



**HAL**  
open science

# Analyse automatique de textes en vietnamien pour l'intelligence économique et l'information scientifique et technique

Philippe Lambert

► **To cite this version:**

Philippe Lambert. Analyse automatique de textes en vietnamien pour l'intelligence économique et l'information scientifique et technique. École thématique. Semaine NooJ à l'INALCO, INALCO, 2011, pp.22. cel-00677799

**HAL Id: cel-00677799**

**<https://cel.hal.science/cel-00677799v1>**

Submitted on 9 Mar 2012

**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.



Analyse automatique de textes en vietnamien  
pour l'intelligence économique et  
l'information scientifique et technique

Philippe Lambert

19 janvier 2011

# Plan

## 1/ Contexte : l'Intelligence économique

Définitions

Cycle

Couches sémantiques (merci NooJ)

Design de l'information

## 2/ Le module de vietnamien pour NooJ

2 études de cas

## 3/ A vous de jouer (TP)

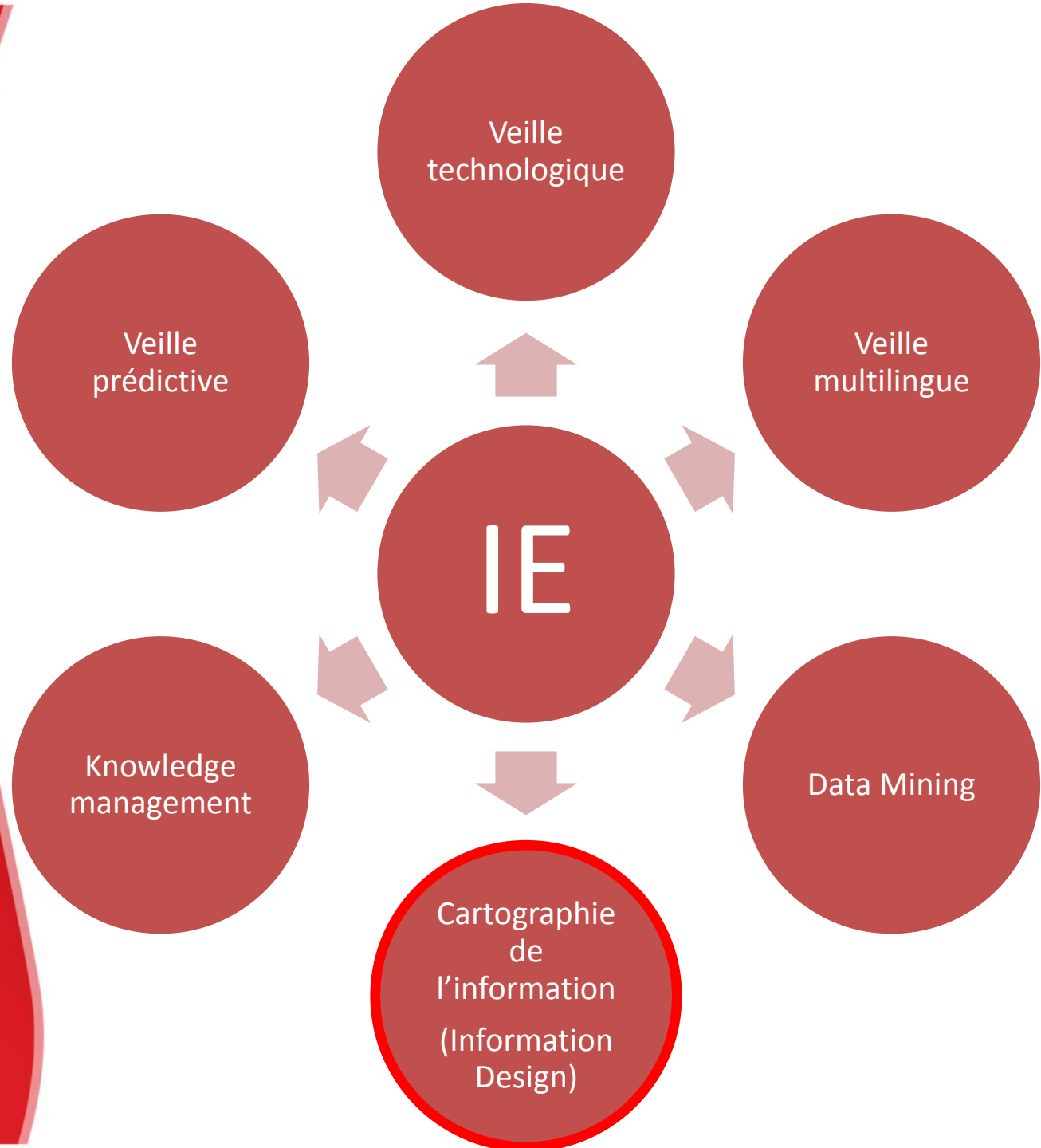
# Définition de l'intelligence économique (IE):

