

CNAPSI
**Corpus de inscripciones nabateas, norarábigas,
palmirenas y siriacas antiguas. Un proyecto en red**

Juan P. Monferrer Sala / Fuensanta Guerrero Carmona / Manuel Marcos Aldón
Universidad de Córdoba
ff1mosaj@uco.es, cc1gucaf@uco.es, lr1maalm@uco.es

Fecha de recepción: 15.06.2012
Fecha de aceptación: 01.09.2012

Resumen: Presentación lingüística, morfológica y descriptiva de un *corpus* de inscripciones en dialectos arameos y norarábigo para su difusión en red con sus representaciones correspondientes.

Palabras clave: Traducción, Nabateo, Norarábigo, Palmireno, Siríaco, Difusión en red, MySQL, Recuperación de información

CNAPSI
**Corpus of Nabatean, North Arabian, Palmyrene and (Old) Syriac
Inscriptions. An on-line project**

Abstract: Our aim in the present article is to give a linguistic, morphologic and descriptive presentation of a corpus of inscriptions in Aramaic and North-Arabian dialects for their dissemination in Internet, together with their respective iconographic representations.

Keywords: Translation, Nabataean, North Arabian, Palmyrene, Old Syriac, on-line diffusion, MySQL, Information Retrieval

Sumario: 1. El marco lingüístico. 2. Desarrollo técnico del proyecto CNAPSI. 2.1. CNAPSI (Modelo de base de datos. 2.2. Normalización. 2.3. Metodología para el diseño de bases de datos de CNAPSI. 2.4. Edición de aplicación. 2.5. Planificación del proyecto. 3. Modelización de datos y herramientas auxiliares. 3.1. Posicionamiento y difusión. 3.2. Control de formación.

1. El marco lingüístico.

Los cuatro registros lingüísticos que concurren en el presente Proyecto de difusión se corresponden respectivamente con cuatro *corpora* textuales correspondientes todos ellos a materiales inscripcionales. Estos cuatro registros, todos ellos registros medios, corresponden a tres dialectos arameos (nabateo, palmireno y siríaco antiguo) y a uno árabe (norarábigo antiguo).

Aun cuando la lengua de los nabateos era la lengua de un grupo étnico árabe, sin embargo utilizaron el arameo en los ámbitos literario e inscripcional¹. El dialecto árabe utilizado por los nabateos suplió al arameo, que fue progresivamente arabizado en el siglo IV AD.² Este proceso gradual es perceptible, por ejemplo, en los préstamos árabes que figuran en los textos arameos, que son, esencialmente, términos técnicos seleccionados por el epigrafista o por quien diseñara el texto que acabó siendo labrado³.

Las afinidades lingüísticas que se aprecian entre lo que podríamos calificar como 'dialecto arameo-árabe nabateo' y el dialecto utilizado por los grupos árabes mesopotámicos durante el periodo neoasirio parecen representar la médula dialectal de la que surgió la lengua de los nabateos⁴. Este 'dialecto protoárabe' que utilizaron los nabateos debe ser claramente distinguido del que usaban sus vecinos árabes de Qēdār y Hagar, por lo que la *Nabataea* debió ser un área lingüística y culturalmente separada de la sociedad norarábigo del momento⁵.

Los materiales inscripcionales nabateos, de variada tipología textual⁶, son de enorme importancia. Tenemos documentadas, en la actualidad, más de un millar de inscripciones de tipo funerario y votivo, las más importantes procedentes de las ciudades de Petra, Basora y al-Ḥiǧr, en las cercanías de Madā'in Šāliḥ, en la Península Arábiga. También contamos con más de tres mil inscripciones breves fragmentarias procedentes del sur de la Península del Sinaí, así como contratos privados y un fragmento hallado en una de las

¹ Jean Starcky, "The Nabataeans: a historical sketch", *Biblical Archaeologist* 18 (1955), cols. 924-926.

