Katabase

In Search of Lost Manuscripts

Ljudmila Petkovic[1,2] Alexandre Bartz[2] Simon Gabay[1]
Matthias Gille Levenson[3] Lucie Rondeau du Noyer

[1]University of Geneva
name.surname@unige.ch

[2]Sorbonne Université
name.surname@chartes.psl.eu

[3]Ecole normale supérieure de Lyon – Casa de Velázquez, Madrid
name.surname@ens-lyon.fr

June 10, 2021
JeRTEH Seminar, Belgrade, Serbia

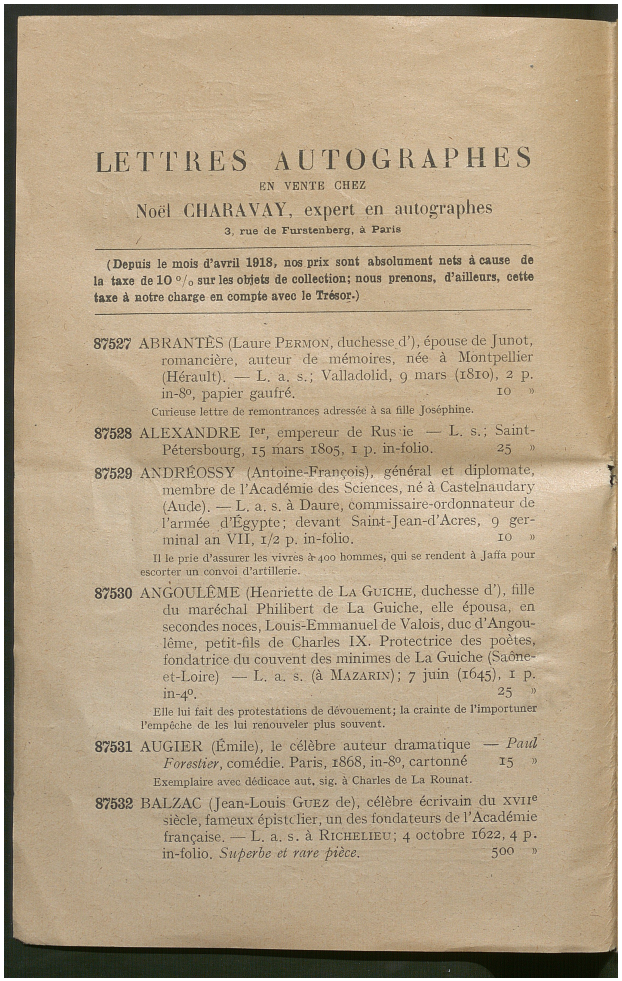
https://katabase.huma-num.fr/

Introduction

Manuscript market

- In Paris, the manuscript market appears c. 1820.
- Manuscript sales catalogues (*i.e.* "mss"):
  - fixed-price (with the price indication);
  - auction (without the price indication).
- Interest for those published by the Charavay dynasty (the brothers Jacques and Gabriel, and their descendants).
- Different types of documents:
  - Mss of great literary value (M^{me} de Sévigné, Voltaire...);
  - Mss of historical value (Napoléon, Robespierre...);
  - And many others (receipts, contracts, documents of forgotten academicians...).
- Formalised entry structure facilitates automatic processing.

Two-fold objective



LETTRES AUTOGRAPHES

EN VENTE CHEZ

Noël CHARAVAY, expert en autographes

3, rue de Furstenberg, à Paris

(Depuis le mois d'avril 1918, nos prix sont absolument nets à cause de la taxe de 10 °/o sur les objets de collection; nous prenons, d'ailleurs, cette taxe à notre charge en compte avec le Trésor.)

87527 ABRANTÈS (Laure PERMON, duchesse d'), épouse de Junot, romancière, auteur de mémoires, née à Montpellier (Hérault). — L. a. s.; Valladolid, 9 mars (1810), 2 p. in-8°, papier gaufré. 10 »

Curieuse lettre de remontrances adressée à sa fille Joséphine.

87528 ALEXANDRE Ier, empereur de Russie — L. s.; Saint-Pétersbourg, 15 mars 1805, 1 p. in-folio. 25 »

87529 ANDRÉOSSY (Antoine-François), général et diplomate, membre de l'Académie des Sciences, né à Castelnaudary (Aude). — L. a. s. à Daure, commissaire-ordonnateur de l'armée d'Égypte; devant Saint-Jean-d'Acres, 9 germinal an VII, 1/2 p. in-folio. 10 »

Il le prie d'assurer les vivres à 400 hommes, qui se rendent à Jaffa pour escorter un convoi d'artillerie.

87530 ANGOULÊME (Henriette de LA GUICHE, duchesse d'), fille du maréchal Philibert de La Guiche, elle épousa, en secondes noces, Louis-Emmanuel de Valois, duc d'Angoulême, petit-fils de Charles IX. Protectrice des poètes, fondatrice du couvent des minimes de La Guiche (Saône-et-Loire) — L. a. s. (à MAZARIN); 7 juin (1645), 1 p. in-4°. 25 »

Elle lui fait des protestations de dévouement; la crainte de l'importuner l'empêche de les lui renouveler plus souvent.

87531 AUGIER (Émile), le célèbre auteur dramatique — *Paul Forestier*, comédie. Paris, 1868, in-8°, cartonné 15 »

Exemplaire avec dédicace aut. sig. à Charles de La Rounat.

87532 BALZAC (Jean-Louis GUEZ de), célèbre écrivain du XVIIe siècle, fameux épistolier, un des fondateurs de l'Académie française. — L. a. s. à RICHELIEU; 4 octobre 1622, 4 p. in-folio. *Superbe et rare pièce.* 500 »

Figure 1: *Lettres autographes et documents*, january 1919.

From the catalogues like the ones to the left of the image, two challenges:

- **Philological**: reconstitute the history of documents;
- **Technological**: transform an image into usable information as quickly as possible.

Data acquisition

Workflow

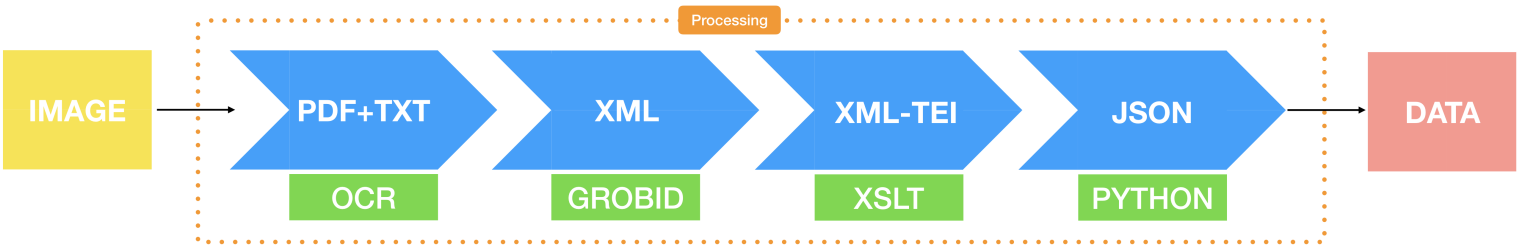


Figure 2: Workflow.

