à Dravey 11 rue Darcy
 de 197⁴ 02 terrain etoile Dravey n° 23 du plan S^m A. P. 40110
 pris 9756,65

Baer court Berthe Isabelle Henriette Bourdin ep. de Richard
 Philippe Anché à Paris 10⁸ de Courcelles

Coté Raymond à Paris 6 rue S'Amsterdam en l.p.
 l. venche

- Deux modèles génériques avec un taux d'erreur par caractère inférieur à 10%.
- Des modèles spécialisés qui résultent de l'affinage des modèles génériques sur des lots de données plus petits et plus uniformes, atteignant un taux d'erreur par caractère égal voire inférieur à 5%.

Conserver la structure logique des répertoires

Nos	DATES	NATURE ET ESPÈCE DES ACTES :		NOMS, PRÉNOMS ET DOMICILES DES PARTIES	RELATION DE	
		BREVETS	EN MINUTES		Enregistrement	DROITS
				INDICATIONS, SITUATIONS ET PRIX DES BIENS	DATES	
505	31		Donation	An 1901, mois de Mai		
506	31		Dr.	Yer (par Louis Chédoré) à Paris, rue Beauregard 61, à Louise Chédoré, sa femme	"	"
507	31		Notoriété	Dr. (par Mlle) son à son mari	"	"
508	31	Procuration		Mars 1900 Maries (après décès de Marie Louise Berthe), décédée à Paris (Eaubou) du 22	5	3,75
509	31	Procuration		De Kirschstein (par Léonie Eugénie) à Neuilly s. Seine, rue de Labourville	1	3,75
510		Procuration		Guillaumin (par Charles Joseph) à Neuilly s. Seine, rue de Labourville	7	3,75
511		Certificat de propriété		Mois de Juin 1901 Legenre (après décès de 1500 ^{fr} au nom de Rose Frot, née Martial Paul Eugénie)	1	Judic
512		Procuration		Dr. (après le décès du dit) à Levallois Perret, rue du Bois 65 du 11 ⁸ arr ¹⁰ 1900	1	3,75
513				Oubaseq (par Justine Adèle Robinson, née de Siles) à Paris, rue de Valenciennes	4	3,75

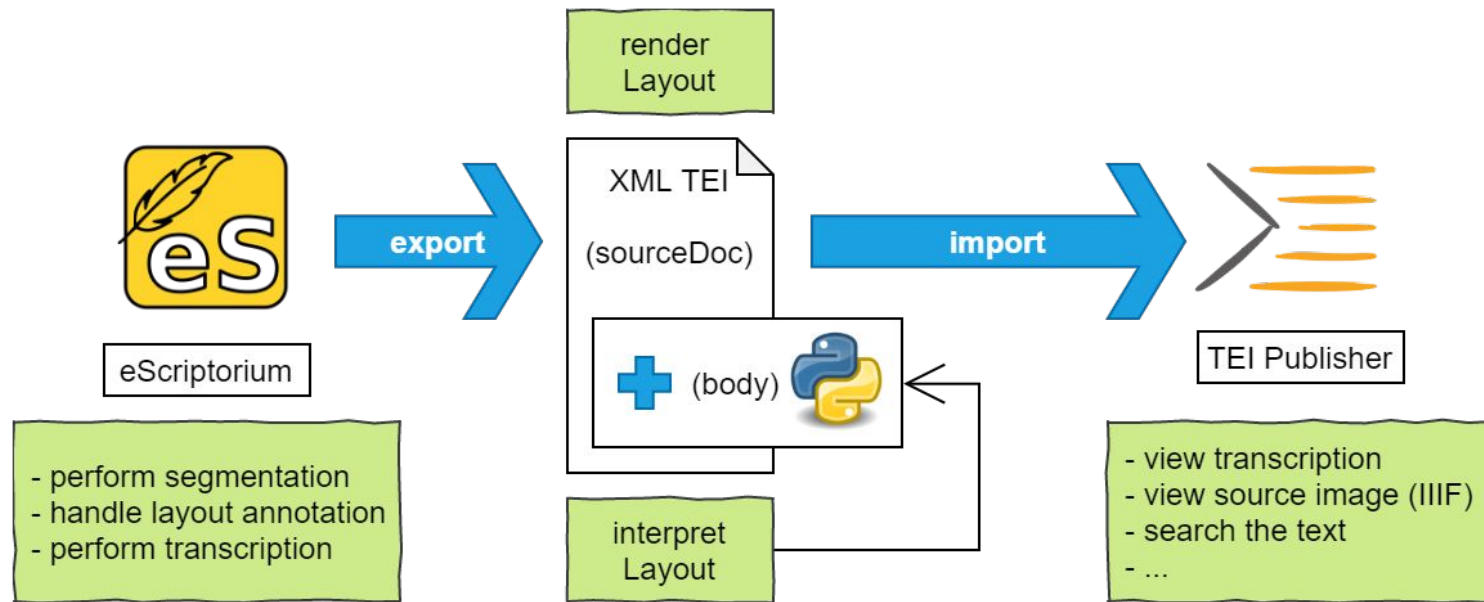
Détection des lignes de texte dans un répertoire de notaire