² A. F. L. Beeston, "The Inscription Jaussen-Savignac 71", *Proceedings of the Seminar for Arabian Studies* 3 (1973), pp. 69-72.

³ M. O'Connor, "The Arabic Loanwords in Nabatean Aramaic", *Journal of Near Eastern Studies* 45 (1986), pp. 213-229.

⁴ David F. Graf, "The Origin of the Nabateans", *Aram* 2 (1990), pp. 59-68.

⁵ Federico Corriente, "The Phonemic System of Semitic from the Advantage Point of Arabic and its Dialectology", *Aula Orientalis* 23 (2005), p. 171 y la bibliografía allí señalada.

⁶ Cf. Francisco del Río Sánchez, *Textos epigráficos en arameo palmireno, hatreo y nabateo*, «Estudios de Filología Semítica» 6, Barcelona: Universitat de Barcelona, 2006, pp. 97-113.

cuevas del Mar Muerto⁷. Pero a la vuelta del siglo IV AD se advierte la coexistencia de textos árabes y nabateos, además de la mezcla de escritura nabatea y lengua árabe: son los casos, por ejemplo, de la inscripción greco-nabatea –procedente de 'Umm al-Ġimāl– de la tumba de Fih̄r bar Šullay, preceptor del filarca Ġađīmah, de la tribu de los Banū Tanūḥ⁸, así como la célebre inscripción de al-Namārah, escrita con caracteres nabateos, pero a excepción de una palabra (*bar*) en lengua árabe, que sigue suscitando diversos posicionamientos en torno a su correcta lectura⁹

En el caso del dialecto arameo utilizado en Palmira éste representa una variante de la lengua literaria que tiene sus orígenes en el período más antiguo, que desde la época del arameo imperial se erige en como lengua de comunicación en todo el ámbito medio-oriental. Los rasgos específicos del dialecto arameo palmireno aparecen sobre todo en el campo léxico, v.gr. presencia de abundantes préstamos del griego, también del latín, significado particular que adoptan algunos términos con respecto a otros dialectos arameos, así como el sentido incierto que presentan algunos sustantivos relacionados con el comercio, la construcción y una serie de títulos militares. El arameo de Palmira se escribía con una caligrafía que recuerda a la cuadrática hebrea en sus modalidades monumental y simple. Los caracteres aparecen normalmente aislados, con diferencia entre los grafemas inicial/medial y final únicamente en un caso (/n/) y la direccionalidad de la escritura seguía el curso derecha–izquierda.

En cuanto a los textos siríacos, el registro lingüístico empleado en este material inscripcional, calificado como 'siríaco antiguo', se inscribe dentro de la diversidad dialectal de la lengua más hablada en el medio oriental del momento, el arameo, siendo, en esencia, una variante del arameo clásico, pero sin haber desarrollado las características propias del siríaco literario posterior¹⁰, que generará un vasta y rica literatura entre los siglos III y XIII AD. Este registro se encuentra, por lo tanto, en un estadio intermedio entre el arameo clásico y el siríaco literario tardío. Sin embargo,

⁷ F. Rosenthal (ed.), *An Aramaic Handbook*. 4 vols., Wiesbaden: Otto Harrassowitz, 1967.

⁸ Enno Littmann, *Semitic Inscriptions*, Leyden: E. J. Brill, 1949, IV/A, pp. 37-40, n° 41.

⁹ Cf. por ejemplo Jean Cantineau, *Le nabatéen*. 2 vols., Osnabrück: Otto Zeller, ²1978 (= Paris, 1930-32), II, pp. 49-50; W. Caskel, "Die Inschrift von en-Nemāra- Neu gesehen", *Mélanges de l'Université St. Joseph* 45 (1969), pp. 367-379; Irfan Shahīd, "Philological Observations on the Namāra Inscription", *Journal of Semitic Studies* 24 (1979), pp. 33-42; J. Bellamy, "A New Reading of the Namārah Inscription", *Journal of the American Oriental Society* 105/1 (1985), pp. 31-74.

