

ANALYSE DE LA BASE DE DONNÉES ISOC-ARTE PAR INGÉNIERIE DE LA CONNAISSANCE : MÉTHODE ET MÉTRIQUES

José PINO-DÍAZ (*), Nuria RODRÍGUEZ-ORTEGA (**), Antonio CRUCES-RODRÍGUEZ (**), Carmen TENOR-POLO (**),
Rafael BAILÓN-MORENO (***), Teresa ABEJÓN-PEÑA (****)

jpinod@uma.es, nro@uma.es, antonio.cruces@uma.es, carmen.tenor@gmail.com, bailonm@ugr.es, teresa.abejon@cchs.csic.es

(*) [Dépt. Économie](#), Université de Málaga, Campus El Ejido, s/n. (Espagne),

(**) [Dépt. Histoire de l'Art](#), Université de Málaga, Campus de Teatinos, s/n. (Espagne),

(***) [Dépt. Ingénierie Chimique](#), Université de Granada, Campus de Fuentenueva (Espagne).

(****) [Centre de Sciences Humaines et Sociales-CSIC](#), C/Albasanz, 26-28. Madrid (Espagne).

Mots clefs :

Veille scientifique et technologique, gestion des connaissances, ingénierie des connaissances, innovation, Evaluation scientifique et technologique, histoire de l'art numérique, humanités numériques, systèmes de connaissances, fouille de textes, analyse des mots associés, réseaux tecno-scientifiques, Théorie de l'Acteur-Réseau.

Keywords:

Scientific and technological watch, knowledge management, knowledge engineering, innovation, KDD, text mining, cword analysis, digital art history, digital humanities, text mining, co-word analysis, Actor-Network Theory, knowledge systems, techno-scientific network.

Palabras clave:

Vigilancia científica y tecnológica, inteligencia competitiva, gestión del conocimiento, ingeniería del conocimiento, sistemas de conocimiento, historia del arte digital, minería de textos, análisis de palabras asociadas, Teoría Actor-Red, redes tecno-científicas.

Résumé

Il a été étudié par des techniques et des outils d'ingénierie des connaissances la base de données ISOC-ARTE afin d'étudier leurs réseaux de connaissances, les sujets de recherche les plus pertinents, les auteurs et les revues qui leur sont associés, et leur évolution dans le temps. On présente la méthodologie utilisée dans l'analyse et les premiers résultats. ISOC-ART réunit plus de 45 000 références, principalement des articles de recherche en histoire de l'art, produites entre 1975 et 2014. L'étude est le résultat de l'accord de coopération signé entre le Centre de Sciences Humaines et Sociales du Conseil Supérieur de la Recherche Scientifique (CCHS-CSIC), propriétaire des bases ISOC, et l'Université de Málaga, et est la responsabilité du groupe de recherche iArtHis_Lab, en collaboration avec le groupe de recherche "Techné, ingénierie de la connaissance et le produit" de l'Université de Grenade.

1 Introduction

Les bases de données ISOC ont été établis à la fin de 1975 afin de rendre accessibles les informations publiées dans les revues espagnoles de Sciences Sociales et Humaines. ISOC-ART est l'une des bases de données de documents de l'ISOC ; est une base de données de références bibliographiques qui recueille des articles publiés dans des revues espagnoles sur Beaux-Arts (Abejón-Peña, 1993); contient plus de 50 493 références d'articles publiés dans des revues des Beaux-Arts, divers et d'études locales (comprend également les procédures de conférence et articles de compilations); pour l'indexation emploie descripteurs des Beaux-Arts et le Thesaurus de l'ISOC (<http://bddoc.csic.es:8080/index.jsp>)

En Juin 2014, la Université de Málaga et le Centre des Sciences Humaines et Sociales du CSIC ont signé une coopération scientifique et technique pour l'analyse des réseaux de production et de connaissances scientifiques dans le domaine de l'histoire de l'art. Suite à l'accord de le CCHS-CSIC a fourni au groupe de recherche iArtHis_Lab UMA un fichier électronique contenant toute l'ISOC-ART ; fichier avec des informations sur les auteurs, titre d'article, année de publication, lieu de travail, magazine, descripteurs, identificateurs et noms de lieux.

