

# Internationnal Conference NOOJ 2021

## University of Franche-Comté



Linguistics, applied research and NLP :  
using NooJ in a technical-operational context. Case-  
study, analysis and perspectives

Nicolas BOFFO



<https://communaute-cimi.fr/>

Philippe LAMBERT



**UNIVERSITÉ  
DE LORRAINE**

[www.univ-lorraine.fr](http://www.univ-lorraine.fr)

June 11, 2021 Intervention

# International Conference NOOJ 2021

## University of Franche-Comté



- I. Introduction
- II. Theoretical axes
- III. Define and conceptualize the technical-operational (or functional) approach.
- IV. Case-study : NOOJ in a technical-operational context
  - (1) Spatio-temporal treatment in a technical-operational (or functional) approach : the case of the Vietnamese press.
  - (2) DATAVIZ : KMS
- V. Conclusion
- VI. Perspectives



# International Conference NOOJ 2021

## University of Franche-Comté



### I) INTRODUCTION

CIMI has chosen to experiment NOOJ in a technical-operational(functional) approach on a poorly endowed language (Vietnamese) with the support of the research engineer Philippe LAMBERT from the University of Lorraine.

Initial question:

Is NOOJ accessible for users working on language data with a need for DATAVIZ and Mapping in technical-operational context ?



UNIVERSITÉ  
DE LORRAINE

# International Conference NOOJ 2021

## University of Franche-Comté



## II) THEORETICAL AXES

### Theoretical linguistic axes :

#### About the discourse :

A praxematic approach.

A sociolinguistic approach (William LABOV).

« Discourse frames and their boundaries » M. CHAROLLES



UNIVERSITÉ  
DE LORRAINE

# International Conference NOOJ 2021

## University of Franche-Comté



### II) THEORETICAL AXES

#### Theoretical linguistic axes :

##### Formalization of temporality:

Logical studies of time and aspect J.L DESCLES.

Approach of J. BRES (cf. " Temps de l'indicatif en Français ").

Semantics of Temporality by GOSSELIN.

Work of Delphine BATISTELLI.

S-Languages of Sylviane D. SCHWER.

TimeML.

Description by Danh Thanh Do-Hurinville in "Temps, Aspect et Modalité en vietnamien" (2009).

Description by Kim Phuong in " Ngu phap tieng viet hung van de thoi, the "(2008).

Linguistic approach of Cao Xuan Hao.



UNIVERSITÉ  
DE LORRAINE



# International Conference NOOJ 2021

## University of Franche-Comté



## II) THEORETICAL AXES

### NLP :

- Formalizing Natural Languages: the Nooj Approach of Max Silberztein
- Works of Delphine BASTISTELLI.



# International Conference NOOJ 2021

## University of Franche-Comté



### III) DEFINE AND CONCEPTUALIZE THE TECHNICAL-OPERATIONAL (OR FUNCTIONAL) APPROACH IN NLP

TECHNICO-OPERATIONNAL (FUNCTIONAL) APPROACH IN NLP (observation of business needs)

What is it ?

Missions

People

Technical

Functionnal

NLP

Linguistic

Quick execution of tasks.

(One-shot task)

Limited time

Limited cost

Possible applications :

Investigations

Geopolitical analysis

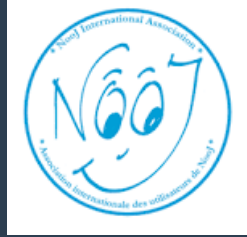
Discourse analysis

OSINT

Etc...