¹⁰ Han J. W. Drijvers, *Old Syriac (Edessan) Inscriptions*, Leiden: E. J. Brill, 1972; H. J. W. Drijvers – John F. Healey, *The Old Syriac Inscriptions of Edessa and Osroene. Texts, Translations, and Commentary*, Leiden: E. J. Brill, 1998.

hay que precisar que siendo el arameo la lengua más hablada, no era la más escrita, pues el griego representaba la lengua de prestigio.

Finalmente, el término ‘norarábigo’ describe a un grupo de dialectos que se presentan, a primera vista, como lingüísticamente homogéneos desde su fase arcaica.¹¹ El repertorio fonético, morfológico y sintagmático halla mayor afinidad con ese grupo dialectal que con cualquier otro externo a este grupo. El ‘norarábigo clásico’, el que conocemos tradicionalmente como ‘árabe clásico’ o simplemente como ‘árabe’, está documentado en un reducido número de inscripciones del siglo IV AD, así como por medio de ejemplos de dialectalismos recogidos en fuentes árabes ulteriores compuestas por autores musulmanes.¹² La más antigua de estas inscripciones es una hallada en el Hawrān, la célebre inscripción de al-Namārah, datada el año 328 AD, en la que el lapicida se sirvió del alfabeto nabateo para la labra del texto.

Los datos norarábigos que se hallan contenidos en materiales inscripcionales surarábigos están constituidos, esencialmente, por antropónimos y topónimos, así como por otros elementos de tipo gramatical.¹³ Contamos, además, con inscripciones qahṭānías,¹⁴ en las que también se encuentran elementos norarábigos, al igual que sucede con las inscripciones halladas en Naḡrān y, sobre todo, en el denominado ‘Himno de Qāniya’ en honor de la diosa Šams, de finales del siglo I AD¹⁵

2.-Desarrollo técnico del proyecto CNAPSI.

2.1.CNAPSI (Modelo de Base de Datos).

La base de datos del CNAPSI consiste en un conjunto de información organizada, de manera que posteriormente se pueda utilizar de forma sencilla y simple. Esta información está relacionada, y está agrupada o estructurada.

Cada base de datos incorporada en el modelo CNAPSI se compone de una o más tablas que guardan un conjunto de datos. Cada tabla consta de

¹¹ M.C.A. Macdonald, “Ancient North Arabian”, en Roger D. Woodard (ed.), *The Ancient Languages of Syria-Palestine and Arabia*, Cambridge: Cambridge University Press, p. 181; para una completa descripción del ‘norarábigo antiguo’: pp. 179-224.

¹² Chaim Rabin, *Ancient West-Arabian*, London: Taylor’s Foreign Press, 1951, *passim*.

¹³ Walter W. Müller, “Das Altarabische und das Klassische Arabisch”, en Wolfdietrich Fischer – Helmut Gätje (eds.), *Grundriss der arabischen Philologie. I. Sprachwissenschaft*, Wiesbaden: Ludwig Reichert, 1982, pp. 26-28.

¹⁴ Se trata del material inscripcional procedente de la célebre ciudad surarábigo de Qahṭān, Christian Robin, *L’arabie antique de Karib’l à Mahomet: nouvelles données sur l’histoire des Arabes grâce aux inscriptions*, Aix-en-Provence: Edisud, 1991-93, pp. 120-121.

¹⁵ C. Robin, *L’arabie antique*, pp. 122-125.