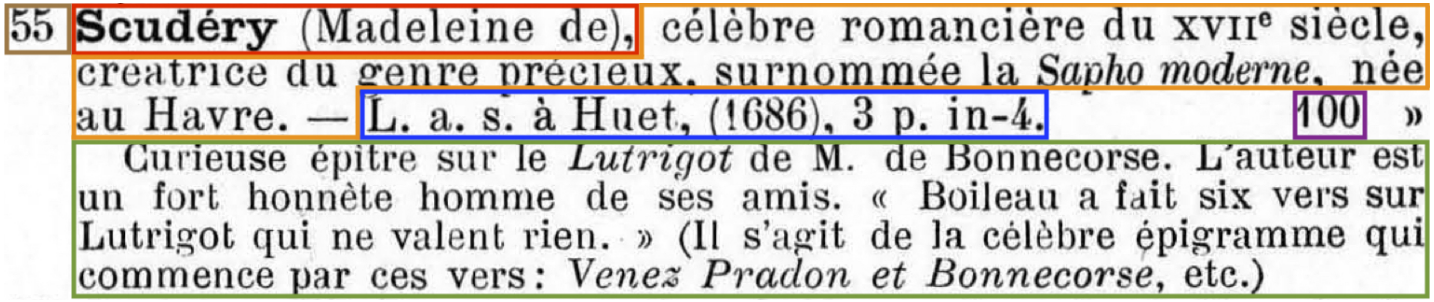
- Use of standard formats (TEI)
- Use of open-source solutions (*GROBID*-dictionaries, *Kraken* in the future)
- Free data publication

[S. Gabay/L. Querciagrossa, "Quantifizierung Des Unbekannten", *Ringvorlesung: Einblicke in die Digital Humanities*, Bern, 2020, *cf.* https://einblicke.hypotheses.org/264]

Data modelling

Modelling

Semantic annotations in order to train *GROBID*


55 **Scudéry** (Madeleine de), célèbre romancière du XVIIe siècle, créatrice du genre précieux, surnommée la *Sapho moderne*, née au Havre. — L. a. s. à Huet, (1686), 3 p. in-4. 100 »

Curieuse épître sur le *Lutrigot* de M. de Bonnecorse. L'auteur est un fort honnête homme de ses amis. « Boileau a fait six vers sur Lutrigot qui ne valent rien. » (Il s'agit de la célèbre épigramme qui commence par ces vers : *Venez Pradon et Bonnecorse*, etc.)

Figure 3: *RDA*, n°67 (March 1881), lot N°55.

- brown: lot n°
- red + typographical information: author's name
- orange: short description of an author
- blue: philological description of a document
- violet: price of a sold ms
- green: additional note (optional)

Expected outcome

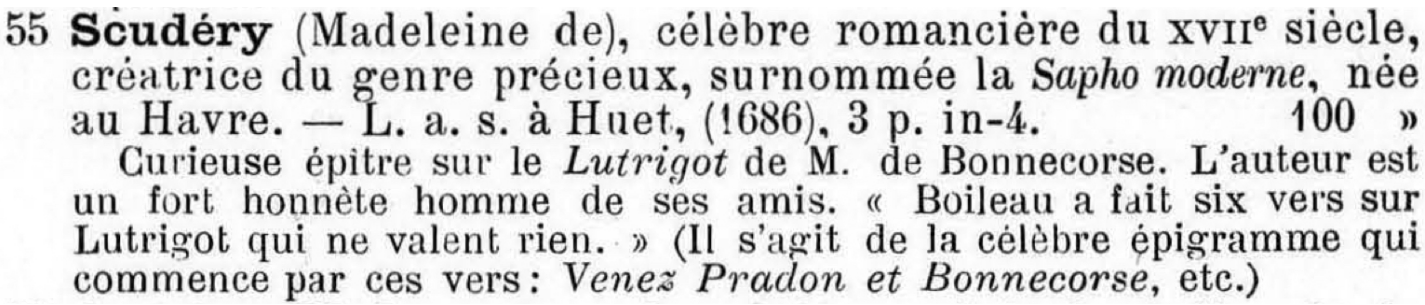
Full semantic XML-TEI encoding for interoperability and sustainability of data.

```
<item n="55" xml:id="CAT_000037_e55">
 <num>55</num>
 <name type="author">Scudéry (Madeleine de),</name>
 <trait>
  <p>célèbre romancière du XVIIe siècle, créatrice du genre précieux,
  surnommée la  Sapho moderne, née au Havre.</p>
 </trait>
 <desc>L.a.s. à Huet, (1686), 3 p. in-4. 100 »</desc>
 <note>Curieuse épître sur le Lutrigot de M. de Bonnecorse. L'auteur
 est un fort honnête homme de ses amis. « Boileau a fait six vers sur
 Lutrigot qui ne valent rien.»  (Il s'agit de la célèbre épigramme
 qui commence par ces vers: Venez Pradon et Bonnecorse, etc.)</note>
</item>
```

OCR

OCR

- Transkribus
- 0.7% CER on 167,985 tokens (17,497 lines)
- Correction of the OCR output



55 **Scudéry** (Madeleine de), célèbre romancière du XVII[e] siècle, créatrice du genre précieux, surnommée la *Sapho moderne*, née au Havre. — L. a. s. à Huet, (1686), 3 p. in-4. 100 »
Curieuse épître sur le *Lutrigot* de M. de Bonnecorse. L'auteur est un fort honnête homme de ses amis. « Boileau a fait six vers sur Lutrigot qui ne valent rien. » (Il s'agit de la célèbre épigramme qui commence par ces vers : *Venez Pradon et Bonnecorse*, etc.)