# International Conference NOOJ 2021

## University of Franche-Comté

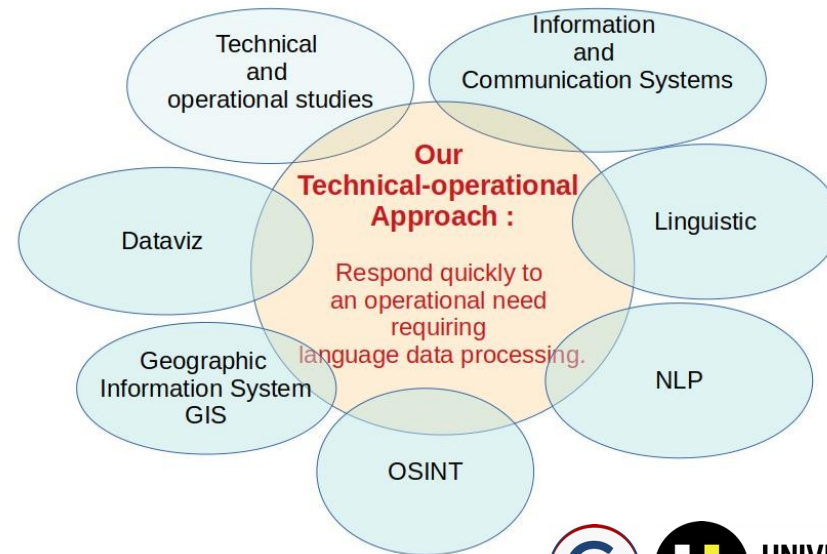


### III) DEFINE AND CONCEPTUALIZE OUR « TECHNICAL-OPERATIONAL (OR FUNCTIONAL) APPROACH »

Break down the silos (Dismantling , « désilotage » ) :



### Scope of our « Technical-operational approach » concept





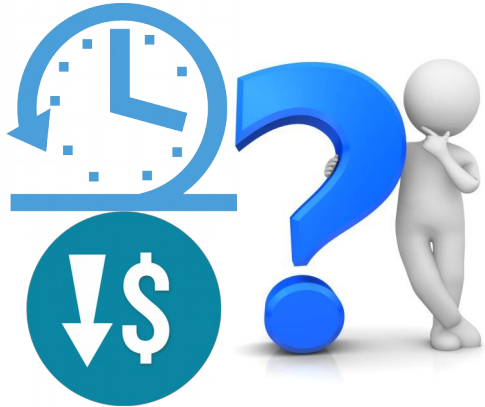
# International Conference NOOJ 2021

## University of Franche-Comté

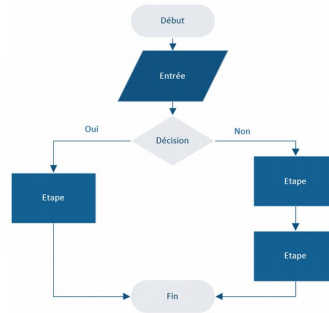


### III) DEFINE AND CONCEPTUALIZE OUR « TECHNICAL-OPERATIONAL (OR FUNCTIONAL) APPROACH »

#### REASONING AND PROCESS OF TECH-OP APPROACH



Solve a problem in a limited time and cost.



Use and develop a toolkit / Reflex cards in the form of a flow chart.



Validation of steps / Analysis/Dataviz/Reporting.



UNIVERSITÉ DE LORRAINE



## **IV) CASE-STUDY : NOOJ IN TECHNICAL- OPERATIONAL (FUNCTIONAL) CONTEXT**

- 1) Choose NOOJ as toolkit**
- 2) Choose a poorly endowed language : Vietnamese.**
- 3) Apply «Reasonning and process of TECH-OP  
Approach »**



## **IV) CASE-STUDY : NOOJ IN TECHNICAL- OPERATIONAL (FUNCTIONAL) CONTEXT**

### **1) NOOJ :**

**« Develop linguistic resources to formalize various linguistic phenomena at the orthographical, lexical, morphological, syntactic and semantic levels, for any natural language »**

# International Conference NOOJ 2021

## University of Franche-Comté



### IV) CASE-STUDY : NOOJ IN TECHNICAL- OPERATIONAL (FUNCTIONAL) CONTEXT

#### 1) Vietnamese as a poorly endowed language :

- Very few annotated corpora.
- Vietnamese is a tonal and monosyllabic language.
- In Vietnamese there are no verbal tenses.
- There are tense, aspect and modality.
- There are aspectual-temporal and modal markers.
- There are also Temporal Frames (according to M. Charolles' definition)





## IV) CASE-STUDY : NOOJ IN TECHNICAL- OPERATIONAL (FUNCTIONAL) CONTEXT

### Temporality in vietnamese :

Đã, đang et sẽ ne sont pas des marqueurs temporels mais aspectuels (Cao Xuan Hao (1998), Danh Thanh Do-Hurinville (2008)).

Ex:

Thẻ đã mua, xin miễn đổi lại. (Once bought, this card is not exchangeable.)

đã = action is accomplished/finished.

Hôm qua khi tôi đang xem TV thì điện thoại reo. (Yesterday, when I was watching TV, the phone rang.)

đang = action is “during”.

Nêu bạn vượt đèn đỏ thì bạn sẽ bị phạt. (Si vous brûlez le feu rouge, vous êtes pénalisé.)