1/ L'ensemble des activités coordonnées de collecte, de traitement et de diffusion de l'information utile aux acteurs économiques en vue de son exploitation

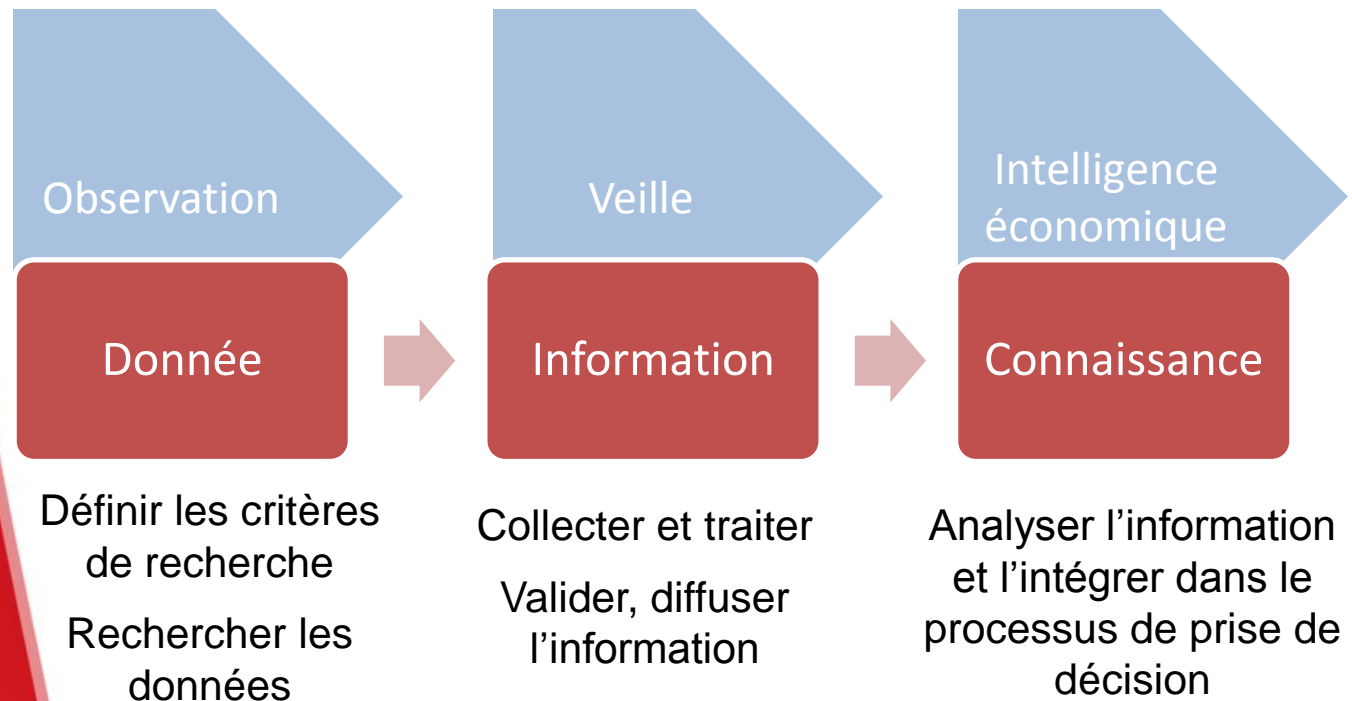
*(Rapport Martre 1994)*

2/ Réduction du niveau de l'incertitude dans la prise de décision

*(Nombreux auteurs)*



# Le cycle de l'IE



# Application de couches sémantiques dans le processus IE (merci NooJ) (1)

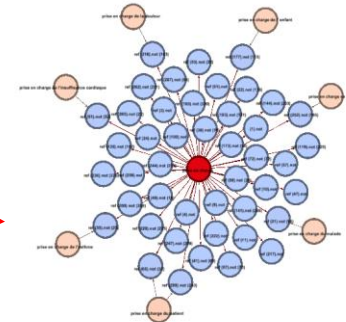
Base de données ISTIE



Mappage (alignement)  
des différents descripteurs  
(homogénéisation)