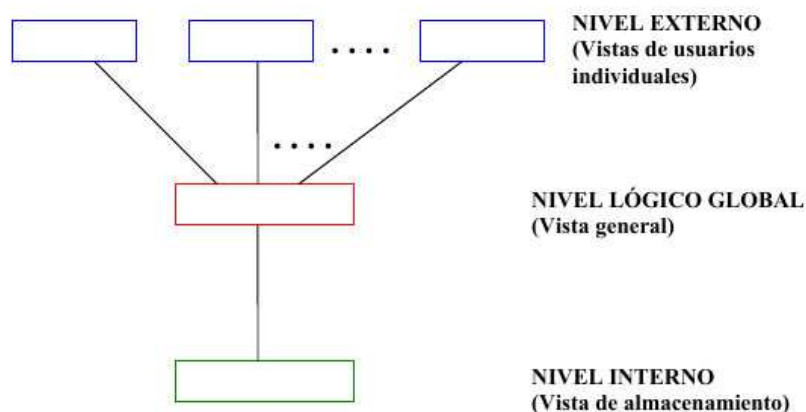
una o más columnas y filas. Las columnas guardan una parte de la información sobre cada elemento que está guardado en la tabla, y cada fila conforma un registro.

Entre las principales características de los sistemas de base de datos podemos mencionar:

- Independencia lógica y física de los datos.
- Redundancia mínima.
- Acceso concurrente por parte de múltiples usuarios.
- Integridad de los datos.
- Consultas complejas optimizadas.
- Seguridad de acceso y auditoría.
- Respaldo y recuperación.
- Acceso a través de lenguajes de programación estándar.

Un sistema de administración de bases de datos (DBMS, Database Management System) integrado en el sistema CNAPSI proporciona los mecanismos necesarios para almacenar, organizar, obtener y modificar datos para muchos usuarios. Los sistemas de administración de bases de datos permiten el acceso y almacenamiento de datos sin necesidad de preocuparse por su representación interna. Proporciona una forma de almacenar y recuperar la información de una base de datos de forma práctica y eficiente¹⁶.

Se puede representar la estructura del DBMS de CNAPSI de la siguiente forma:



¹⁶ M. J. Donahoo, *SQL. Practical Guide for Developers*, USA: Morgan Kaufmann Elsevier Inc, 2005, pp. 2-3.

El nivel interno es el más cercano al almacenamiento físico, es decir, es el relacionado con la forma en que los datos están realmente almacenados. El nivel externo es el más cercano a los usuarios, por lo tanto es el relacionado con la forma en que los datos son vistos por cada usuario individualmente. El nivel lógico global es un nivel intermedio entre los dos anteriores.

En la actualidad, los sistemas de bases de datos más populares son las bases de datos relacionales. En el proyecto CNAPSI el lenguaje empleado en su creación es SQL (lenguaje de consulta estructurado o Structured Query Language) pues es el estándar internacional que se utiliza casi universalmente con las bases de datos relacionales para realizar consultas y manipular datos. Dentro de estos estándares se ha optado por el sistema de administración de bases de datos relacionales MySQL¹⁷, el cual es un motor de base de datos relacional, muy rápido, multihilo, multiusuario y robusto, ampliamente utilizado no sólo en proyectos de software libre, sino también en multitud de desarrollos comerciales¹⁸, ya que dispone de dos versiones, una comercial y otra libre con licencia GNU GPL (del inglés GNU General Public License, Licencia Pública General de GNU)¹⁹. Hasta abril de 2009 pertenecía a Sun Microsystems, pero en esa fecha ha pasado a formar parte de Oracle, ya que ésta empresa ha adquirido a Sun Microsystems, por lo que se ha optado por mantener este sistema que permite tanto aprovechar los logros e implementaciones del Software Libre y el mantenimiento del software propietario.

2.2. Normalización

Debemos normalizar la BBDD del CNAPSI para que no haya redundancia superflua. La teoría de la normalización se ha desarrollado para obtener estructuras de datos eficientes que eviten las anomalías de actualización. El concepto de normalización fue introducido por E.D. Codd²⁰ para aplicarse a sistemas relacionales. La normalización es la expresión formal del modo de realizar un buen diseño pues provee los medios necesarios para describir la estructura lógica de los datos en un sistema de información.

¹⁷ T. Converse, *PHP 5 and MySQL Bible*, USA: Wiley Publishing, 2004, pp 245-247.