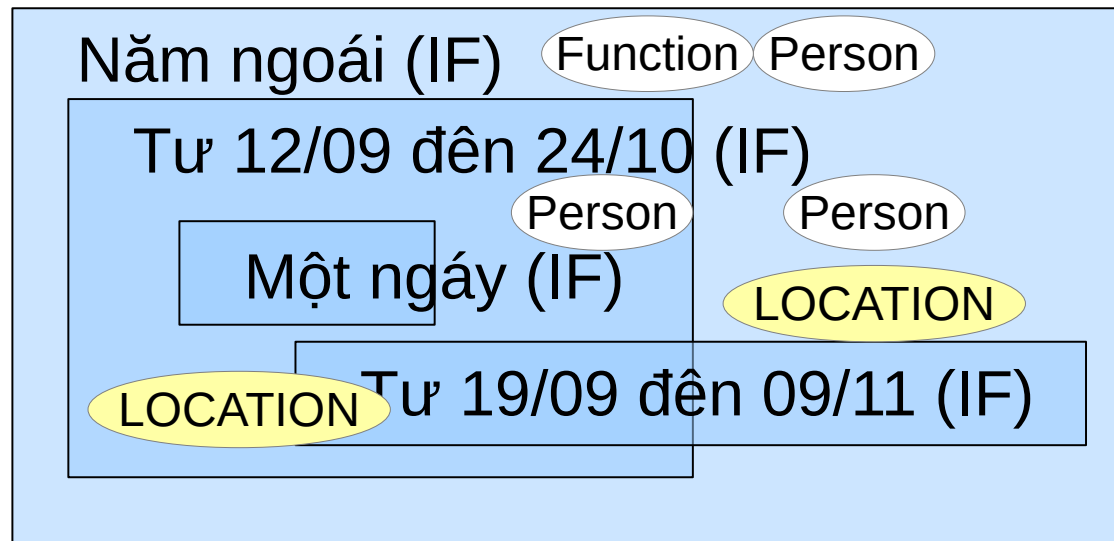
sẽ = action is after.





## IV) CASE-STUDY : NOOJ IN TECHNICAL- OPERATIONAL (FUNCTIONAL) CONTEXT

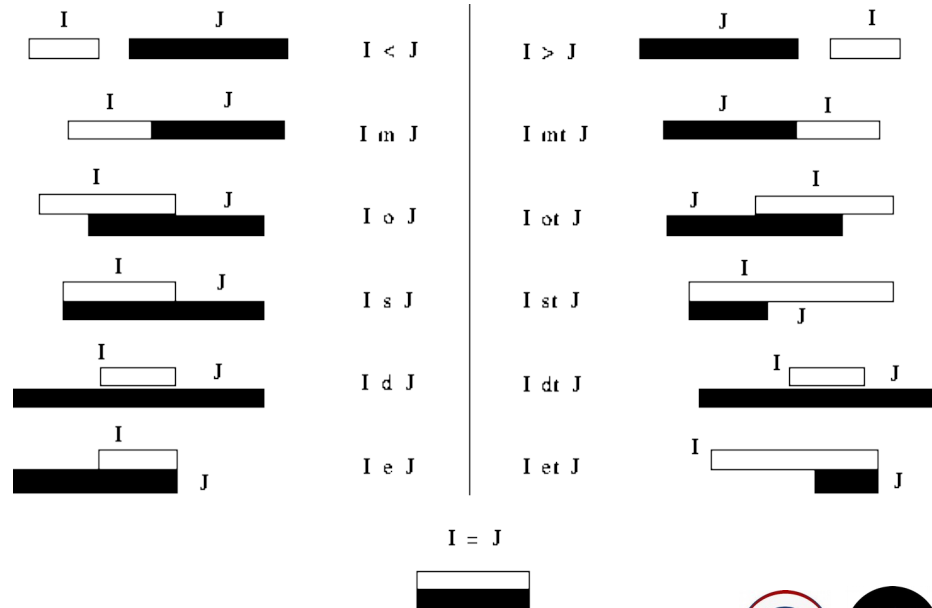
### Temporal frames in vietnamese





### IV) CASE-STUDY : NOOJ IN TECHNICAL- OPERATIONAL (FUNCTIONAL) CONTEXT

- Catégories Temporelles:
- Passé
- Present
- Futur
- Neutre
- 
- Relations temporelles :
- Relations d'Allen
- 
- Catégories Aspectuelles:
- Accompli (đã)
- Progression (đang)
- Iterative (V+lại)
- Terminative (xong, hết, thôi, etc.)
- Imminente (sắp)



# International Conference NOOJ 2021

## University of Franche-Comté



### IV) CASE-STUDY : NOOJ IN TECHNICAL- OPERATIONAL (FUNCTIONAL) CONTEXT

#### Construction of Vietnamese temporal grammars and dictionaries

##### Linguistic work:

- Observation and analysis of aspect and tense.
- Vietnamese grammars and linguistic studies by Tran Kim Phuong, D.T. Do-Hurinville, Thomson, Cao Xuan Hao, Nguyen Van Hue, etc.
- Vietnamese foreign language methods (Hanoi University and Ho Chi Minh City University, etc.)
- Dictionaries of Le Kha Ke, Nguyen Trong Bau.
- Lexicons of "Trung tam bien soan dich thuat sach sai gon".
- Coocurence: Use of the TRAMEUR of Paris 3.