**NOOJ**

Résultats



**DESIGN DE L'INFORMATION**

# DESIGN DE L'INFORMATION

## « Utilisateur proactif »

« l'art et la science de la préparation de l'information afin qu'elle puisse être utilisée par les êtres humains avec efficacité »

*Benjamin Jotham Fry. (2004). Computational information design. bfa Communication Design, in Computer Science Carnegie Mellon University. Massachusetts Institute of Technology, April 2004*

## Signaux faibles

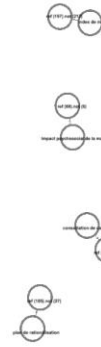
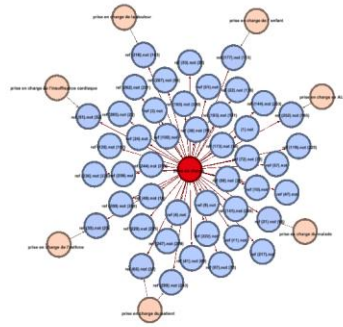
« une alerte (externe ou interne), des événements et développements qui sont encore trop incomplets pour permettre une estimation fine de leur impact et/ou pour déterminer une réponse pleinement adaptée »

*H. I. Ansoff, "Strategic response in turbulent environments," Handbook of business strategy, 1985.*

« **Soudaineté et étonnement** » (« effet chocolat »)



# DESIGN DE L'INFORMATION



## Objectifs :

- Signaux faibles
- Relations thématiques latentes
- Diagnostic



# Démonstration pour les maladies chroniques



## 2. Le module de vietnamien pour NooJ

Principal problème : Une douzaine d'encodage  
(unicode, VPS, VNI, ABC, ...)

**(2) Enter a Text:**

ê ê

Value (decimal):

**101 204 163 32 225 186 185**

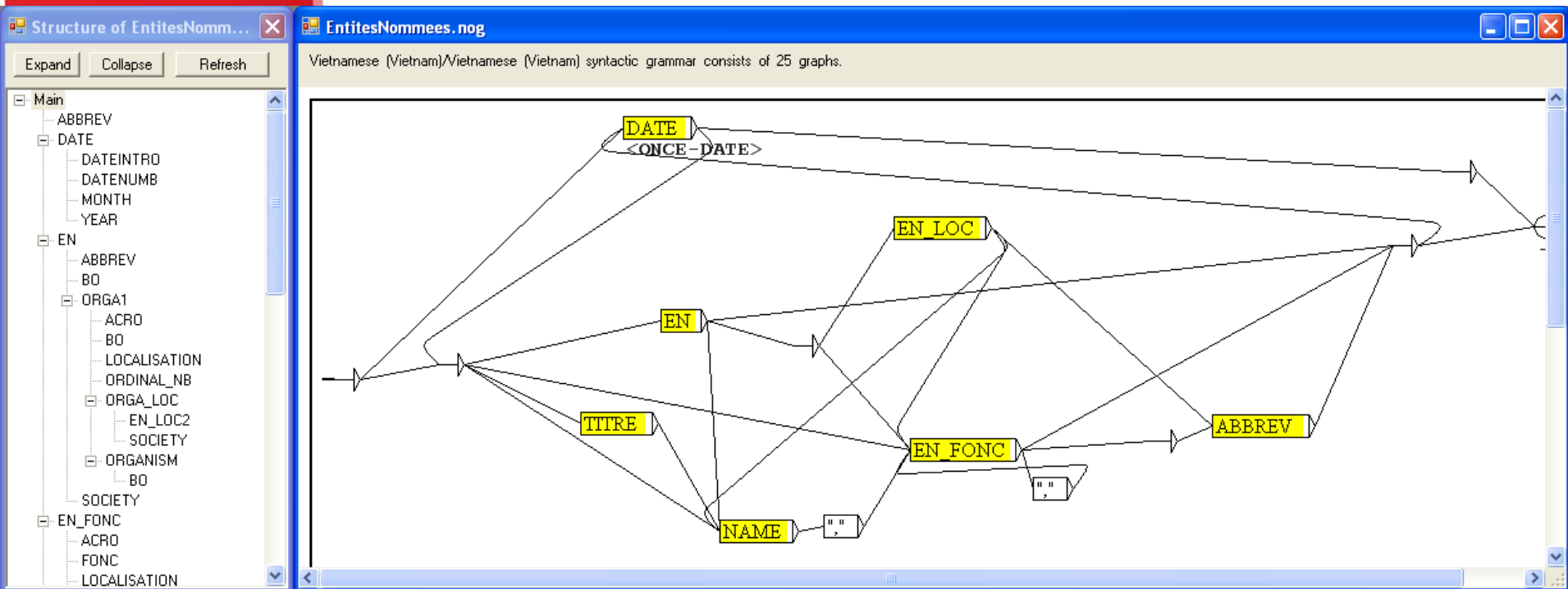
- Une trentaine de graphes
- Une vingtaine de dictionnaires thématiques (économie, Noms propres, spécifiques, etc.)