Le groupe iArtHis_Lab emploie méthodes et techniques de l'ingénierie des connaissances dans leurs recherches sur les réseaux de connaissances en histoire de l'art en Espagne. L'ingénierie des connaissances est la branche de l'intelligence artificielle dédiée à la conception et au développement des systèmes de connaissances (Polanco, 1996). Récents travaux de iArtHis_Lab dans cette recherche sont: "*How can knowledge engineering (KE) analysis strategies open new interpretations in the field of art theory and criticism? A case study*" (Rodríguez-Ortega, 2014) et "Application de l'analyse dynamique de réseaux scientifiques à l'étude de l'évolution de la recherche liée au descripteur « histoire de l'art » pendant la période 1976-2012, par ISOC "(Pino-Díaz, 2014-a).

Dans le traitement de la base de données ISOC-ARTE a été utilisé le système de connaissances Techné Coword du groupe de recherche "Techne, ingénierie de la connaissance et du produit", de la Université de Grenade ; un logiciel qui a été très efficace dans le traitement de grands volumes documentaires (voir Pino-Díaz, 2014-b). Ce système expert utilise analyse de mots associés, méthode d'extraction de texte utilisée pour obtenir des groupes de descripteurs (sujets de recherche) les plus pertinentes et importantes du corpus documentaire.

Le but de cet article est de présenter la méthodologie et les premiers résultats, notamment les métriques, obtenus dans les deux analyses, stratégique et dynamique, de la base de données ISOC-ART.

2 Methodologie

Le fichier fourni par la base de données ISOC-ART, « *datosarte.pc* », présente pour chaque champ d'enregistrement: le numéro d'identification (ID), l'auteur (UA), le titre de l'article (TI), une institution des auteurs (OG) le magazine (RV), année de publication (AC), les descripteurs (DE), les identificateurs (IF) et noms (TP) (voir l'exemple dans le Tableau 1):

Tableau 1 : Exemple des champ d'enregistrement de fichier « datosarte.pc » (Base de datosISOC © Consejo Superior de Investigaciones Científicas)

ID: 747254
AU: Córdoba Salmerón, Miguel
TI: Lectura iconográfica-teológica del retablo mayor del antiguo colegio de la Compañía de Jesús en Granada
OG: Univ. Granada, Dep. Historia del Arte, España
RV: Proyección: Teología y Mundo Actual
AC: 2013
DE: Arte barroco; Retablos mayores; Escultura religiosa; Iconografía religiosa; Jesuitas; Escultura en madera; Contrarreforma; Eucaristía; Formación; Alfabetización; Descripción; Interpretación teológica
IF: Compañía de Jesús; Colegio de San Pablo. Granada; Díaz de Ribero, Francisco (1592-1670)
TP: Granada

D'abord, il faut passer le fichier « datosarte.pc » à fichier RIS (voir exemple dans le tableau 2):

Tableau 2 : Exemple des champ d'enregistrement de fichier RIS. (Base de datosISOC © Consejo Superior de Investigaciones Científicas)

TY - GENERIC
ID - 747254
AU - CORDOBA SALMERON, MIGUEL
TI - LECTURA ICONOGRAFICA-TEOLOGICA DEL RETABLO MAYOR DEL ANTIGUO COLEGIO DE LA COMPAÑIA DE JESUS EN GRANADA
AD - UNIV. GRANADA, DEP. HISTORIA DEL ARTE, ESPAÑA
JO - PROYECCION: TEOLOGIA Y MUNDO ACTUAL
PY - 2013
KW - ARTE BARROCO / RETABLOS MAYORES / ESCULTURA RELIGIOSA / ICONOGRAFIA RELIGIOSA / JESUITAS / ESCULTURA EN MADERA / CONTRARREFORMA / EUCARISTIA / FORMACION / ALFABETIZACION / DESCRIPCION / INTERPRETACION TEOLOGICA
M1 - COMPAÑIA DE JESUS / COLEGIO DE SAN PABLO. GRANADA / DIAZ DE RIBERO, FRANCISCO (1592-1670)
U1 - GRANADA
ER -

Puis le fichier RIS a été importé par le gestionnaire de référence Procite: la base de données contient 45 178 enregistrements (12 975 descripteurs, 20 675 auteurs et 1 207 sources documentaires), datant de la période 1925-2014, bien que la plupart d'entre eux indexés à partir de 1975 (année de création de la base de données) (voir les figures 1 et 2)