↙

**MT VN>FR dictionaries**  
**Tables N-2 N-1 N0 N+1 N+3**  
**Grammars (NOOJ)**





# International Conference NOOJ 2021

## University of Franche-Comté



### IV) CASE-STUDY : NOOJ IN TECHNICAL- OPERATIONAL (FUNCTIONAL) CONTEXT

#### Dictionnaire marqueurs Présent (6 entrées traduite en français)

```
C:\Users\internet\Documents\Nooj\vi\Lexical Analysis\CTO.dic
Dictionary contains 6 entries

# NooJ V2
# Dictionary
#
# Input Language is: vi
#
# Alphabetical order is not required.
#
# Use inflectional & derivational paradigms' and properties' definition files (
# Special Command: #use properties.def
# Special Command: #use paradigms.nof
#
# Special Features: +NW (non-word) +FXC (frozen expression component) +UNAMB (u
#                   +FLX= (inflectional paradigm) +DRV= (derivational paradigm)
#
# Special Characters: '\ ' "' '+' ',' '#' ' ' '
#
hôm nay,CT+o+FR=aujourd'hui
bây giờ,CT+o+FR=maintenant
lúc nay,CT+o+FR=à présent
ngày nay,CT+o+FR=de nos jours
hiện nay,CT+o+FR=de nos jours
trong hoàn cảnh hiện nay,CT+o+FR=par le temps qui court
```



# International Conference NOOJ 2021

## University of Franche-Comté



### IV) CASE-STUDY : NOOJ IN TECHNICAL- OPERATIONAL (FUNCTIONAL) CONTEXT

#### Dictionnaire marqueurs Neutre (188 entrées traduite en français)

```
C:\Users\internet\Documents\Nooj\vi\Lexical Analysis\CTN.dic
Dictionary contains 188 entries

#                               +FLX= (inflectional paradigm) +DRV= (derivational paradigm)^
#
# Special Characters: '\ ' "' '+' ',' '# ' ' '
#
trước hôm,CT+n+FR=la veille (de)
một thời đại,CT+n+FR=une ère
kỷ nguyên,CT+n+FR=une ère
mười ngày,CT+n+FR=une décade
ba tháng,CT+n+FR=un trimestre
quý,CT+n+FR=un trimestre
một thời kỳ,CT+n+FR=une époque
thời đại,CT+n+FR=une époque
một thế kỷ,CT+n+FR=un siècle
một thời gian,CT+n+FR=une durée/une période
nửa năm,CT+n+FR=un semestre
sáu tháng,CT+n+FR=un semestre
thiên kỷ,CT+n+FR=le millénaire
bình minh,CT+n+FR=l'aube
rạng đông,CT+n+FR=l'aube
hoàng hôn,CT+n+FR=le crépuscule
thời gian,CT+n+FR=le temps
khoảng thời gian,CT+n+FR=un (certain) temps
lúc,CT+n+FR=un moment/un instant
giây,CT+n+FR=la seconde
phút,CT+n+FR=la minute
giờ,CT+n+FR=l'heure
ngày,CT+n+FR=le jour/la journée
```



# International Conference NOOJ 2021

## University of Franche-Comté



### IV) CASE-STUDY : NOOJ IN TECHNICAL- OPERATIONAL (FUNCTIONAL) CONTEXT

#### Dictionnaire marqueurs Futur (14 entrées traduite en français)

```
C:\Users\internet\Documents\Nooj\vi\Lexical Analysis\CTF.dic
Dictionary contains 14 entries

#
# Input Language is: vi
#
# Alphabetical order is not required.
#
# Use inflectional & derivational paradigms' and properties' definition files
# Special Command: #use properties.def
# Special Command: #use paradigms.nof
#
# Special Features: +NW (non-word) +FXC (frozen expression component) +UNAMB
#                   +FLX= (inflectional paradigm) +DRV= (derivational paradigm)
#
# Special Characters: '\ ' "' '+' ', ' #' ' '
#

ngày mai,CT+f+FR=demain
ngày kia,CT+f+FR=après-demain
ngày mốt,CT+f+FR=après-demain
sáng mai,CT+f+FR=demain matin
tối mai,CT+f+FR=demain soir
đêm mai,CT+f+FR=demain soir
sáng ngày kia,CT+f+FR=après demain matin
tối ngày kia,CT+f+FR=après demain soir
sáng hôm sau,CT+f+FR=le lendemain matin
tối hôm sau,CT+f+FR=le lendemain soir
tuần tới,CT+f+FR=la semaine prochaine
tháng tới,CT+f+FR=le mois prochain
```