Figure 4: *RDA*, n°67 (March 1881), lot N°55.

```
55 Scudéry (Madeleine de), célèbre romancière du XVIIe siècle,
créatrice du genre précieux, surnommée la Sapho moderne, née au
Havre. - L.a.s. à Huet, (1686), 3 p. in-4. 100 » Curieuse
épître sur le Lutrigot de M. de Bonnecorse. L'auteur est un fort
honnête homme de ses amis. « Boileau a fait six vers sur
Lutrigot qui ne valent rien.» (Il s'agit de la célèbre épigramme
qui commence par ces vers: Venez Pradon et Bonnecorse, etc.)
```

Improvement of the processing chain

There are several ways to improve the efficiency of the processing chain:

- using bigram models rather than unigram ones;
- with the additional information, such as typographic emphasis (bold, italic)

The last point led to a complexification of the processing chain[1]:

1. development of the OCR model with the typographical information: bold (<b>) and italics (<i>) tags;
2. automatic tag verification and their correction;
3. conforming the ALTO-XML to *GROBID* – problem;
4. ALTO $\rightarrow$ FO $\rightarrow$ PDF (RenderX XEP Engine) – solution.

[1] https://github.com/ljpetkovic/CatMan_ExhibCat_dataset

New OCR model

- 0.9% CER on 91,990 tokens (11,313 lines)

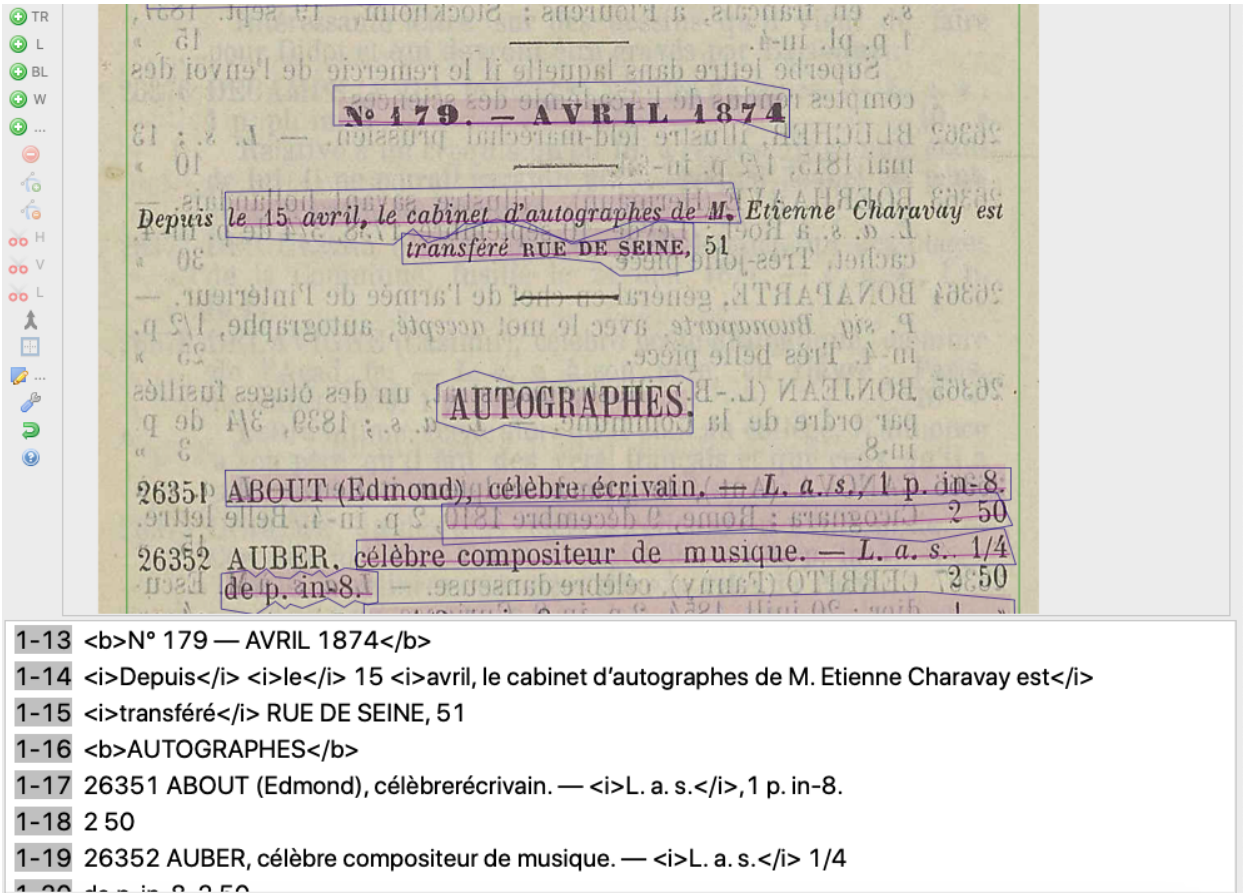


Figure 5: Annotating the catalogues with the typographical information.

Automatic tag verification[2]

In Python and Shell:

1. Check the well-formedness of the tags, whether any open tag is closed, and the tag order.
2. Indicate problems requiring manual correction: `<>foo</>`.
3. Evaluate the OCR performance on three levels (top-down).

The programme's output is shown in three columns, *e.g.*:

- original text line: `<i>Alliot<i>`;
- output code (possible tag scenarios): 3;
- indication of an error or the suggestion for a correction: `MISSING TAGS`.

2 https://github.com/ljpetkovic/CatMan_ExhibCat_dataset/tree/main/scripts/eval_OCR.

Automatic tag verification — statistics

- Three layers:

Type of tags	Count					%				
	LAC_1	LAC_2	LAC_3	LAV_1	LAV_2	LAC_1	LAC_2	LAC_3	LAV_1	LAV_2
Correct	198	469	40	98	155	42.86	20.94	7.09	23.00	19.85
Incorrect	37	224	10	11	22	8.01	10.00	1.77	2.58	2.82
Not automatically correctable	5	55	2	5	10	1.08	2.46	0.35	1.17	1.28
Automatically correctable	32	169	8	6	12	6.93	7.54	1.42	1.41	1.54
No tags	227	1547	514	317	604	49.13	69.06	91.13	74.41	77.34
Initially without problems with the tags	198	469	40	98	155	42.86	20.94	7.09	23.00	19.85
Initially well-formed tags	5	55	2	5	10	1.08	2.46	0.35	1.17	1.28
– Wrong order	0	0	0	0	0	0.00	0.00	0.00	0.00	0.00
– Missing tags	5	55	2	5	10	1.08	2.46	0.35	1.17	1.28
Initially malformed tags	32	169	8	6	12	6.93	7.54	1.42	1.41	1.54
– Well-corrected tags	4	48	3	3	4	0.87	2.14	0.53	0.70	0.51
– Well-corrected tags, bad order	0	0	0	0	0	0.00	0.00	0.00	0.00	0.00
– Well-corrected tags, missing tags	9	39	2	2	4	1.95	1.74	0.35	0.47	0.51
– Well-corrected tags, empty tags	19	82	3	1	4	4.11	3.66	0.53	0.23	0.51

Table 1: The OCR evaluation scores.