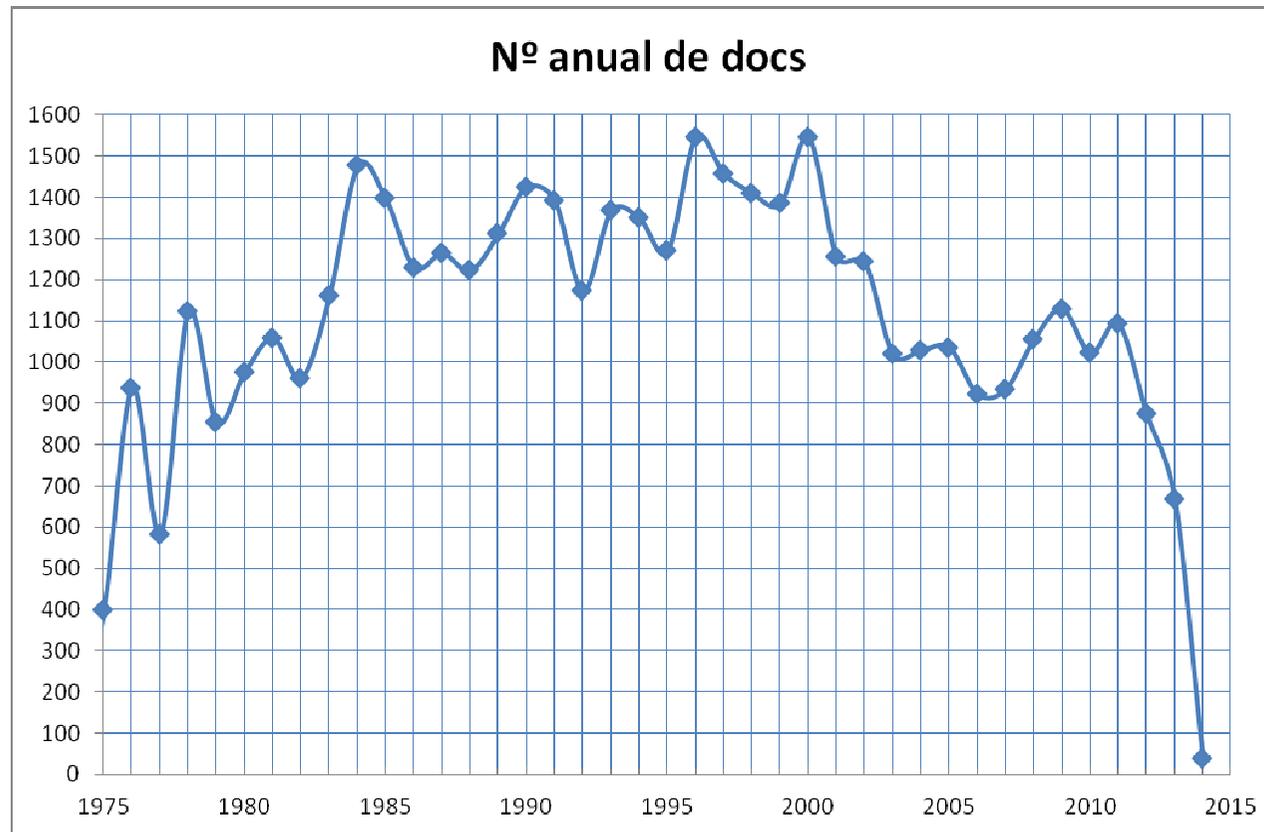


Figure 1 : Nombre annuel de documents

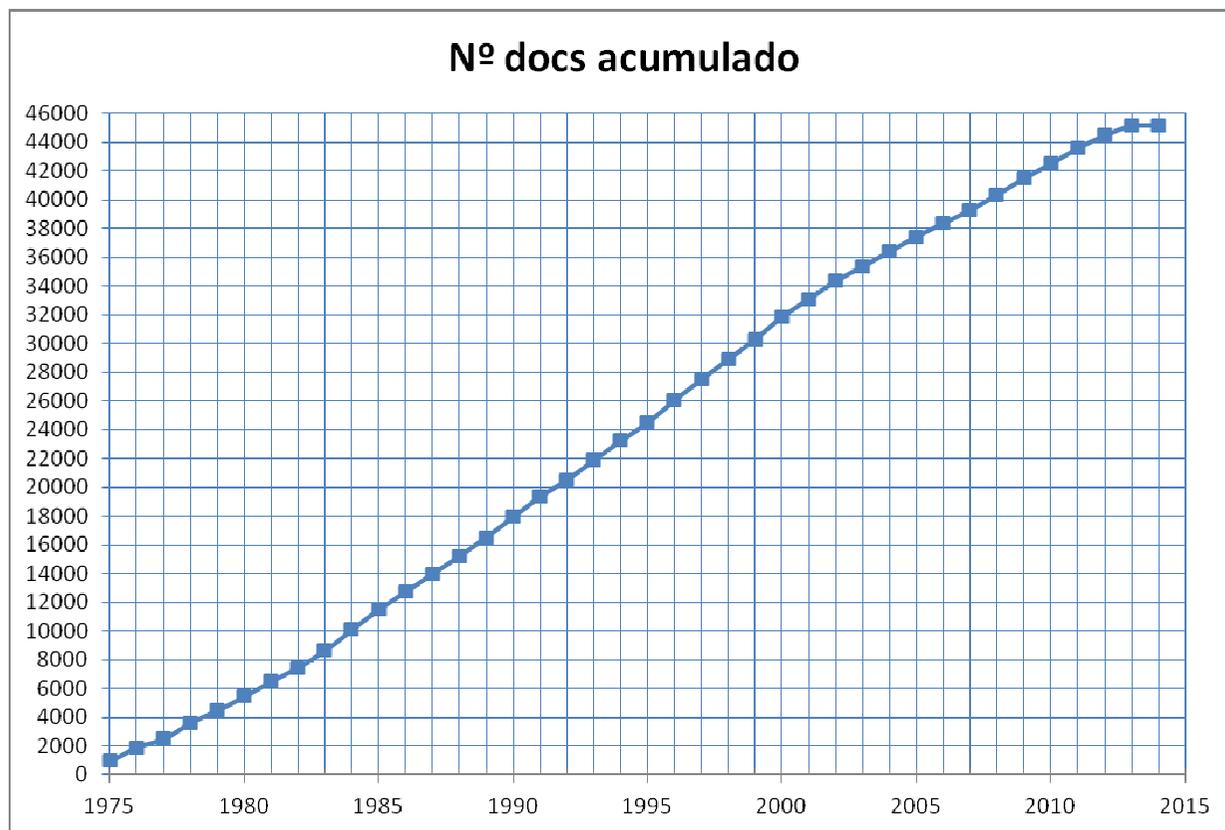


Figure 2 : Nombre annuel de documents accumulés.

Les données ont été exportées vers un fichier TXT pour leur traitement par le système de connaissances Techne Cowords (voir exemple dans le tableau 3):

Tableau 4: Exemple de niveaux de la profondeur de l'analyse stratégique (l'analyse des groupes individuels)

Niveau 1	Niveau 2	Niveau 3	Niveau 4
ISOC –ARTE	Exemple: Groupe de « ARCHITECTURE RELIGIEUSE »	Exemple: Groupe « PEINTURE MURALE », sous-groupe de l'ARCHITECTURE RELIGIEUSE)	Exemple: Groupe « PEINTURE ORNEMENTALE », sous-groupe de la PEINTURE MURALE

c.2) L'analyse des groupes (voir Tableau 5):

Tableau 5: Exemple de niveaux de la profondeur de l'analyse stratégique (l'analyse des plusieurs groupes)

Niveau 1	Niveau 2	Niveau 3	Niveau 4
ISOC –ARTE	Exemple: Groupe de « ARCHITECTURE RELIGIEUSE »	Exemple: Groupe « PEINTURE MURALE + RETABLES + DOCUMENTS INTERGROUPES » (sous-groupes de l'ARCHITECTURE RELIGIEUSE)	Exemple: Groupe « MONASTERES », sous-groupe de [PEINTURE MURALE + RETABLES + DOCUMENTS INTERGROUPES]

c.3) les paramètres de formation des groupes à différents niveaux de profondeur de l'analyse (voir tableau 6):

Tableau 6 : les paramètres formation de groupe à différents niveaux de profondeur de l'analyse.