# International Conference NOOJ 2021

## University of Franche-Comté



### IV) CASE-STUDY : NOOJ IN TECHNICAL- OPERATIONAL (FUNCTIONAL) CONTEXT

#### Dictionnaire marqueurs Passé (14 entrées traduites en français)

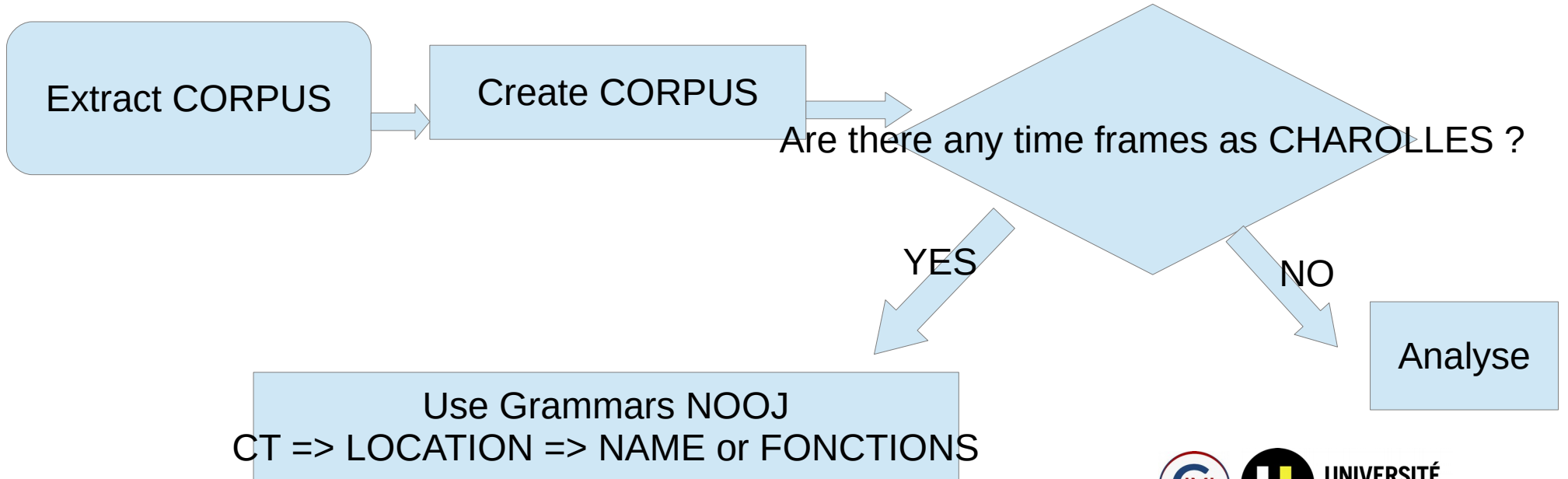
```
C:\Users\internet\Documents\NooJ\vi\Lexical Analysis\CTP.dic
Dictionary contains 12 entries

# NooJ V2
# Dictionary
#
# Input Language is: vi
#
# Alphabetical order is not required.
#
# Use inflectional & derivational paradigms' and properties' definition files (
# Special Command: #use properties.def
# Special Command: #use paradigms.nof
#
# Special Features: +NW (non-word) +FXC (frozen expression component) +UNAMB (u
#                   +FLX= (inflectional paradigm) +DRV= (derivational paradigm)
#
# Special Characters: '\' ''' '+' ',' '#' ' ' '
#
hôm qua,CT+p+FR=hier
hai hôm trước,CT+p+FR=l'avant-veille (de)
đêm qua,CT+p+FR=hier soir
tối qua,CT+p+FR=hier soir
đêm trước,CT+p+FR=la nuit dernière
tối hôm trước,CT+p+FR=la nuit dernière
tối hôm trước,CT+p+FR=la veille au soir
sáng hôm trước,CT+p+FR= la veille au matin
tuần trước,CT+p+FR=la semaine passé
tháng trước,CT+p+FR=le mois passé
năm ngoái,CT+p+FR=l'an passé
năm ngoái cũng vào thời điểm này,CT+p+FR=l'an dernier à pareille
```