## 2 études de cas :

- Traçage des personnalités de l'actualité VN
- Identification d'experts VN

# Traçage des personnalités de l'actualité VN

Corpus : 1000 textes de l'actualité vietnamienne (source AVI)  
→ Application de la grammaire syntaxique de reconnaissance des entités nommées (noms de personnes, de lieux, d'organisations, les dates ou les unités monétaires)





Réseau sémantique de l'actualité vietnamienne

## **Cas 2 : Identification d'experts**

**Objectif principal : identifier les spécialistes d'un domaine**



# Thư viện KHOA HỌC TỔNG HỢP THÀNH PHỐ HỒ CHÍ MINH



bạn đang ở: trang chủ

### Các trang chính

- Tìm kiếm tài liệu
- Các Dịch vụ
- Hướng dẫn & hỗ trợ
- Hoạt động mạng lưới
- Hoạt động khiếm thị
- Nghề thư viện
- Chính sách
- Liên kết web sites

### Liên kết

Thư viện trong nước

### Làm thế nào...?

mượn sách thư viện



Số người truy cập :  
**2882635**

### Tin tức

- » THÔNG BÁO THƯ VIỆN TẠM NGỪNG PHỤC VỤ - ĐẠI HỘI CÁN BỘ, VIÊN CHỨC THƯ VIỆN 2011
- » THÔNG BÁO TẠM NGỪNG PHỤC VỤ CÁC NGÀY LỄ TẾT NGUYÊN ĐÁN TÂN MÃO
- » VIETTEL ĐỒNG HÀNH CÙNG THƯ VIỆN
- » LỚP TẬP HUẤN NGHIỆP VỤ - TIÊU ĐỀ CHỦ ĐỀ THÁNG 11-2010
- » Cuộc thi "PHẦN MỀM ỨNG DỤNG VÌ PHÁT TRIỂN"
- » TRAO TẶNG TÀI LIỆU CHO TRẺ EM KHIẾM THỊ
- » THƯ VIỆN LƯU ĐỘNG PHỤC VỤ CHO NGƯỜI KHIẾM THỊ
- » THƯ VIỆN KHOA HỌC TỔNG HỢP TP. HCM GIỚI THIỆU BỘ SÁCH TIÊU ĐỀ CHỦ ĐỀ
- » THÔNG TIN VỀ CÁC HOẠT ĐỘNG CỦA THƯ VIỆN KHOA HỌC TỔNG HỢP TẠI HUYỆN CẦN GIỜ NHÂN LỄ HỘI NGHINH ỒNG VÀ TẾT TRUNG THU NĂM 2010
- » HỘI THẢO "THỰC TRẠNG VÀ GIẢI PHÁP PHÁT TRIỂN VĂN HÓA ĐỌC Ở VIỆT NAM"

>>> >>

### Giới thiệu thư viện

#### Lịch sử hình thành

Thành lập năm 1868: thư viện của các đô đốc thống đốc. Trước 1975: Thư viện Quốc gia của Nam Việt Nam. Hiện nay: Thư viện Khoa học Tổng hợp Thành phố Hồ Chí Minh (Quyết định của Ủy ban Nhân dân TpHCM ngày 14.04/1978)

#### Cơ cấu tổ chức

Hệ thống, sơ đồ tổ chức, cơ cấu các phòng ban

#### Các dự án

Các Dự án phục vụ cộng đồng thư viện đã thực hiện, xây dựng và phát triển,...

#### Địa chỉ và bản đồ

Cung cấp các thông tin giúp bạn đọc định vị đường đến thư viện

#### Tham quan thư viện

Mời bạn tham quan thư viện trực tuyến qua các hình ảnh, lời giới thiệu,...