56 Procuration
57 Audtorisation
58 Certificat de ppte
59 Procuration
60 Mainlevée
61 de signification
62 Mainlevée
63 Mainlevée
64 Mainlevée
65 Mainlevée
66 Obligation
67 Dépôt
68 Transport
69 Vente
70 Constitution
71 deSequestre
72 Décharge
73 Procuration
74 Substitution
75 Quittance
76 Vente
77 An 1914, mois d. Fevrier
78 à Paris 34 rue Sedaine
79 Lamour (par Jn B^e) employé à la Caserne de la G^de Républicaine
80 à Paris rue St Etienne du Mont en blanc pour recueillir SS^on
81 Boettcher (par Henry Emile) Ing^t conseil à Paris 39 Bd St Martin
82 à Léontine Jeanne Blanchot safe dt avec lui pour gérer Comce
83 Boileau-Desombres (parlaSté) à Paris 73 rue Lafayette d'inxn c/
84 Sophie Marie Louise Dessoliers épse de Edouard Lavelaine de Maubeuge
85 Gront (par Jean Eugène André) direstt d'usme à Bahigny route des
86 Petits Ponts à André Valern Fortin ind^el à Paris 34 rue Sedaine

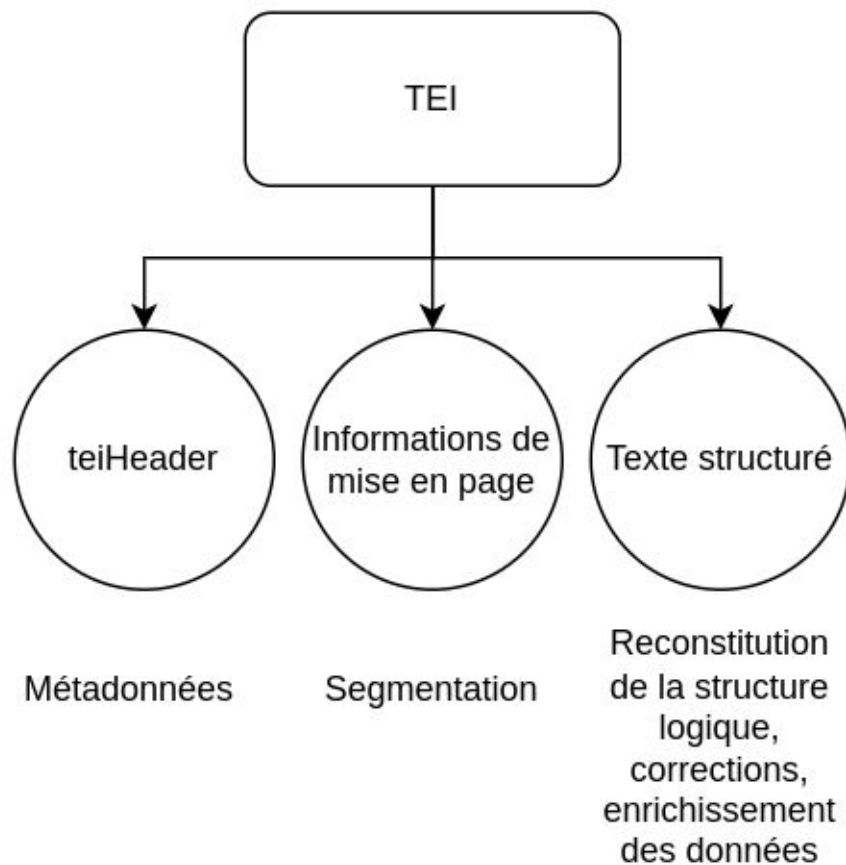
Extrait d'une transcription à plat,
exportée en texte brut.