### FLOWCHART & PREPARATION OF PROCESSING OF SPATIO-TEMPORAL INFORMATION



# International Conference NOOJ 2021

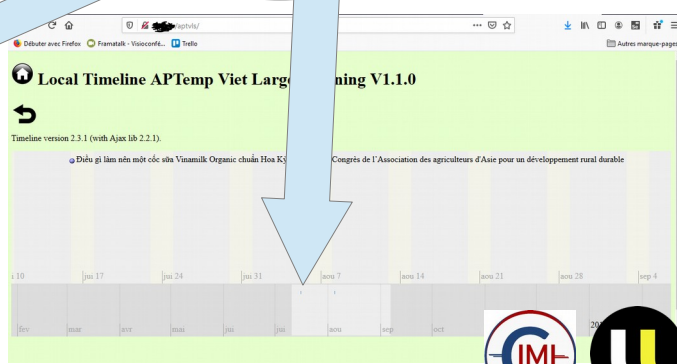
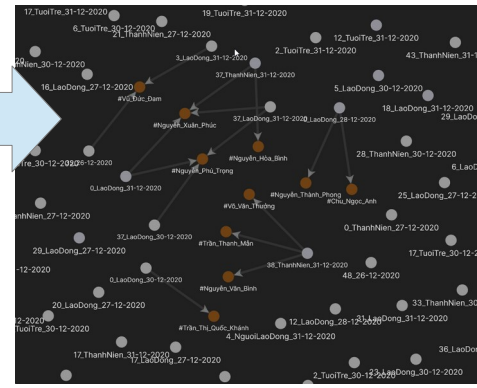
## University of Franche-Comté



### PROCESSING

Constitution du corpus web  
Scripts Python

NOOJ  
Traitement du corpus  
Grammaires  
Dictionnaires



UNIVERSITÉ DE LORRAINE



## VIEW SPATIO-TEMPORAL DEMO





**DATAVIZ : KMS (Knowledge Management System) complementary approach**

**by**

**Philippe LAMBERT**

**&**

**VIEW KMS DEMO**

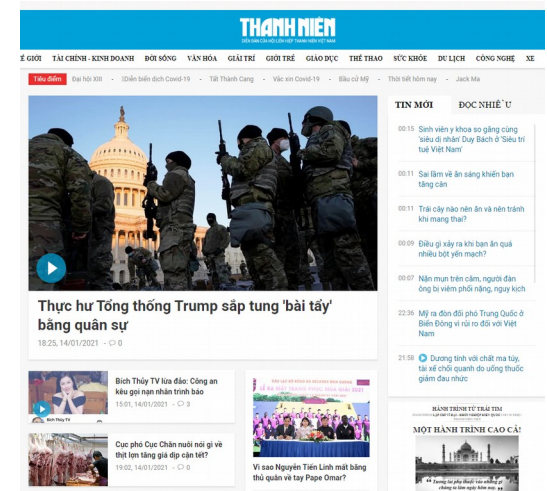




# DATAVIZ example (Philippe LAMBERT)

## Process

- Extraction of articles from VN press sites:
- Transformation into an XML file
- Processing of XML nodes by NooJ (<content>)
- Extraction of concordance terms
- Injection of concordance terms as #tags into an information visualization solution.





# Conceptual Framework and objectives

## Personal Knowledge Management System

1. Rapid implementation of an information system for the analyst
2. Analyst / Information system interactivity for knowledge capitalization
3. Information need Analysis for monitoring information
4. Portability