Figure 6: The OCR evaluation scores for the *LAC* and *LAV* catalogues.

GROBID-conforming ALTO-XML files[3]

In Transkribus:

- manually correct the non-automatically correctable tags;
- export the ALTO-XML files at the word-level.

In Python and Shell:

1. correct the malformed tags, according to the corrections indicated by the previous script;
   e.g. `<b>Août > <b>Août</b>`;
2. transform them into the *GROBID*-conforming ALTO-XML files:
   - specify the fonts within the `<Styles>` tag
   - remove the redundant tags (`TopMargin, Shape...`)
   - `<String>` needs the incremental `ID`;
   - `<String>` needs the `STYLEREFS="FONT{0,1,2}"`.

3 https://github.com/ljpetkovic/CatMan_ExhibCat_dataset/tree/main/scripts/trans_ALTO.

GROBID

GROBID-dictionaries

Conversion into the TEI with *GROBID-dictionaries* [4]

- Five-level model training; CRF
- Satisfactory results of the models (with the typo info) [5]
  - precision, F-1 $>$ 90% at almost all levels
  - $\approx$ 80% at the `lexicalEntry` level: `<lemma>` et `<sense>`

```
<item n="55" xml:id="CAT_000037_e55">
 <num>55</num>
 <name type="author">Scudéry (Madeleine de),</name>
 <trait>
  <p>célèbre romancière du XVIIe siècle, créatrice du genre précieux,
  surnommée la  Sapho moderne, née au Havre.</p>
 </trait>
 <desc>L.a.s. à Huet, (1686), 3 p. in-4. 100 »</desc>
 <note>Curieuse épître sur le Lutrigot de M. de Bonnecorse. L'auteur
 est un fort honnête homme de ses amis. « Boileau a fait six vers sur
 Lutrigot qui ne valent rien.»  (Il s'agit de la célèbre épigramme
 qui commence par ces vers: Venez Pradon et Bonnecorse, etc.)</note>
</item>
```

[4] https://github.com/MedKhem/grobid-dictionaries.

[5] https://github.com/katabase/GROBID_typo.

Training GROBID with the ALTO-XML files[6]

The main idea is to train **two** *GROBID*-dictionaries models:

- with the typographical information;
- without the typographical information.

Sampling criteria:

- fixed-price catalogues published by Jacques Charavay (*i.e. LAC*) and Auguste Laverdet (*i.e. LAV*)
- the oldest/most recent catalogues
- 4 pages in training, 1 page in validation set

Annotation and training:

- create the training data at five levels of granularity
- train and evaluate *GROBID* after each level

[6] `https://github.com/katabase/GROBID_typo`

Five levels of annotation

1. `dictionary-segmentation`: separate the body
2. `dictionary-body-segmentation`: entry
3. `lexical-entry`: parts of entry
4. `form`: name, author's biography
5. `sense`: ms description, note

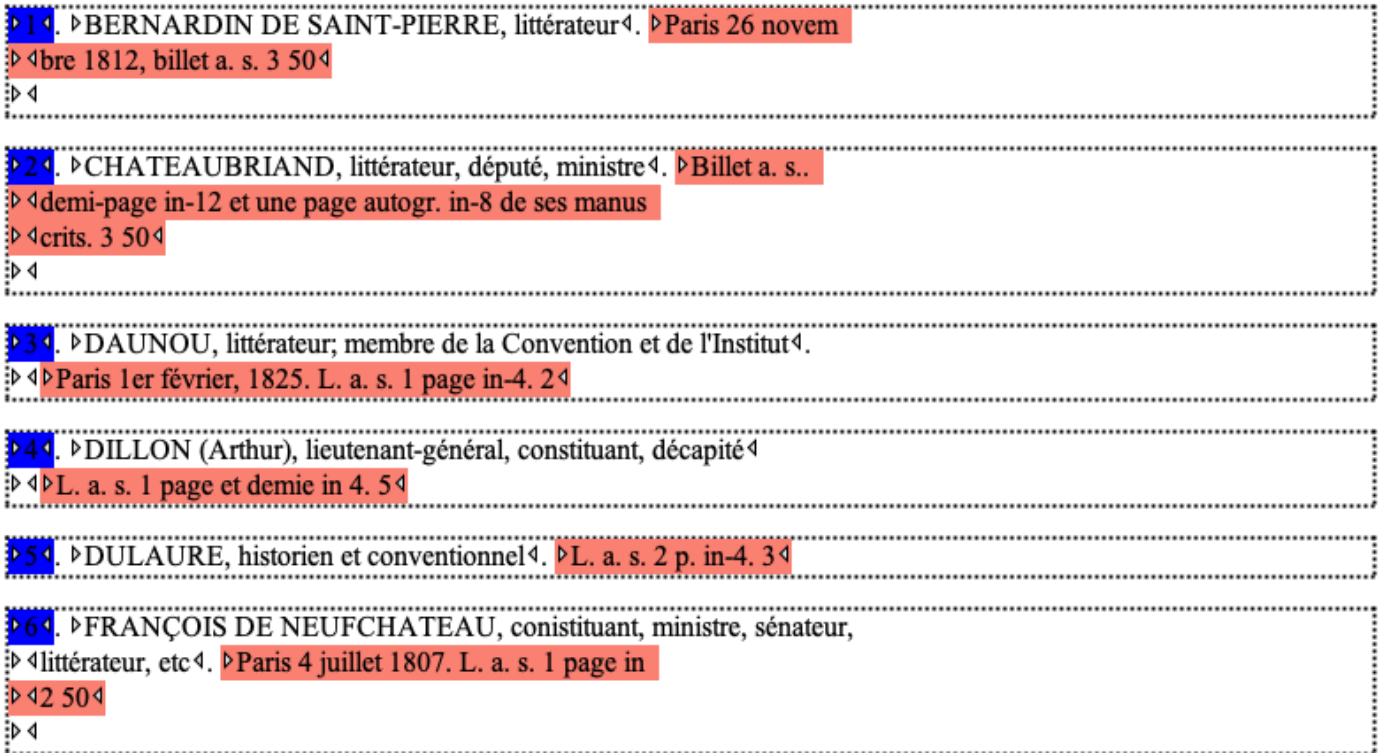


Figure 7: Annotating the catalogues on the `lexical-entry` level.