N ^{os} DU RÉPERTOIRE	DATES	NATURE ET ESPÈCE DES ACTES		INDICATIONS, SITUATIONS ET PRIX DES BIENS	RELATION DE l'Enregistrement.	
		ARTICLES	EN BREVETS		DATES	DROITS
1358	8			An 1927, mois d'Août Baladilhe (concernant 5 extraits d'inscrip ^{on} au total de 1419 ^x de rente f ^{se} 3% au même nom que ci-dessus	11	22 50
1359	8			Baladilhe (concernant divers extraits d'inscrip ^{on} de rente f ^{se} au nom de Lefebvre - Desvallières Jeanne Elisabeth Georgina, f ^e de Emile / surnommé	11	22 50
1360	8			Baladilhe (concernant divers extraits d'inscrip ^{on} de rente f ^{se} au même nom que ci-dessus	11	22 50
1361	8	(son M ^e Delarue)	Mainlevée	Blonde et Boigaud (par la S ^{te} / de Paris 14 rue de Nâcon - d'inscrip ^{on} cf Raphaël Boigaud et H ^e ép ^{te} de Paris 24 f ^{ce} du Marché St Honoré		
1362	8	Procuration	(son M ^e Delarue)	Wathoy (par Fernand / de Paris 113 rue du Chemin Vert en blanc - pour vendre		
1363	8	Mainlevée	d ^o	Tricotel (par Georges / de Paris 8 bd Poissonnière - d'inscrip ^{on} cf André Lécraud et H ^e ép ^{te} de Paris 207 rue de Crimée		

Détection des différentes régions de texte dans un répertoire de notaire



La chaîne de traitement LEPIDEMO



N ^{os} DU RÉPERTOIRE	DATES	NATURE ET ESPÈCE DES ACTES		INDICATIONS, SITUATIONS ET PRIX DES BIENS	RELATION DE l'Enregistrement.	
					DATES	DROITS
1358	8	Certificat de pp ^{te}		An 1927, mois d'Avril Baladilhe (concernant 5 extraits d'inscriptions au total de 1419 ^x de rente 3 ^o / ₁₀₀ au même nom que ci-dessus	11	22 50
1359	8	d ^o		Baladilhe (concernant divers extraits d'inscriptions de rente 3 ^o / ₁₀₀ au nom de Sébastien Desvallières Jeanne Elisabeth Georgina, f ^e de Emilie / subnomme	11	22 50
1360	8	d ^o		Baladilhe (concernant divers extraits d'inscriptions de rente 3 ^o / ₁₀₀ au même nom que ci-dessus	11	22 50
1361	8	(son N ^o Delarue)	Mainlevée	Blonde et Boigaud (par la 3 ^o / ₁₀₀ de Paris 14 rue de Mécon d'inscriptions et 4 ^o / ₁₀₀ de Paris 24 p ^o du Marché St Honoré		
1362	8	Procuration	(son N ^o Delarue)	Wathiez (par Fernand / de Paris 113 rue du Chemin Vert en blanc pour vendre		
1363	8	Mainlevée	d ^o	Tricotel (par Georges / de Paris 8 bd Poissonnière d'inscriptions et André Lévraud et 4 ^o / ₁₀₀ de Paris 207 rue de Crémier		
1364	9	Certificat de vic	d ^o	Bourcier (concernant Louis / de Paris 43 imp ^o du Poussieu		

Reconstituer des unités logiques pour signaler l'information avec des traitements par règle ou avec de la reconnaissance d'entités nommées.

Numéros du répertoire	Dates des actes	Actes en brevets	Actes en minutes	Noms, prénoms et domiciles des parties ; indication, situations et prix des biens	Date de l'enregistrement	Droits de l'enregistrement	Autres
292	15	P. V. ouv. Cof.		Bajul au décès de Désiré Alexandre Marius à Romainville 7 et 9 rue des Oseraies arrivé le 10 8 ^{bre} 1938	22	-	
293	15		Inventaire	Jadin au décès de Dieudonné Joseph Adolphe à Paris 6 rue des Batigolles y arrivé le 22-12-1938 ep [^] x de Henriette Eugénie Montigny	22	35 -	
294		16	Vente	Pierrat et Anderson à Jules Menny à Drancy 11 rue Daisy de 297 [^] f 02 terrain étoile Drancy n°203 du plan XXX A. P n°110 prix 9756,65	23	1747 04	
295		17 Certif Sign.		Baer conct Berthe Isabelle Henriette Bourdin ep [^] se de Richard Philippe André à Paris 20 B [^] d de Courcelles	23	35 -	

Conclusion

- La segmentation est essentielle pour bien délimiter des unités de texte, mais également pour comprendre ces unités une fois transcrites au sein de la mise en page.
- La transcription permet d'obtenir la donnée, et plus un modèle sera robuste, plus la donnée sera de qualité.

Déployer eScriptorium à Inria

- 2019 : machine virtuelle, sans carte graphique.
- 2020 : serveur traces6, 2 cartes graphiques, plus de stockage.
- 2021-2022 : CREMMA, 2 cartes graphiques supplémentaires, plus de stockage, plus de mémoire vive.

Merci pour votre attention !