## Data Vizualisation

1. SNA – Social Network Analysis
2. Analyst / Information system interactivity for knowledge capitalization



md

- CORPUS
- COVID

**Corpus**

**NooJ Extraction  
Implemented in the system**



|                        |    |
|------------------------|----|
| Covid-19               | 64 |
| Ban_chi_dạo            | 15 |
| Bộ_Y_tế                | 14 |
| Số_Y_tế                | 14 |
| Ban_Quản_lý            | 7  |
| Bộ_Công_an             | 7  |
| Bộ_Kế_hoạch_và_Đầu_tư  | 3  |
| Lê_Văn_Thành           | 3  |
| Báo_Lao_Động           | 2  |
| Báo_Thanh_Niên         | 2  |
| Covid                  | 2  |
| Nguyễn_Thanh_Long      | 2  |
| Nguyễn_Văn_Nên         | 2  |
| Số_Giáo_dục_và_Đào_tạo | 2  |
| Võ_Văn_Thương          | 2  |
| Bà_Nguyễn_Thị_Lan      | 1  |
| bà_Tô_Thị_Bích_Châu    | 1  |
| Ban_Tuyên_giáo         | 1  |
| Báo_Người_Lao_Động     | 1  |
| Bộ_Thương_mại          | 1  |
| Bộ_Xây_dựng            | 1  |
| Covid_-19              | 1  |
| Đà_Nẵng                | 1  |
| Huỳnh_Văn_Tấn          | 1  |
| Lê_Quân                | 1  |
| Ngô_Minh_Châu          | 1  |
| Nguyễn_Ngọc_Thiện      | 1  |
| Nguyễn_Thị_Ánh_Hồng    | 1  |
| ông_Lê_Quân            | 1  |
| Phạm_Minh_Chính        | 1  |
| PNJ                    | 1  |
| Quảng_Ngãi             | 1  |
| Số_Kế_hoạch_và_Đầu_tư  | 1  |
| Số_Tư_pháp             | 1  |
| Số_Xây_dựng            | 1  |
| Thủ_tướng_Chính_phủ    | 1  |
| TP_Quảng_Ngãi          | 1  |

**No file is open**

- Create new file (Ctrl + N)
- Go to file (Ctrl + O)
- See recent files (Ctrl + O)

**Each keyword refers to a documentary source  
for information mining**

Documentary source

---

tag : [bà\_Tô\_Thị\_Bích\_Châu, Covid-19, Covid, Huỳnh\_Văn\_Tấn, PNJ]  
url : [https://thanhnien.vn/.../su/pnj-ung-ho-2-ti-dong-kinh-phi-mua-vac-xin-phong-chong-dich-covid-19-1396389.html]

---

PNJ ủng hộ 2 tỉ đồng kinh phí mua vắc xin phòng chống dịch Covid-19  
Đây là hoạt động thể hiện sự quan tâm cùng phát triển cùng nhau, chung tay cùng toàn dân chiến thắng đại dịch.

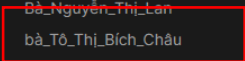
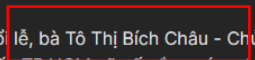
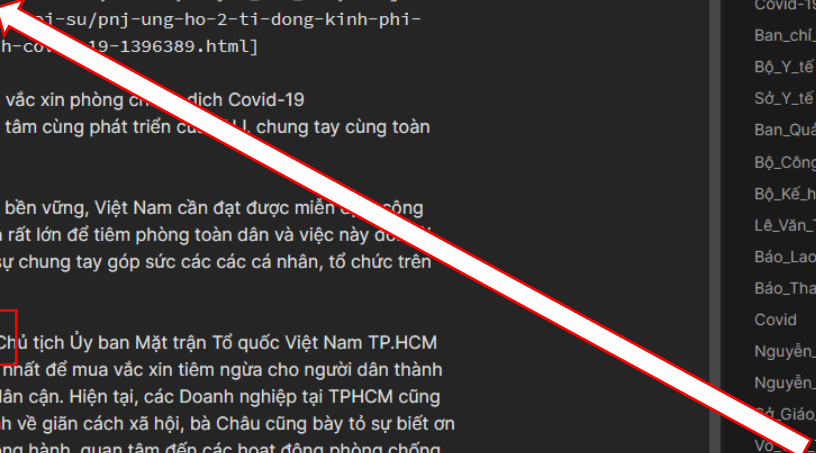
Để ngăn chặn dịch bệnh một cách bền vững, Việt Nam cần đạt được miễn dịch cộng đồng, do đó cần một lượng vắc xin rất lớn để tiêm phòng toàn dân và việc này đòi hỏi nguồn kinh phí lớn, do đó rất cần sự chung tay góp sức các cá nhân, tổ chức trên toàn quốc.

Tại buổi lễ, bà Tô Thị Bích Châu - Chủ tịch Ủy ban Mặt trận Tổ quốc Việt Nam TP.HCM cho biết, TP.HCM sẽ cố gắng sớm nhất để mua vắc xin tiêm ngừa cho người dân thành phố cũng như chia sẻ với các tỉnh lân cận. Hiện tại, các Doanh nghiệp tại TPHCM cũng nghiêm túc chấp hành các quy định về giãn cách xã hội, bà Châu cũng bày tỏ sự biết ơn đối với Doanh nghiệp vì đã luôn đồng hành, quan tâm đến các hoạt động phòng chống dịch của Mặt trận Tổ quốc Việt Nam TP.HCM.