#### Ban đọc lưu ý

#### Vai trò & nhiệm vụ

Tên gọi chính thức của thư viện và tên giao dịch quốc tế, vai trò & chức năng, nhiệm vụ, cơ quan chủ quản, các hoạt động chính và hoạt động hỗ trợ...

#### Đối ngoại

Liên kết, hợp tác, trao đổi với các thư viện, tổ chức, ... trong và ngoài nước

#### Thủ tục cấp thẻ

Thời gian làm thẻ, Đối tượng, Các loại thẻ và chi phí, Thủ tục cần thiết, Thời hạn sử dụng thẻ, Ghi chú

#### Giờ hoạt động

Giờ giấc, ngày nghỉ, thông báo ngưng phục vụ,...

#### Các tiện ích

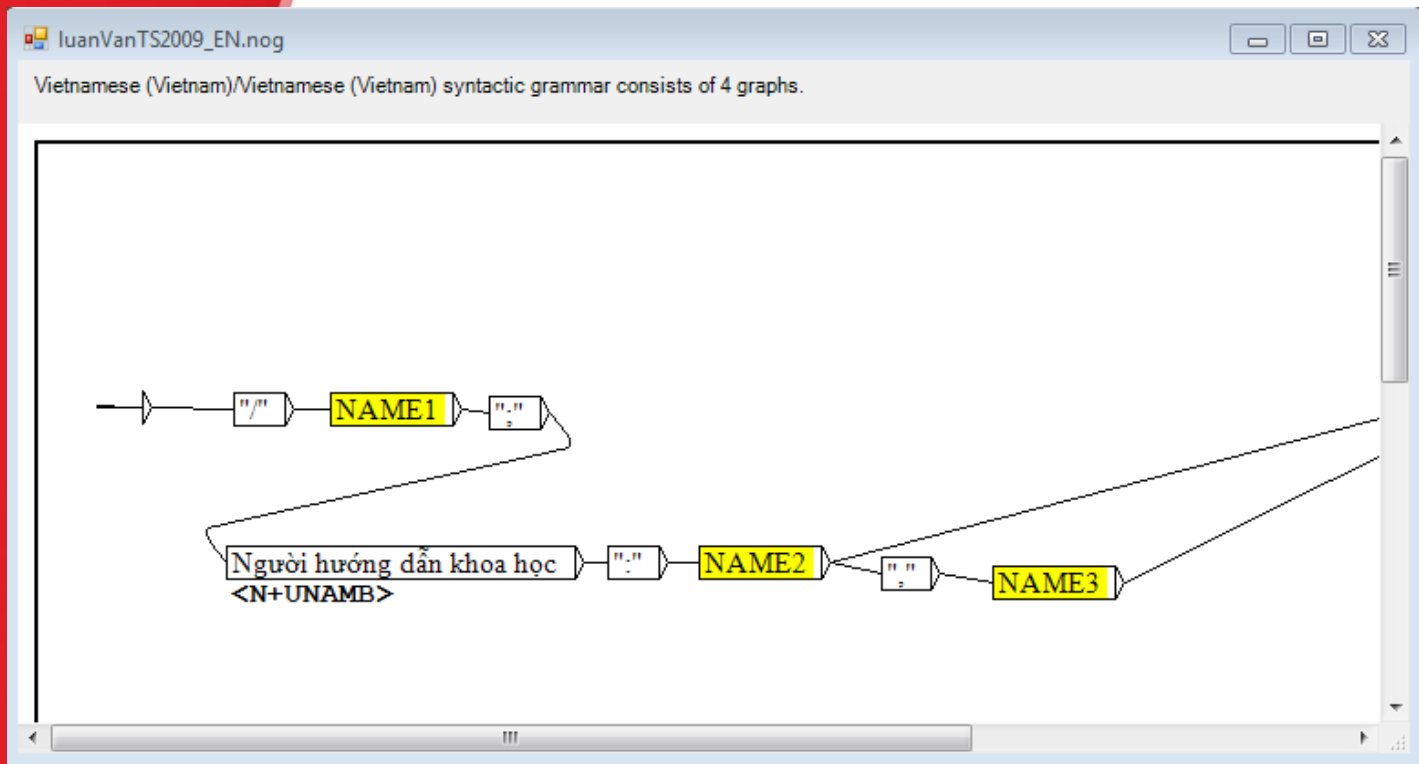
Phương tiện, trang thiết bị giúp bạn đọc khai thác và sử dụng thư viện