¹⁸ A. Cobo, *PHP y MySQL: tecnologías para el desarrollo de aplicaciones web*, Madrid: Ediciones Díaz de Santos, 2007, pp 339-341.

¹⁹ GNU es un acrónimo recursivo que significa GNU No es Unix, proyecto iniciado en 1983 por Richard Stallman, con el objetivo de crear un sistema operativo completamente libre.

²⁰ E. F. Codd, *A Relational Model of Data for Shared Data Banks*, USA: IBM Research Laboratory, 1970.

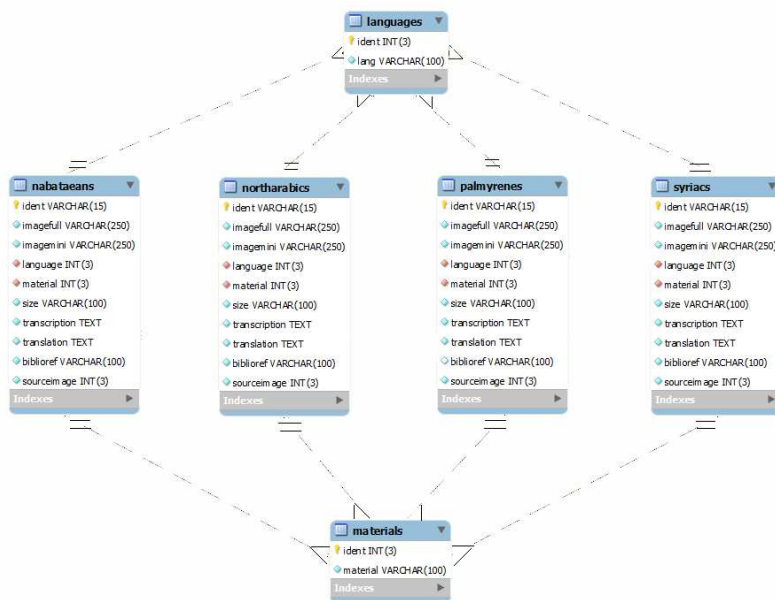
La normalización convierte una relación en varias sub-relaciones, cada una de las cuales obedece a reglas. Estas reglas se describen en términos de dependencia. Una vez que hayamos examinado las distintas formas de dependencia, encontraremos procedimientos a aplicar a las relaciones de modo tal que las mismas puedan descomponerse de acuerdo a la dependencia que prevalece. Esto nos llevará indefectiblemente a formar varias sub-relaciones a partir de la única relación preexistente.

2.3. Metodología para el diseño de bases de datos de CNAPSI

- Determinación de entidades y atributos
- Normalización de entidades
- Determinación de relaciones (DER)
- Obtención del modelo lógico global de los datos
- Diseño físico de la base de datos

Cuando se va a realizar el diseño de la base de datos para un sistema, es necesario determinar los datos que se deben tomar en cuenta y las dependencias funcionales existentes entre ellos.

Después de la normalización, las tablas de la base de datos, así como las relaciones entre ellas, son las siguientes:



2.4. Ejemplo de la aplicación.

Una vez realizada la base de datos, normalizadas las tablas y llevadas a la Web, la página inicial de la aplicación se divide en:

- una cabecera en la parte superior
- los menús para realizar búsquedas en cada una de las tablas
- la zona principal en la que aparece el texto y las imágenes
- la parte lateral derecha en la que aparecen enlaces de interés e información de contacto

CNAPSI
Catalogue
of Nabataean, North Arabian,
Palmyrene & Syriac
Inscriptions

HOME NABATAEAN NORD ARABIC PALMYRENE SYRIAC

CNAPSI
 The objective of the CNAPSI project, which began in 2011 is to collect the published Ancient Nabataean, Palmyrene, Nord Arabic and Syriac inscriptions. New corpora have been added in recent years: the catalogue of the British Museum inscriptions, the Louvre Museum, the Hermitage Museum, the Samlinger Museum, Palmyra Museum and other sites. Up to the present, the published corpus of texts from Nabataean, Palmyrene, Nord Arabian and Syria is relatively small, but a big edition of new texts is going to be published.

