

## Content

Our work is mainly concentrated around two kinds of documents:

- Web documents, parsing and studying both multimedia content and markup to find information relevant to our goals: subject of discussion, prices...
- Ancient documents, our main field of study being intertextuality (multi-structure, viewpoint confrontation and quotations).

Passage citant : 4.4.1 8 - 4.6.2 11

ἡμᾶς ἐν τοῖς εὐαγγελίοις ἄχρι νῦν διετάρασεν ἀβασανίστως καὶ  
δημαρτημένως ὑπὸ νηπιότητος ἀκρωμένους ἐκπορευομένου αὐτοῦ εἰς  
ὁδὸν προσελθῶν τις ἐγνονυπέτει λέγων διδάσκαλε ἀγαθὲ τί  
ποιήσω ἵνα ζωὴν αἰώνιον κληρονομήσω ὃ δὲ Ἰησοῦς  
λέγει τί με ἀγαθὸν λέγεις οὐδεὶς ἀγαθὸς εἰ μὴ εἶς ὁ  
θεὸς τὰς ἐντολάς οἶδας μὴ μοιχεύσης μὴ φονεύσης μὴ  
κλέψης μὴ ψευδομαρτυρήσης τίμα τὸν πατέρα σου  
καὶ τὴν μητέρα ὃ δὲ ἀποκριθεὶς λέγει αὐτῷ πάντα  
ταῦτα ἐφύλαξα ἐκ νεότητός μου ὃ δὲ Ἰησοῦς ἐμβλέψας  
ἠγάπησεν αὐτὸν καὶ εἶπεν ἔν σοι ὑστερεῖ εἰ θέλεις τέλειος εἶναι πώλησον  
ὅσα ἔχεις καὶ διάδος πτωχοῖς καὶ ἕξεις

Passage cité : Lc 18.18-22

ὡμῖν ὅς ἂν μὴ δέξηται τὴν βασιλείαν τοῦ θεοῦ ὡς παιδίον οὐ μὴ εἰσέλθῃ  
εἰς αὐτήν καὶ ἐπηρώτησέν τις αὐτὸν ἄρχων λέγων διδάσκαλε  
ἀγαθὲ τί ποιήσας ζωὴν αἰώνιον κληρονομήσω εἶπεν  
δὲ αὐτῷ ὁ Ἰησοῦς τί με λέγεις ἀγαθὸν οὐδεὶς ἀγαθὸς εἰ  
μὴ εἶς ὁ θεὸς τὰς ἐντολάς οἶδας μὴ μοιχεύσης μὴ  
φονεύσης μὴ κλέψης μὴ ψευδομαρτυρήσης τίμα τὸν  
πατέρα σου καὶ τὴν μητέρα ὃ δὲ εἶπεν ταῦτα πάντα  
ἐφύλαξα ἐκ νεότητος ἀκούσας δὲ ὁ Ἰησοῦς εἶπεν αὐτῷ  
ἔτι ἔν σοι λείπει πάντα ὅσα ἔχεις πώλησον καὶ διάδος πτωχοῖς καὶ ἕξεις  
θησαυρὸν ἐν τοῖς οὐρανοῖς καὶ δεῦρο ἀκολουθεῖ μοι

Detecting Quotations [Greek Reuse Toolkit, for the  
ANR Project Biblindex]

## Collaborations

The DIRE Platform is a step forward in our partnerships within many projects around the concept of document and its multiple uses, and would not exist without these international, european, national and regional partnerships. Among them:

- ANR Project IPRI (Internet : Pluralisme et Redondance de l'Information)
- ANR Project Biblindex
- COST Action Towntology
- ACI Cognitive / Programme Société de l'Information (CNRS STIC/SHS) : Modélisation opératoire d'un système d'interrogation du document d'architecture en archéologie
- Equipe Projet multi-laboratoires n°50 du RTP 33 (Document) : Du partage de corpus documentaires structurés à la confrontation de points de vue
- Projet Corpus Numériques du Cluster 13 de la région Rhône-Alpes

## Contacts

**Sylvie CALABRETTO**  
sylvie.calabretto@liris.cnrs.fr

**Elöd EGYED-ZSIGMOND**  
elod.egyed-zsigmond@liris.cnrs.fr

**Samuel GESCHE**  
samuel.gesche@liris.cnrs.fr

**DIRE** retrieval  
Information  
Document and

Document and  
Information  
REtrieval :  
A modular platform  
for Document  
Engineering