Niveau 1 → Niveau 2	Niveau 2 → Niveau 3	Niveau 3 → Niveau 4
<ul style="list-style-type: none"> - Occurrence minimum, 60 - Co-occurrence minimum, 30 - Taille du groupe minimum, 4 descripteurs - Taille du groupe maximum, 15 descripteurs 	<ul style="list-style-type: none"> - Occurrence minimum, 4 - Co-occurrence minimum, 2 - Taille du groupe minimum, 4 descripteurs - Taille du groupe maximum, 15 descripteurs 	<ul style="list-style-type: none"> - Occurrence minimum, 3 - Co-occurrence minimum, 2 - Taille du groupe minimum, 4 descripteurs - Taille du groupe maximum, 15 descript.

d) Les paramètres de formation des groupes dans chacune des neuf périodes d'analyse dynamique:

- L'occurrence minimum: 7
- La co-occurrence minimum: 4

- La taille minimum du groupe : 4 descripteurs
- La taille maximum du groupe: 15 descripteurs

3 Résultats

Résultat de l'analyse stratégique des groupes des tableaux 4 et 5 sont obtenus les chiffres suivants des documents, des descripteurs, des auteurs et des sources (descripteurs non liés sont exclus) (voir les Tableaux 7 et 8).

Tableau 7: Les chiffres obtenus dans différents niveaux de profondeur de l'analyse stratégique (l'analyse de groupes individuels de chaque niveau)

Niveau 1	Niveau 2	Niveau 3	Niveau 4
ISOC –ARTE	Exemple: Groupe de « ARCHITECTURE RELIGIEUSE »	Exemple: Groupe « PEINTURE MURALE », sous-groupe de l'ARCHITECTURE RELIGIEUSE.	Exemple: Groupe « PEINTURE ORNEMENTALE », sous-groupe de la « PEINTURE MURALE »
45 178 documents 12 975 descripteurs 20 675 auteurs 1 207 revues	2 943 documents 2 316 descripteurs 1 991 auteurs 314 revues	153 documents 453 descripteurs 163 auteurs 75 revues	24 documents 15 descripteurs 28 auteurs 4 revues

Tableau 8 : Les chiffres obtenus dans différents niveaux de profondeur de l'analyse stratégique (l'analyse de pairs de groupes de chaque niveau)

Niveau 1	Niveau 2	Niveau 3	Niveau 4
ISOC –ARTE	Exemple: Groupe de « ARCHITECTURE RELIGIEUSE »	Exemple: Groupe « PEINTURE MURALE + RETABLES + DOCUMENTS INTERGROUPES », sous-groupes de l'ARCHITECTURE RELIGIEUSE	Exemple: Groupe « MONASTERES », sous-groupe de [PEINTURE MURALE + RETABLES + DOCUMENTS INTERGROUPES]
45 178 documents 12 975 descripteurs 20 675 auteurs 1 207 revues	2 943 documents 2 316 descripteurs 1 991 auteurs 314 revues	262 documents 639 descripteurs 242 auteurs 94 revues	35 documents 15 descripteurs 37 auteurs 26 revues

Dans l'analyse dynamique ont été obtenus les « métriques » suivants de chaque période (descripteurs non liés sont exclus) (voir le Tableau 9):

Tableau 9 : Les périodes de l'analyse dynamique et les « métriques » de chaque période

1925-1974	1975-1979	1980-1984	1985-1989	1990-1994	1995-1999	2000-2004	2005-2009	2010-2014
569 docs. 27 descript. 321 auteurs 36 revues	3 895 docs. 120 descript. 1 647 auteurs 188 revues	5 634 docs. 348 descript. 2 415 auteurs 220 revues	6 428 docs. 354 descript. 3 233 auteurs 300 revues	6 710 docs. 441 descript. 3 808 auteurs 385 revues	7 070 docs. 528 descript. 4 799 auteurs 449 revues	6 095 docs. 636 descript. 4 705 auteurs 438 revues	5 079 docs. 611 descript. 3 977 aut. 431 revues	3 697 docs. 385 descript. 3 431 aut. 431 revues

De l'analyse de la base référentielle on a aussi obtenu les informations suivantes:

- a) Rapports sur les réseaux de connaissances de chaque période (doc)
- b) Diagramme stratégique du réseau de la connaissance (bmp) (voir les exemples de diagrammes stratégiques de la Figure 3)
- c) Listes de descripteurs (alphabétiques et fréquences) (doc et xls)
- d) Listes des auteurs et des revues (classées par nombre d'occurrences) (doc et xls)
- e) Les graphes des groupes de mots dans le diagramme stratégique (bmp)
- f) Diagramme dynamique de la série chronologique (voir la Figure 4)