Nabataean Inscriptions
[Corpus of Inscriptions](#)

Palmyrene Inscriptions
[Corpus of Inscriptions](#)

Nord Arabic Inscriptions
[Corpus of Inscriptions](#)

Syriac Inscriptions
[Corpus of Inscriptions](#)

Related Links

[Corpus of South Arabian Inscriptions](#)

[Dipartimento di Scienze Storiche del Mondo Antico, University of Pisa](#)

[Oriental Institute of the University of Chicago](#)

[Department Oriental Antiques, Louvre Museum](#)

[Department of the Middle East, British Museum](#)

[The Oriental Department, Hermitage Museum](#)

Contact Information

Partner Leader:
 Juan Pedro Monferrer Sala,
 University of Córdoba

Scientific Staff:
 Francisco del Río, University of Barcelona
 Manuel Marcos Aldón, University of Córdoba
 Fuensanta Guerrero Carmona, University of Córdoba

W3C quality assurance: [xhtml 1.0 strict](#) | [CSS validator](#) | [credits](#)

Como ejemplo, se muestra el resultado de una búsqueda:

CNAPSI

Catalogue
of Nabatean, North Arabian,
Palmyrene & Syriac
Inscriptions



UNIVERSIDAD DE CORDOBA

HOME
NABARBEAN
NORD ARABIC
PALMYRENE
SYRIAC

CNAPSI
Palmyrene Inscriptions

Palmyrene Bust



Materials: limestone
Techniques: relief
Production place: Made in Palmyra (Asia, Middle East; Levant, Syria, Central Syria, Homs (governorate), Palmyra)
Place (findspot): Excavated/Findspot Palmyra (Asia, Middle East, Levant, Syria, Central Syria, Homs (governorate), Palmyra)
Date: 50-150
Description: Limestone bust from a Palmyrene funerary relief; seated woman; left hand holding distaff and spindle; right hand raised with palm outward and upward; inscription; 6 ll; worn; abraded; edges badly chipped.
Inscription Translation: Marta, daughter of Malku, son of Yarhibolē.
Dimensions:
 Height: 54.61 centimetres
 Width: 40.18 centimetres

Bibliography: R. F. Harper, "Babylonian Letter. The Joseph Shemtob Collection of Babylonian Antiquities recently purchased for the University of Pennsylvania", *Hebraica* 5 (October 1888), pp. 74-76.
Biography: Dealer in antiquities; sold large numbers of cuneiform tablets and a smaller quantity of other objects from Mesopotamia to The British Museum between 1877 and 1893, plus a number of tablets to the University of Philadelphia in Pennsylvania. NB There may be some confusion as some collections are registered as being published from one Mr Raphael Shemtob of Islington (e.g. 1877-11-14 coll.).

Related Links

- Corpus of South Arabian Inscriptions
- Dipartimento di Scienze Storiche del Mondo Antico. University of Pisa
- Oriental Institute of the University of Chicago
- Department Oriental Antiques. Louvre Museum
- Department of the Middle East. British Museum
- The Oriental Department. Hermitage Museum