Evaluating *GROBID*

Without typographical information:

d) `form` :

```
===== Field-level results =====

label                accuracy     precision    recall       f1           support

<desc>               93.33        90.91        86.96        88.89        23
<name>               92           87.5         87.5         87.5         24
<pc>                 97.33        95.45        95.45        95.45        22

all (micro avg.)     94.22        91.18        89.86        90.51        69
all (macro avg.)     94.22        91.29        89.97        90.61        69

===== Instance-level results =====

Total expected instances:    22
Correct instances:           19
Instance-level recall:       86.36


Evaluation for form model is realized in 382 ms
[INFO] ------------------------------------------------------------------------
[INFO] BUILD SUCCESS
[INFO] ------------------------------------------------------------------------
[INFO] Total time:  7.345 s
[INFO] Finished at: 2020-08-23T14:53:08Z
[INFO] ------------------------------------------------------------------------
```

Figure 8: Results for the catalogues *LAC + LAV*.

Evaluating *GROBID*

With the typographical information:

d) `form`:

```
===== Field-level results =====

label                accuracy     precision    recall       f1           support

<desc>               95.65        90.91        95.24        93.02        42
<name>               96.38        95.35        93.18        94.25        44
<pc>                 96.38        91.11        97.62        94.25        42

all (micro avg.)     96.14        92.42        95.31        93.85        128
all (macro avg.)     96.14        92.46        95.35        93.84        128

===== Instance-level results =====

Total expected instances:   45
Correct instances:          41
Instance-level recall:      91.11


Evaluation for form model is realized in 457 ms
[INFO] ------------------------------------------------------------------------
[INFO] BUILD SUCCESS
[INFO] ------------------------------------------------------------------------
[INFO] Total time:  16.376 s
[INFO] Finished at: 2020-08-23T22:09:15Z
[INFO] ------------------------------------------------------------------------
```

Figure 9: Results for the catalogues *LAC + LAV*.

Evaluating *GROBID*

- Better results for the *LAC* compared to the *LAV*
- The results slightly deteriorate with the addition of the new data (more difficult for the model to generalise).
- The results for the *LAC + LAV* with the typographical information are still satisfactory.
- The precision and the F-1 measure scores are above the 90% at almost all levels (except the `lexical-entry` level, where the scores for `<lemma>` and `<sense>` lean more to the 80%).

Post-processing

Post-processing

- Conversion of the lexicographical-like tags (*e.g.* <sense>...) into those required by the catalogues (*e.g.* <desc>...)
- Additional information (standardisation of dates, formats, types of documents)
- Validation by specific schemas throughout the process to ensure data quality

```
<item n="55" xml:id="CAT_000037_e55">
 <num>55</num>
 <name type="author">Scudéry (Madeleine de),</name>
 <trait>
  <p>célèbre romancière du XVIIe siècle, créatrice du genre précieux,
  surnommée la  Sapho moderne, née au Havre.</p>
 </trait>
 <desc>L.a.s. à Huet, (1686), 3 p. in-4. 100 »</desc>
 <note>Curieuse épître sur le Lutrigot de M. de Bonnecorse. L'auteur
 est un fort honnête homme de ses amis. « Boileau a fait six vers sur
 Lutrigot qui ne valent rien.»  (Il s'agit de la célèbre épigramme
 qui commence par ces vers: Venez Pradon et Bonnecorse, etc.)</note>
</item>
```

Figure 10: Result of the XSLT transformations.

Further extraction

Minor adjustments are made thanks to a python script, mainly to extract additional data from the desc:

```
<item n="55" xml:id="CAT_000037_e55">
 <num>55</num>
 <name type="author">Scudéry (Madeleine de),</name>
 <trait>
  <p>célèbre romancière du XVIIe siècle, créatrice du genre précieux,
  surnommée la  Sapho moderne, née au Havre.</p>
 </trait>
 <desc><term>L.a.s.</term> à Huet, (<date>1686</Date>),
   <measure type="length">3 p.</measure> <measure type="length">in-4
   </measure>. <measure commodity="currency">100</measure> »</desc>
 <note>Curieuse épître sur le Lutrigot de M. de Bonnecorse. L'auteur
 est un fort honnête homme de ses amis. « Boileau a fait six vers sur
 Lutrigot qui ne valent rien.»  (Il s'agit de la célèbre épigramme
 qui commence par ces vers: Venez Pradon et Bonnecorse, etc.)</note>
</item>
```

Further extraction

Key information is extracted in json for further data mining and online publication:


```
"CAT_000037_e55_d1": {
    "desc": "L.a.s. à Huet, (1686), 3 p. in-4. 100",
    "price": 100,
    "date": 1686,
    "number_of_pages": 3,
    "format": 4,
    "term": 7,
    "author": "Scudéry",
    "sell_date": "Mars 1881"
}
```

Use cases

Reconciliation

Reconstitution of manuscripts

Over the time, some manuscripts have been cut into several pieces.

207 **Sévigné** (Marie de *Rabutin-Chantal*, marquise de), la célèbre épistolaire. — Fin de lettre aut. à sa fille Mme de Grignan ; aux Rochers, 12 août 1685, 3 p. in-4, suivie de 2 pages aut. d'*Emmanuel de Coulanges*. 200 »

Précieuse pièce où elle parle longuement de son séjour aux Rochers, en compagnie d'Emmanuel de Coulanges, et du prochain départ de ce dernier avec Charles de Sévigné pour les Etats de Bretagne. « Mon fils a une petite lanternerie d'émotion qui l'a empêché d'aller aux Etats. Il prend de cette tisane des capucins que vous connoissez, et dont je me suis si bien trouvée ; il compte cependant de partir demain avec M. de Coulanges. »

265 **Sévigné** (Marie de *Rabutin-Chantal*, marquise de), la célèbre épistolaire. — Fin de lettre aut. à sa fille Mme de Grignan ; aux Rochers, 12 août 1685, 3 p. in-4, suivie de 2 pages aut. d'*Emmanuel de Coulanges*. 200 »

Précieuse pièce où elle parle longuement de son séjour aux Rochers, en compagnie d'Emmanuel de Coulanges, et du prochain depart de ce dernier avec Charles de Sévigné pour les Etats de Bretagne.

Figure 11: Part 1: *RDA*, May 1894, lot n°166 and July 1897, lot n°200

201 **Sévigné** (Marie de *Rabutin Chantal*, marquise de), la célèbre épistolaire, petite-fille de Sainte-Chantal, née à Paris en 1626, morte à Grignan en 1696. — Fragment de let. aut. à sa fille Mme de Grignan, 12 août 1685, 2 p. in-4. *Rare*. Précieuse pièce. 125 »

Figure 12

Figure 13: Part 2: *RDA*, April 1902, lot n°257

A manuscript fingerprint

To reconcile entries or reconstitute manuscripts, we need to compare the items sold. To do so, we used the data contained in the `desc` of each `item`. Such as a social number, using our date or our place of birth, we can create a unique ID for each manuscript with the following information:

- Author's name
- Type of document
- Writing date
- Length
- Format
- Price

Comparing IDs

RDA, May 1894 (N°166)	RDA, July 1897 (N°200)	RDA, April 1902 (N°257)
Sévigné	Sévigné	Sévigné
Fin de lettre aut.	Fin de lettre aut.	Fragment de let. aut.
12 août 1685	12 août 1685	12 août 1685
3 p.	3 p.	2 p.
in-4	in-4	in-4
200 francs	200 francs	125 francs

Table 1: Key information of three sold items from catalogues

Last reconciliation

Some manuscripts that have been bought by libraries, and are now accessible. It is the case for the two manuscripts of Sévigné we presented *supra*, which are now kept in Paris and Princeton.