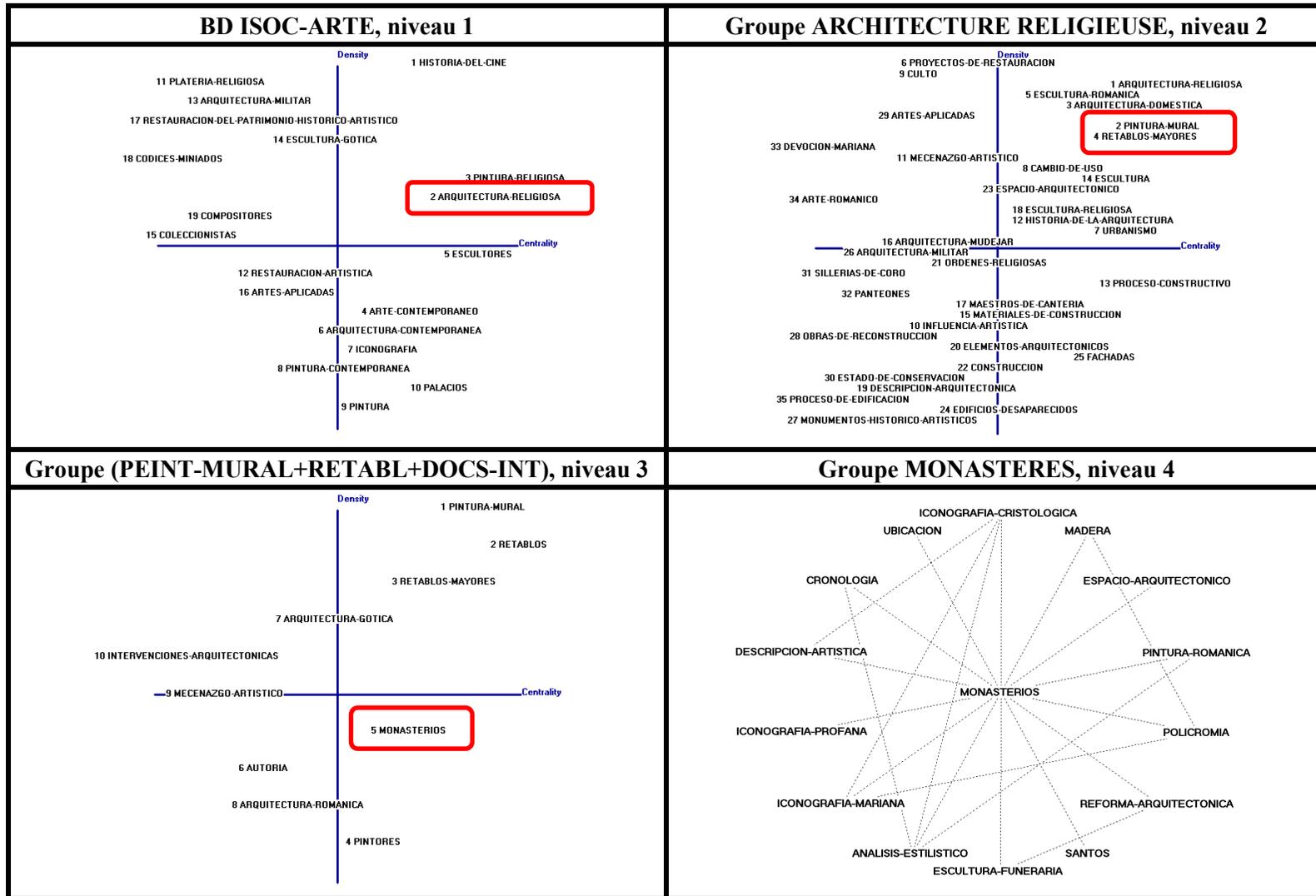


Figure 3 :Le diagramme stratégique des réseaux des différents niveaux de profondeur de l'analyse stratégique.