Contact Information

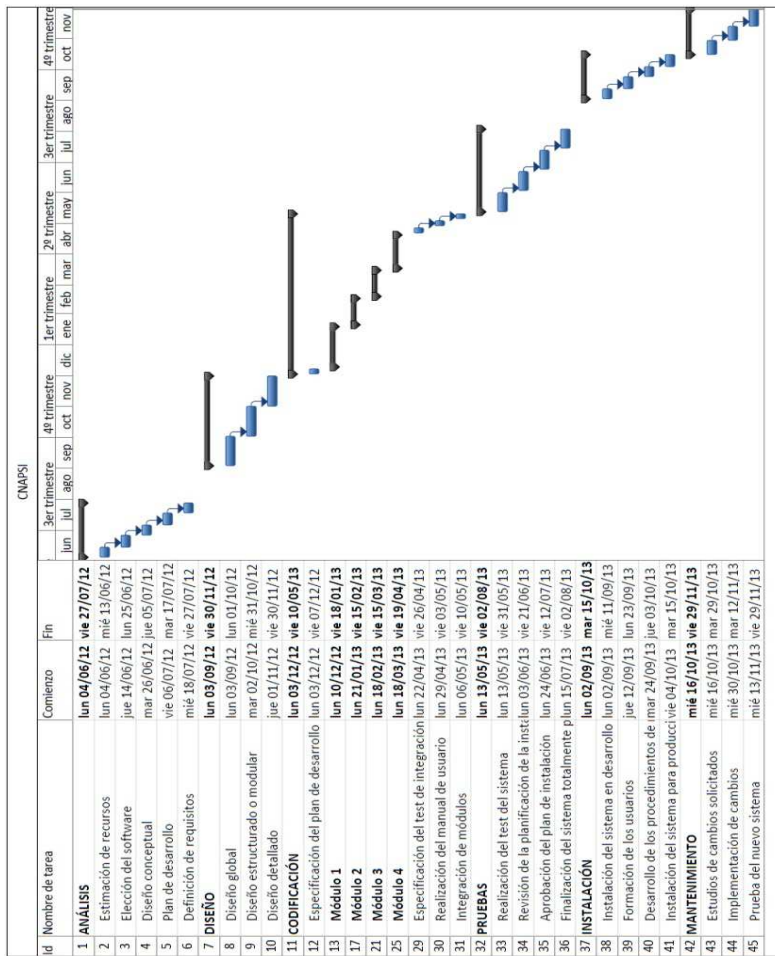
Partner Leader:
Juan Pedro Monferrer Sala. University of Córdoba

Scientific Staff:
Francisco del Río. University of Barcelona
Manuel Marcos Aldón. University of Córdoba
Fuensanta Guerrero Carmona. University of Córdoba

W3C quality assurance: [xhtml 1.0 strict](#) | [CSS validator](#) | [credits](#)

2.5. Planificación del proyecto.

En un proyecto de esta envergadura se debe realizar una planificación adecuada, ajustada y realista, teniendo en cuenta los recursos materiales y humanos, así como el esfuerzo requerido para que se cumplan los plazos establecidos y se lleve a cabo en el tiempo acordado. La programación del proyecto es la siguiente:



3. Modelización de datos y herramientas auxiliares.

Para llevar a cabo este proyecto en el Laboratorio de Ingeniería Documental de la Universidad de Córdoba (LabInDoc) se desarrollaron modelos que se sustentaban en diferentes formas en las que se tratan la representación, almacenamiento, organización y acceso de items de información, generalmente documental y ofrecer al usuario sugerencias específicas sobre items o acciones dentro de un dominio específico²¹.

Entre los sistemas de integración de datos hemos optado en la implantación en CNAPSI de un sistema de recomendación, es decir, una herramienta construida sobre la implicación de una red bayesiana: este tipo de sistemas de recomendación consta de dos componentes a) uno cualitativo, representado por un grafo dirigido y acíclico compuesto por nodos que representan las variables aleatorias del problema que se desea resolver, y arcos que indican causalidad, relevancia o relaciones de dependencia entre variables, y (b) otro cuantitativo, codificado mediante distribuciones de probabilidad condicionadas. Éstas permiten cuantificar nuestra incertidumbre sobre las relaciones existentes entre las variables del problema. Resolviendo así los problemas entrópicos.