#### Hỗ trợ liên lạc



DỊCH VỤ THƯ VIỆN  
CHO NGƯỜI KHIẾM THỊ  
CẢM ƠN THỰC HÀNH TỐT NHẤT



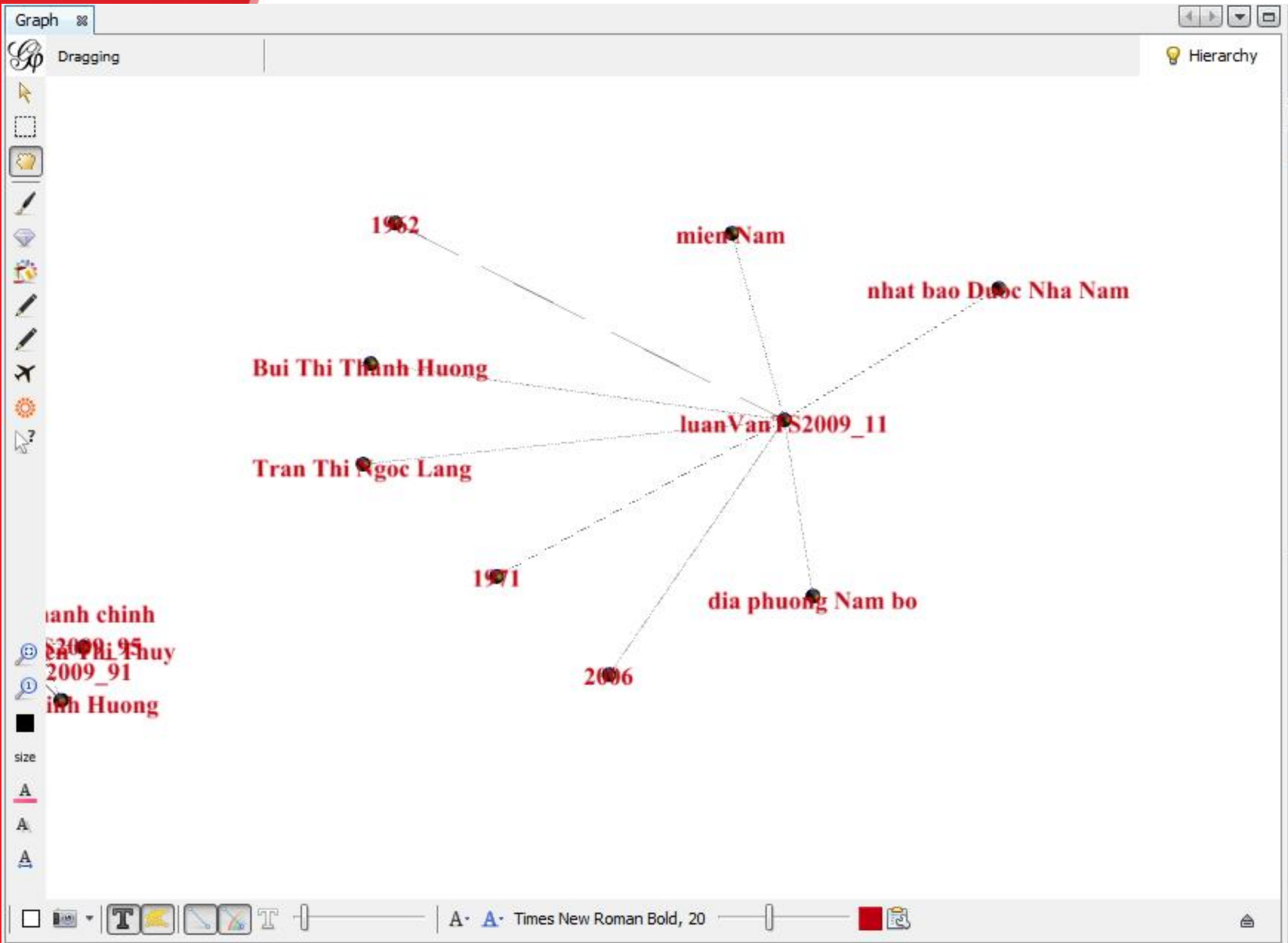


Source	syntagm	TF
luanVanTS2009_Titre_01	vong mac dai thao duong	0,419434073700293
luanVanTS2009_Titre_01	ung dung quang dong toan vong mac	1.20E+14
luanVanTS2009_Titre_02	Pham tru noi dong	1.84E+14
luanVanTS2009_Titre_03	thanh pho Ho Chi Minh	0,271929963352074
luanVanTS2009_Titre_03	Quan he bien chung	0,30403738628168
luanVanTS2009_Titre_05	Bieu hien tu tuong ton giao tin nguong	0,883156217294405
luanVanTS2009_Titre_05	kien truc ton giao Nam bo	0,566125780316926
luanVanTS2009_Titre_06	van ban tin tieng Anh	0,697466961350453
luanVanTS2009_Titre_06	tieng Viet	0,270134593180383
luanVanTS2009_Titre_06	van ban tin tieng Anh	1.33E+14
luanVanTS2009_Titre_07	nguồn viên trợ phi chính phủ nước ngoài	1.73E+14
luanVanTS2009_Titre_07	thanh pho Ho Chi Minh	0,356725328268312
luanVanTS2009_Titre_08	Phuong phap dich Anh-Viet	1.84E+14
luanVanTS2009_Titre_10	dong bang ven bien Nam Trung	0,50644268957843
luanVanTS2009_Titre_101	Dan chu Nhan dan Lao	0,307566841843823
luanVanTS2009_Titre_101	tinh mien Trung nuoc	0,307566841843823
luanVanTS2009_Titre_101	tinh mien Trung nuoc	0,322945183936014
luanVanTS2009_Titre_102	hanh vi trai phap luat	0,423694883063359
luanVanTS2009_Titre_103	huong hoan thien	0,428196953839712
luanVanTS2009_Titre_105	Tieng Viet	0,4083428989691277
luanVanTS2009_Titre_105	To hop cu phap dang lap	0,939188876289937
luanVanTS2009_Titre_106	tieng Viet	0,14160281094133
luanVanTS2009_Titre_106	cau don tieng Anh	0,565030053000184
luanVanTS2009_Titre_106	cau don tieng Anh	0,299133557470686

**Restructuration des données**



**Entrepôt de données  
(analyse multidimensionnelle)**



## Logiciels de type Open Source pour la visualisation des données :

1/ Gephi (<http://gephi.org>)

2/ Pajek (<http://vlado.fmf.uni-lj.si/pub/networks/pajek/>)

3/ Netdraw (<http://www.analytictech.com/netdraw/netdraw.htm>)