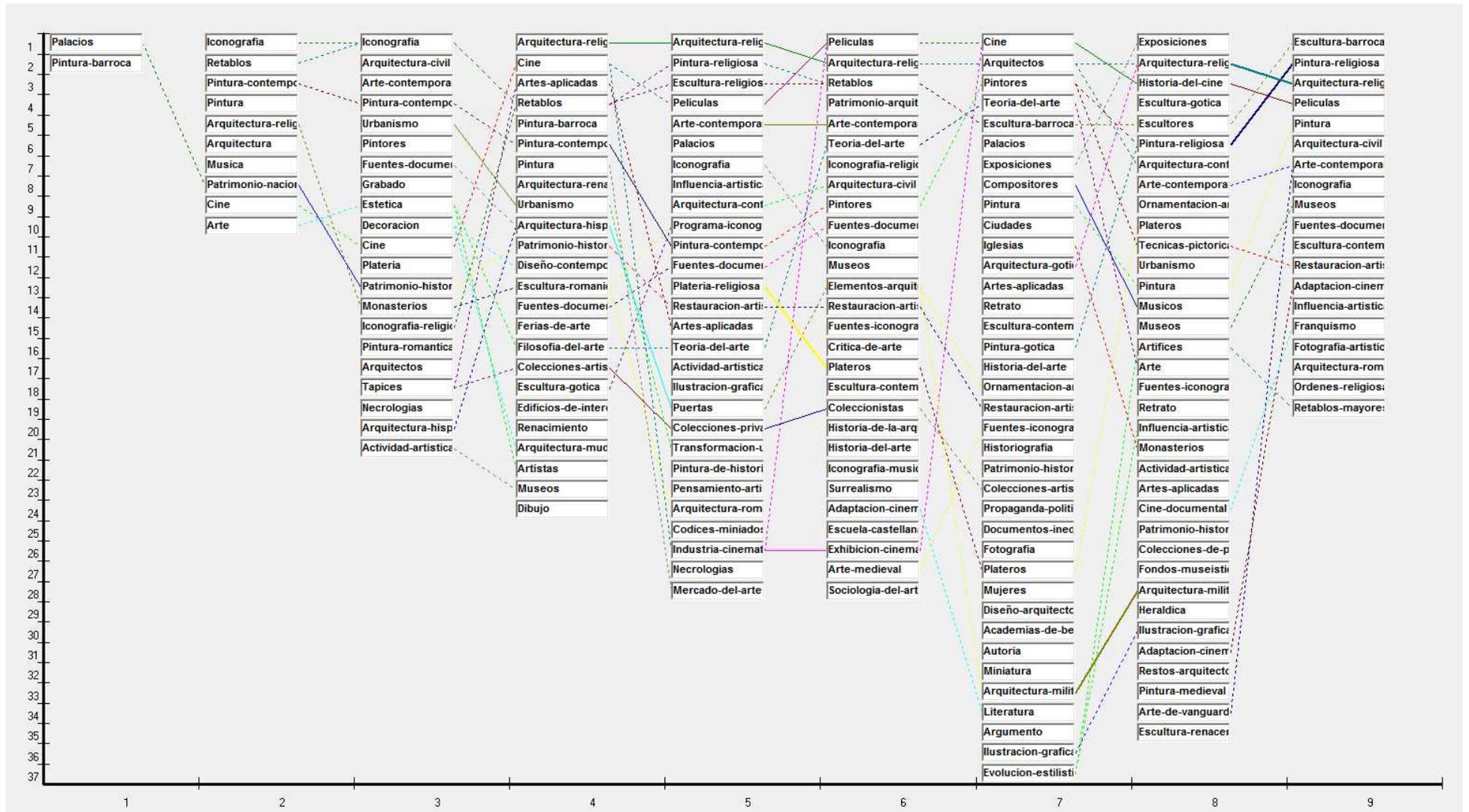


Figure 4 : Le diagramme dynamique des groupes du réseau de la connaissance de chaque période.