Princeton, Rare Books and Special Collections, C0710, vol. 4, f°57/N°4	Paris, BNF, NAF 717
Sévigné	Sévigné
Fin de lettre aut.	Fragment de let. aut.
12 août 1685	12 août 1685
3 p.	2 p.
in-4	in-4

Table 2: Key information of two manuscripts

Identifying manuscripts

It is therefore possible to extend the reconciliation process to records of libraries, that we have encoded in TEI.

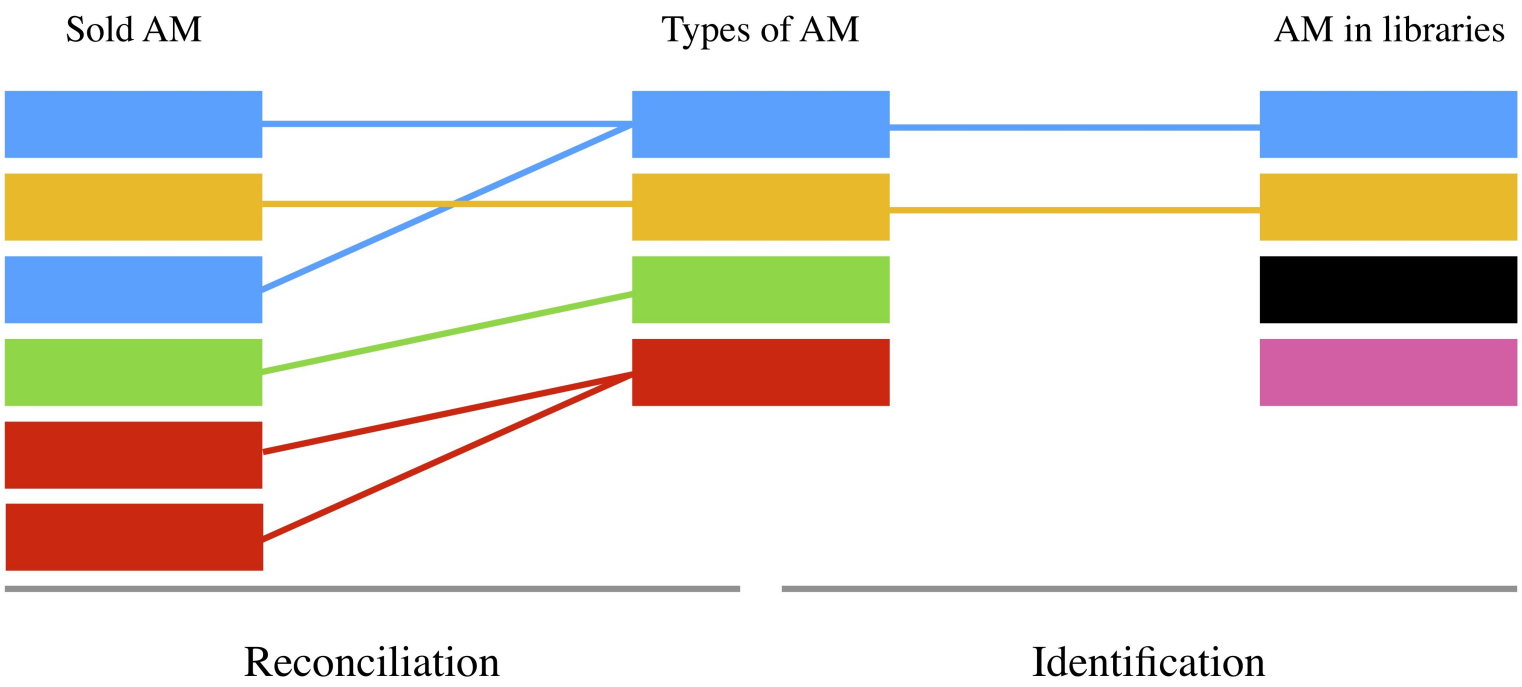
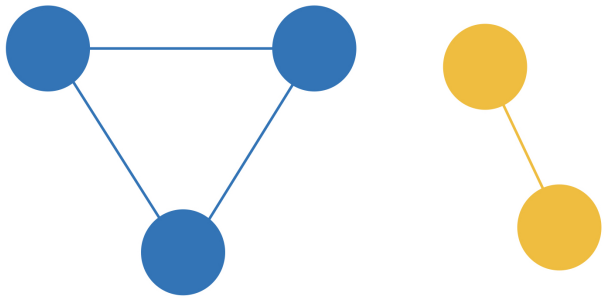


Figure 14: Reconciliation-identification process

From the TEI to a graph


```
"CAT_000127_e12_d1": {
    "price": 22.0,
    "author": "Augier",
    "date": "1845",
    "number_of_pages": 2.0,
    "format": 4,
    "term": 3,
    "sell_date": "1889-12",
    "desc": "Pièce aut.; (1845), 2 p. in-4."
},
```


(a) Catalogue entry encoded in JSON format from the TEI.



(b) Reconciliation mechanism (edges) between manuscripts (points) using triplets.

Figure 15: Data transformation.

Theoretical principle

- The graphs make it possible to simplify the reconciliation mechanism:
  1. we compare the degree of similarity between a ms A (represented by a node) and a ms B (represented by another node);
  2. if the similarity is sufficient we can connect these two nodes by an edge;
  3. the entries are then reconciled.
- We are talking about similarity and not strict identity between two entries/mss entries, indeed:
  - the same ms can deteriorate over time (and therefore its description can change);
  - OCR errors can creep into the transcription (and the description of the ms is therefore partially wrong).

Using weights

After filtering on a single author, all items found are compared one to another, and a score is offered for each pair.

ID	Bonus	Malus
Type of document	+ 0.2	- 0.1
Writing date	+ 0.5	- 0.5
Length	+ 0.1	- 0.1
Format	+ 0.1	- 0.3
Price	+ 0.1	- 0.1

Table 3: Using a malus or a bonus to create a score

Reconciliation algorithm

The similarity between two mss is determined according to modular criteria:

- some, which do not normally vary over time, are important (date, author);
- others, less stable, are less important (pagination, format).