Ông Huỳnh Văn Tấn, GD Truyền Thông & Đối Ngoại đại diện PNJ trao tặng 2 tỉ đồng kinh phí mua vắc xin phòng chống Covid -19 cho Ủy ban Mặt Trận Tổ Quốc Việt Nam TP.HCM "Số tiền 2 tỉ đồng được trích từ nguồn quỹ các hoạt động xã hội cộng đồng của PNJ, trong đó một phần lớn nguồn quỹ này là từ nguồn đóng góp của mỗi nhân viên hàng quý, mong muốn vắc xin sẽ nhanh chóng được tiêm ngừa cho toàn dân, trước tình hình dịch bệnh bùng phát trở lại càng phức tạp hiện nay, thì việc tiêm vắc xin cho người dân để đảm bảo miễn dịch cộng đồng là vô cùng cấp thiết", Ông Huỳnh Văn Tấn - Giám đốc Truyền Thông & Đối Ngoại Công ty PNJ chia sẻ. "Số tiền 2 tỉ đồng được trích từ nguồn quỹ các hoạt động xã hội cộng đồng của PNJ, trong đó một phần lớn nguồn quỹ này là từ nguồn đóng góp của mỗi nhân viên hàng quý, mong muốn vắc xin sẽ nhanh chóng được tiêm ngừa cho toàn dân, trước tình hình dịch bệnh bùng phát trở lại càng phức tạp hiện nay, thì việc tiêm vắc xin cho người dân để đảm bảo miễn dịch cộng đồng là vô cùng cấp thiết", Ông Huỳnh Văn Tấn - Giám đốc Truyền Thông & Đối Ngoại Công ty PNJ chia sẻ.

Siêu thị 0 đồng của PNJ hỗ trợ công nhân vùng tâm dịch Bắc Giang PNJ là một trong những doanh nghiệp tiên phong trong các hoạt động phòng chống Covid ngay từ khi dịch bệnh mới xuất hiện tại Việt Nam. Luôn tuân thủ các biện pháp phòng chống dịch Covid nhằm bảo vệ sức khỏe an toàn cho nhân viên, khách hàng và cộng đồng. Vào cuối tuần qua, các cán bộ nhân viên PNJ tại Bắc Giang và Bắc Ninh cũng đã có hoạt

- Covid-19 64
- Ban\_chỉ\_đạo 15
- Bộ\_Y\_tế 14
- Số\_Y\_tế 14
- Ban\_Quản\_lý 7
- Bộ\_Công\_an 7
- Bộ\_Kế\_hoạch\_và\_Đầu\_tư 3
- Lê\_Văn\_Thành 3
- Báo\_Lao\_Động 2
- Báo\_Thanh\_Niên 2
- Covid 2
- Nguyễn\_Thanh\_Long 2
- Nguyễn\_Văn\_Nên 2
- Bộ\_Giáo\_dục\_và\_Đào\_tạo 2
- Voi\_Thương 2
- Bà\_Nguyễn\_Thị\_Lan 1
- bà\_Tô\_Thị\_Bích\_Châu 1
- Ban\_Tuyên\_giáo 1
- Báo\_Người\_Lao\_Động 1
- Bộ\_Thương\_mại 1
- Bộ\_Xây\_dựng 1
- Covid\_-19 1
- Đà\_Nẵng 1
- Huỳnh\_Văn\_Tấn 1
- Lê\_Quân 1
- Ngô\_Minh\_Châu 1
- Nguyễn\_Ngọc\_Thiện 1
- Nguyễn\_Thị\_Ánh\_Hồng 1
- ông\_Lê\_Quân 1
- Phạm\_Minh\_Chính 1
- PNJ 1
- Quảng\_Ngãi 1
- Số\_Kế\_hoạch\_và\_Đầu\_tư 1
- Số\_Tư\_pháp 1





### V) Conclusion

- NOOJ compatible with DATAVIZ (prospects/tracks : CSV export, API, PPlugin).
- NOOJ requires a certain expertise, but you can extract named entities without being an expert. (It is recommended to do the NOOJ week).
- We can deal with poorly endowed languages vs AI and statistical model.
- NOOJ can have a technical-operational use.
- **Difficult to evaluate without annotated corpus.**



### VI) Perspectives

- **Develop the technical-operational approach in NLP with NOOJ.**
- **Make NOOJ modular**
- **Develop the interoperability of NOOJ**
- **Present NOOJ to the whole CIMI**
- **Annotation of corpus (spatio-temporal & thematic)**
- **...and many other great perspectives.**



# International Conference NOOJ 2021

## University of Franche-Comté



# THANK YOU !



UNIVERSITÉ  
DE LORRAINE