Chaque groupe de mots est caractérisé par:

- a) Les paramètres généraux du groupe: nombre de documents, la centralité, la densité, le *rank* de la centralité, le *rank* de la densité et le nombre de descripteurs (voir exemple dans le tableau 10)
- b) Liste des descripteurs qui forment le groupe,
- c) Liste des liens internes avec leur valeurs de l'indice d'équivalence
- d) Liste de liens externes avec leurs valeurs,
- e) Liste des liens avec d'autres groupes de mots
- f) Liste des auteurs et les revues du groupe (classés par nombre d'occurrences) et les paramètres généraux (voir le Tableau 10)

Tableau 10 : Exemple des paramètres généraux du groupe ICONOGRAFÍA du période 1980-1984

Paramètres généraux du groupe « ICONOGRAPHIE » période 1980-1984
Nombre d'enregistrements: 570 Centralité: 34854 Densité: 7266 Rank centralité: 95,2% Rank densité: 90,5% Nombre de descripteurs: 15

4 Conclusion

Les données abondantes et les informations obtenues par le système de connaissances Techné Coword sur les réseaux de connaissances de la base de données ISOC-ART, par *cowords analysis*, *clustering* et *information visualization*, seront analysés par des spécialistes en histoire de l'art dans une perspective globale. L'approche de réseau suppose que les nœuds (descripteurs) sont positionnés par rapport à l'autre selon les relations sémantiques établies entre eux; ces relations changent au fil du temps et donnent lieu à de nouveaux réseaux qui sont retestés. Dans le contexte de chaque période, l'importance ou la pertinence de chaque sujet de recherche, depuis un point de vue stratégique, est donnée par les valeurs de centralité (position par rapport à l'autre) et de densité (force de ses liens).

Pour la première fois l'une des bases de données de références du CSIC a été entièrement analysée à l'aide des outils et des méthodes d'ingénierie des connaissances. La ligne de travail qui se ouvre est très ambitieux et les attentes générées très bonnes. L'étude de l'ISOC-ART a permis favorablement tester la méthodologie d'analyse, utilisé précédemment dans les études partielles d'autres bases de données, mais jamais dans une base de données complète; les résultats sont bons et le "bruit" documentaire minimum, en raison de l'excellent travail de l'indexation effectuée par les documentaristes des bases de données ISOC pendant quarante années de travail.

5 Bibliographie

- [1] ABEJÓN-PEÑA T. *Características y método de elaboración de una base de datos referencial especializada en revistas españolas de Bellas Artes : ISOC-ARTE*, à Bibliotecas de arte, arquitectura y diseño / Art, Architecture and Design Libraries: Perspectivas actuales / Current trends. Barcelona. Actas del Congreso organizado por la Sección de Bibliotecas de Arte de la IFLA, Vol 74 de IFA Publications, 1993.
- [2] POLANCO X., *Infometría e Ingeniería del Conocimiento: Exploración de Datos y Análisis de la Información en vista del Descubrimiento de Conocimientos*. 1997, [en ligne] <http://www.oei.es/salactsi/polanco4.htm> [accès: 29 janvier 2015].
- [3] RODRÍGUEZ-ORTEGA N. et PINO-DÍAZ J., *How can knowledge engineering (KE) analysis strategies open new interpretations in the field of art theory and criticism? A case study*. Conference Digital Memories, Oxford (Royaume Uni), 2014. [en ligne] http://www.inter-disciplinary.net/critical-issues/wp-content/uploads/2014/08/NRodriguezOrtega_DADM1.pdf [accès: : 29 janvier 2015].
- [4] PINO DÍAZ J., CRUCES-RODRÍGUEZ A., BAILÓN-MORENO R. et RODRÍGUEZ-ORTEGA N., *Application de l'analyse dynamique de réseaux scientifiques à l'étude de l'évolution de la recherche liée au descripteur « histoire de l'art » pendant la période 1976-2012, par ISOC*, Seminaire Les Journées d'Intelligence Économique, Big Data Mining, Tanger (Maroc), 2014.
- [5] PINO DÍAZ J., CHIADMI-GARCIA L., CEBRIAN-MENCHERO D. et BAILON-MORENO R., *Text Mining of Scientific Big Data for Decision Making in Conservation of Mediterranean Marine Biodiversity*, Seminaire Les Journées d'Intelligence Économique, Big Data Mining, Tanger (Maroc), 2014.
- [6] BAILON-MORENO R., *Ingeniería del conocimiento y vigilancia tecnológica aplicada a la investigación en el campo de los tensioactivos. Desarrollo de un modelo cuantitativo unificado*, PhD thesis, Universidad de Granada, Granada. Spain, 2003, [en ligne] <http://hdl.handle.net/10481/24728>
- [7] MICHELET B., *L'analyse des associations*, PhD Thesis, Paris: Université de Paris 7, 1988