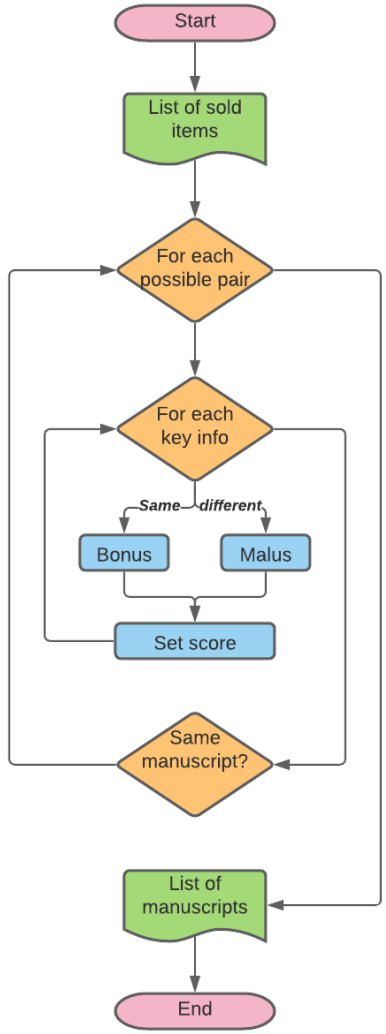


Figure 16: Flowchart of the reconciliation algorithm.

Sample

A first sample have been processed (numbers are still growing):

- 200 fixed-price catalogues of the *Revue des autographes*, directed by Gabriel Charavay.
- 30 fixed-price catalogues of the *Lettres autographes et documents*
- 40 fixed-price catalogues of Laverdet

(All the catalogues date from the 19^{th} c.)

Results : manuscripts of Mme de Sévigné

We can now offer some results:

- 63 sales have been identified until 1903
- 46 AM being sold at least one time, 14 at least two times
- 13 letters out of the 46 sold are not in public libraries or archives

Following these numbers, we can say that:

- c. 1% of the 1,350 letters identified by R. Duchêne are still circulating on the market.
- c. 5% of the total has survived but is inaccessible to scholars, if we add the 62 letters still held in the private collections of the Guitaut family in Burgundy.

Publication

Application: *back end*

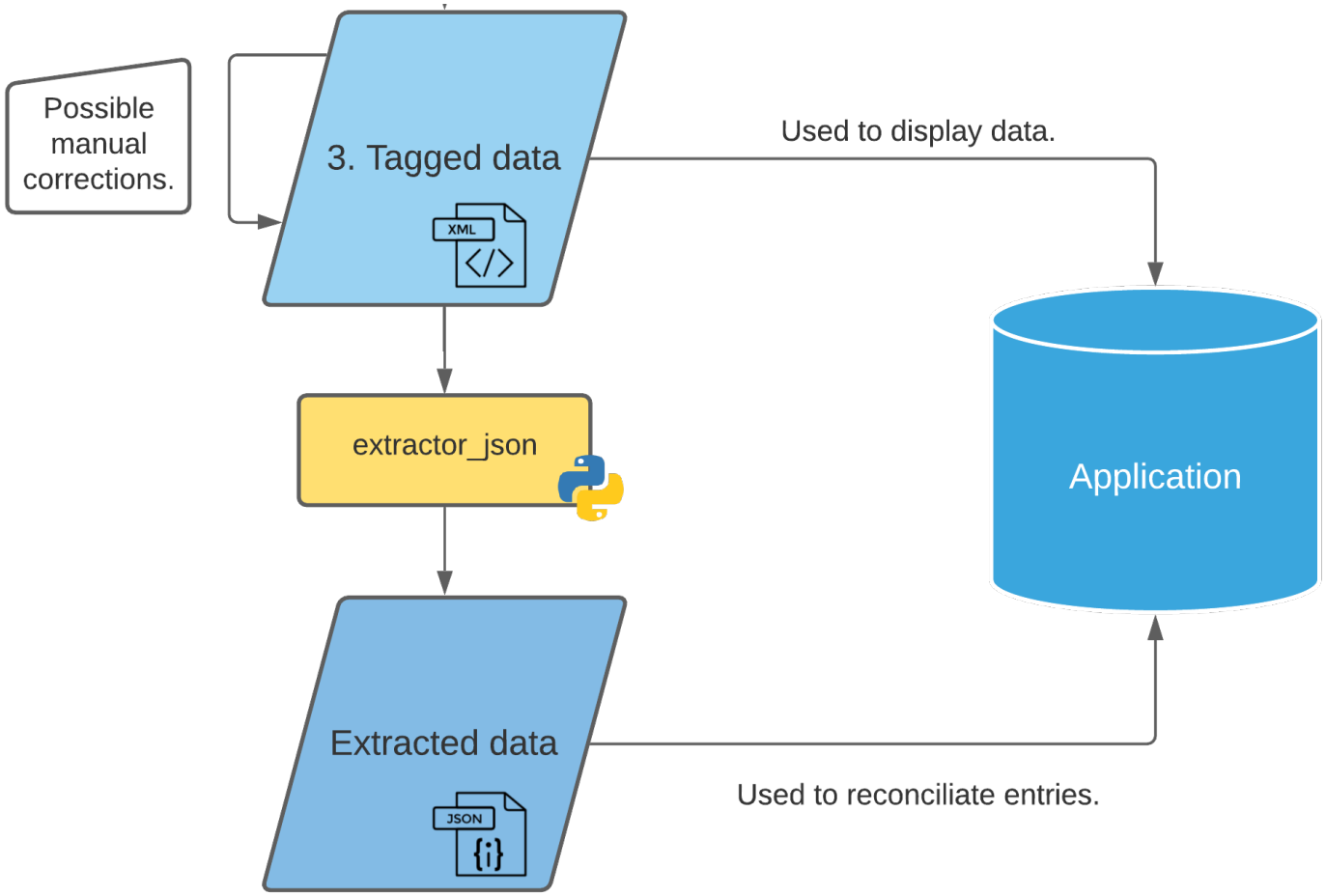


Figure 17: The application *Katabase* works as much with data in `.json` format (for queries) as with those in `.xml` format (for display).

Application: *front end* (reading)

Figure 18: Read mode of the *Katabase* application to read the catalogues.