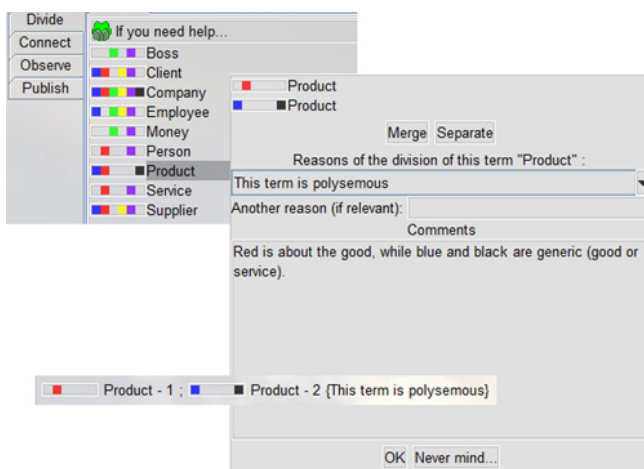
**LIRIS**  
DRIM

## General Description

DIRE's purpose is to gather, unify and disseminate the software and algorithms built by the Document Engineering and Information Retrieval subteam of the LIRIS DRIM team. This platform allows new research, both internally and externally, to quickly iterate from older works.

DIRE provides a stable, modular environment designed around the widely used GATE Text Engineering platform in order to share processing resources with other communities. It also gathers tools developed in previous projects in an open source, packaged form.

Contributions within the DIRE platform are licensed under either the GNU GPL or the GNU LGPL, v3 or later.



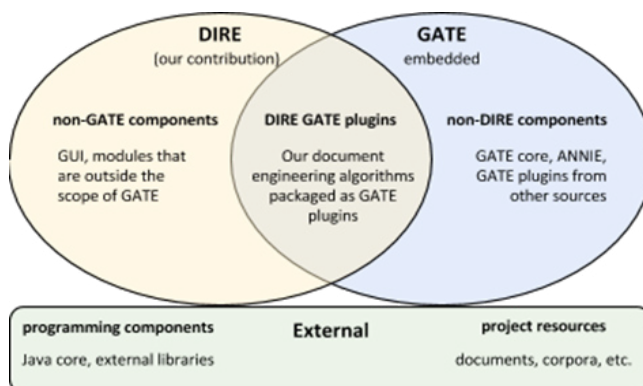
Managing heterogeneity with ArcEnCiel

## The DIRE Development Platform

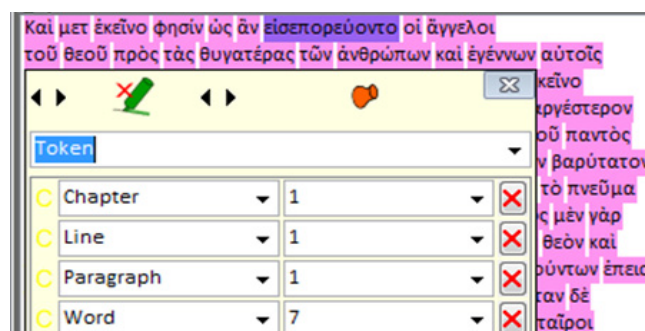
DIRE includes a java platform built around GATE, another platform from the Sheffield University Department of Computer Science.

GATE works as the main Document Processing engine, the abstraction layer allowing a more general project to access document processing functions.

The DIRE Development Platform entails a GATE repository where the GATE abstraction layer can find all our processing modules.



General architecture for DIRE-based software



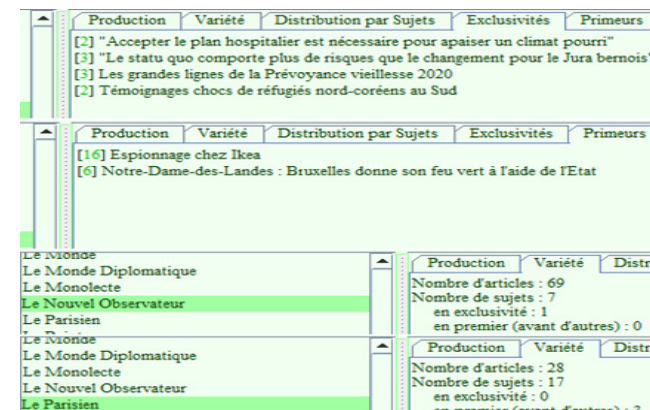
Extracting Text Structure as a GATE Processing Resource  
[Greek Reuse Toolkit, for the ANR Project Biblindex]



Towns are a complex concept  
[ArcEnCiel, for the COST Action Townontology]

## The DIRE Software Repository

The DIRE Software Repository stores full fledged software contributions to various projects. Some predate the DIRE platform and do not make use of the GATE Development Platform, while others are built with a GATE layer. The Repository also contains software libraries that have been built during the projects but do not go well with the GATE document processing paradigm.



Measuring media on the Web  
[IPRINA, for the ANR Project IPRI]

## Website

<http://liris.cnrs.fr/dire>

## GATE plugin repository

<http://liris.cnrs.fr/dire/gate/gate-update-site.xml>