## Formats de fichier pour la visualisation :

1/ CSV

2/ .net (Pajek)

3/ .txt

4/ Graphml

## Processus de traitement de l'info pour la visualisation

1/ l'acquisition (*acquire*) des données depuis un support quelconque,

2/ le découpage (*parse*) → structuration aux données,

3/ le filtrage (*filter*) → les données pertinentes,

4/ la fouille (*mine*) → données dans un contexte mathématique,

5/ la représentation (*represent*) → une représentation simple que les données peuvent revêtir,

6/ le raffinement (*refine*) → rendus visuels plus évolués

7/ l'interaction (*interact*) : l'ajout de méthodes de manipulation de données via la visualisation

# NooJ + Les techniques de visualisation / cartographie de l'information

Quelques pistes de recherche :

- **Modélisations cognitives** des décideurs (entreprise)  
questionnaires / interviews
- **Géographie littéraire** -> cartographie romanesque
- **Traductologie – Analyse** :  
cartographie des thématiques traduites

# Remerciements



**Knowledge, Information and Web  
Intelligence**

**Nouveaux modèles de l'IE**

## Éléments de bibliographie :

1/ Benjamin Jotham Fry. Computational information design. bfa Communication Design, in Computer Science Carnegie Mellon University. Massachusetts Institute of Technology, April 2004.

2/ M. J. Eppler and R. A. Burkhard, “Visual representations in knowledge management: framework and cases,” *Journal of Knowledge Management*, vol. 11, no. 4, pp. 112–122, 2007.

3/ S. Sidhom, “Approche conceptuelle par un processus d'annotation pour la représentation et la valorisation de contenus informationnels en intelligence économique (IE),” 2008. [Online]. Available: <http://hal.archives->

### 3. A vous de jouer (TP)

1. Ouvrir le projet « Lambert »
2. Demander à NooJ une analyse textuelle sur le nœud « Titre » uniquement
3. Appliquer la grammaire SN2
4. Transformer les résultats en une base de données type Excel