Application: *front end* (searching)

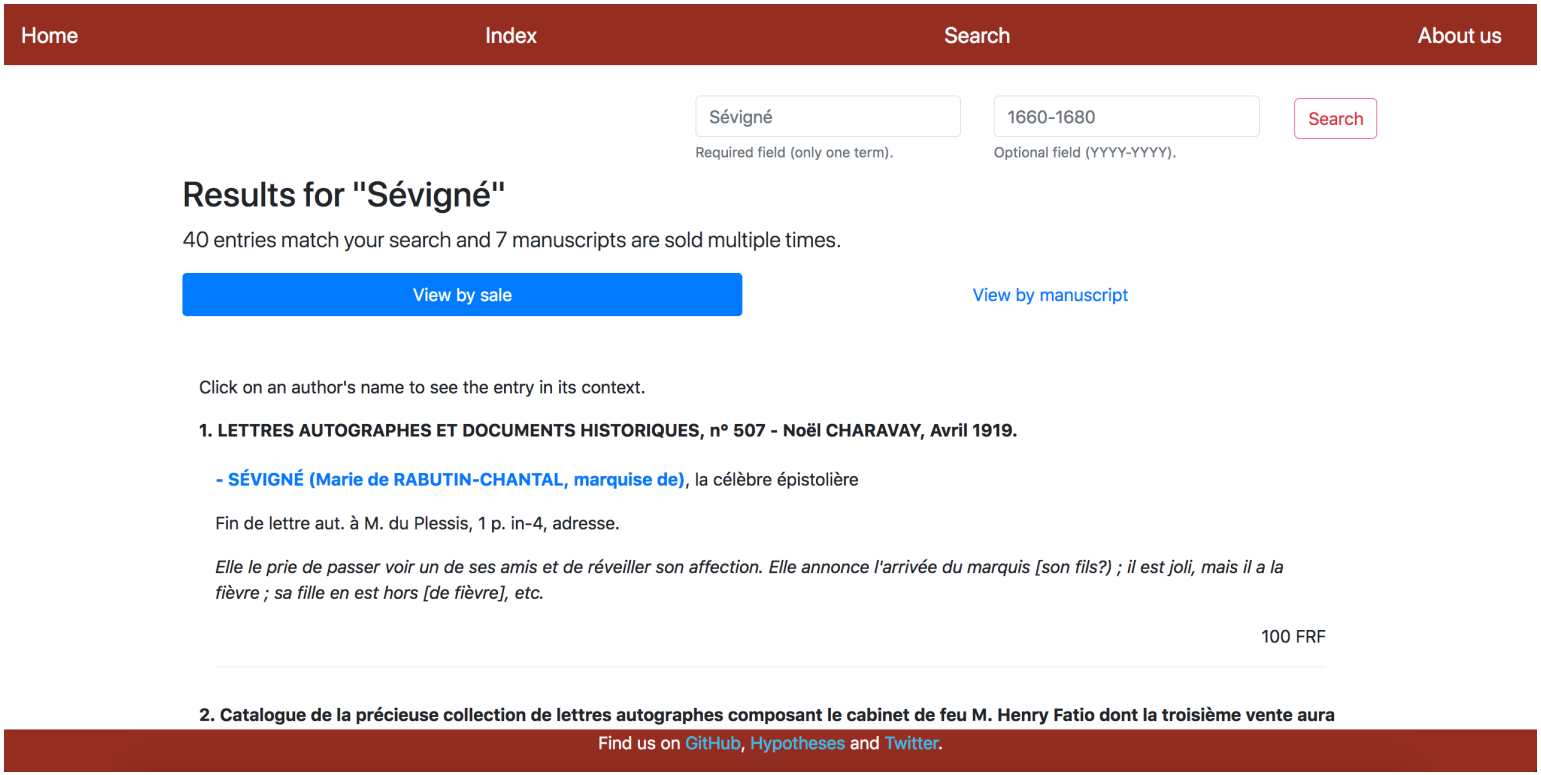


Figure 19: Search mode of the *Katabase* application to find the mss.

Econometrics

First results

- **Number of mss sold**: out of 44,333 mss sold, 3,567 were sold at least twice or several times, ie 7.5% of mss
- **Price evolution**: there is a clear downward trend, as shown by the example of Hugo's mss *infra*.

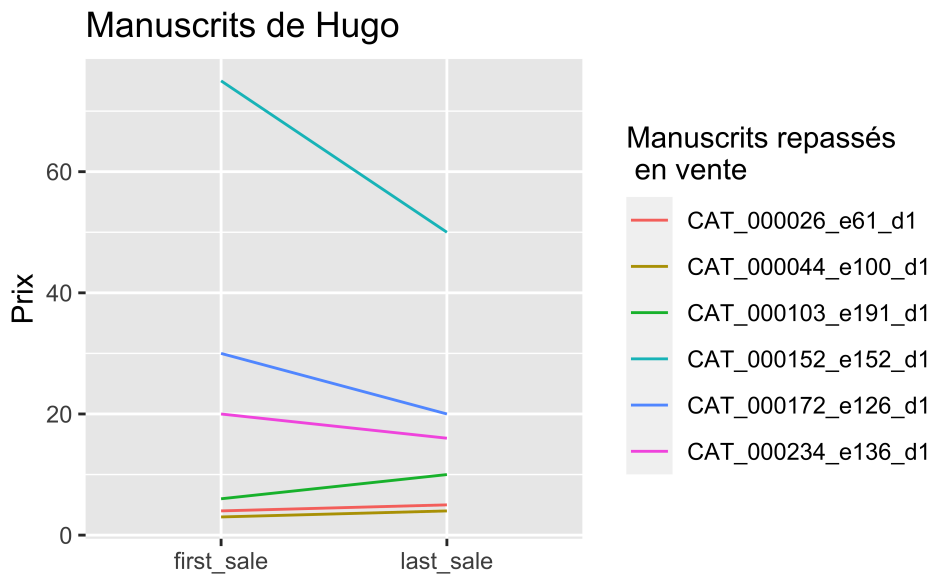


Figure 20: Price evolution of V. Hugo's manuscripts between the first and the last sale.

Conclusion

Conclusion

- Nearly 200 catalogues in preparation:
  1. **123** *auction catalogues* (AUC);
  2. **40** catalogues of *Librairie ancienne et autographes de Charavay* (LAC);
  3. **29** *Catalogues de lettres autographes, manuscrits, documents historiques, etc., d'Auguste Laverdet* (LAV).
- That is, a corpus of more than 400 finely encoded catalogues available to all and freely readable, searchable and downloadable (`https://github.com/katabase`).
- Multiple uses:
  - **philology**: find documents for the editions;
  - **history**: reconstruct the appearance of the market;
  - **literature**: understand the construction of the canon;
  - **economy**: analyse the history of prices.

Acknowledgments

We thank Caroline Corbières, Mohamed Khemakhem, Jean-Baptiste Camps, Laurent Romary, Béatrice Joyeux-Prunel... and you for your attention!