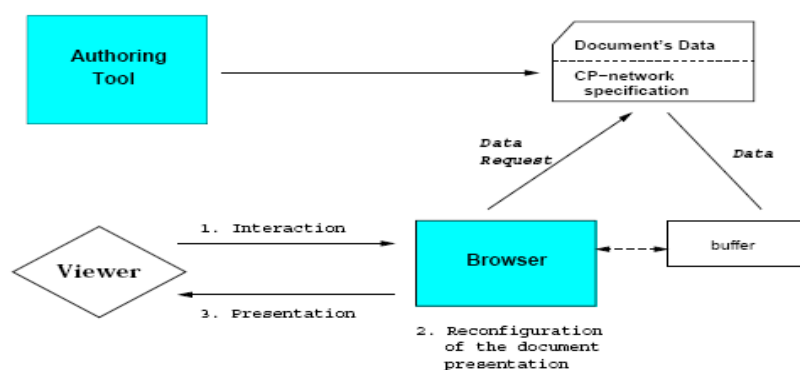
Por tanto, formalmente, una red bayesiana puede considerarse como una representación muy eficiente de una distribución de probabilidad conjunta, que tiene en cuenta las relaciones de independencia condicional codificadas en el grafo; la distribución de probabilidad conjunta puede obtenerse a partir de las distribuciones condicionales mediante:

$$p(X_1, X_2, \dots, X_n) = \prod_{i=1}^n p(X_i | pa(X_i))$$

Estos trabajos se fundamentan en labores básicas de desarrollo en recuperación de información probabilística, en los que se integran algoritmos y funciones de agrupamiento con formalismos de Fuzzy Logic, formalismos que hasta el momento no han sido utilizados para tratar la incertidumbre intrínseca inherente a multitud de problemas de

²¹ Baeza-Yates, R. A., and Ribeiro-Neto, B. (2010). *Modern Information Retrieval* (2nd ed.). Reading, Massachusetts: Addison-Wesley; Croft, W. B., Metzler, D., and Strohman, T. (2010). *Search Engines: Information Retrieval in Practice*. London, England: Pearson.; Grossman, D. A., and Frieder, O. (2004). *Information Retrieval: Algorithms and Heuristics* (2nd ed.). Berlin, Germany: Springer; Hearst, M. A. (2009). *Search User Interfaces*. Cambridge, England: Cambridge University Press; Manning, C. D., Raghavan, P., and Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge, England: Cambridge University Press; Manning, C. D., and Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, Massachusetts: MIT Press; Özsu, M. T., and Liu, L., editors (2009). *Encyclopedia of Database Systems*. Berlin, Germany: Springer.

documentación literaria y se ha escogido para ello un desarrollo de redes bayesianas y en la modificación del algoritmo de recuperación naïve-Bayes, obteniendo un nuevo sistema más voraz y eficaz, aunque en un principio, su empleo depende de la evaluación de las fuentes de datos que se han empleado²².



Esquema de configuración del algoritmo semi naïve-Bayes empleado

Para desarrollar el algoritmo se utilizaron variaciones de métodos ya establecidos teniendo siempre presente que en nuestro caso no es una recuperación de datos sino una aproximación de probabilidades así que se empleó un desarrollo de inferencia estadística, los dos problemas fundamentales que estudia la inferencia estadística son el problema de la estimación y el problema del contraste de hipótesis²³.

La estimación de un parámetro involucra el uso de los datos muestrales en conjunción con alguna estadística. Existen dos formas de llevar a cabo lo anterior: estimación puntual y estimación por intervalo. En la primera se busca un estimador, que, con base en los datos muestrales, de origen a una estimación del valor del parámetro, y que recibe el nombre de estimador puntual. Para la segunda, se determina un intervalo en el que, en forma probable, se encuentra el valor del parámetro. Este intervalo recibe el

²² Artandi, S., "Information Concepts and Their Utility", *Journal of the American Society for Information Science*, 24 (1973), pp. 242-245 Baird, J.C., Information Theory and Information Processing. *Information Processing & Management* 20:3 (1984), pp. 373-381; Belkin, N.J., "Some Soviet Concepts of Information for Information Science", *Journal of the American Society for Information Science* 26 (1975), pp. 56-64; Belkin, N.J., "Information Concepts for Information Science", *Journal of Documentation* 34 (1978), pp. 55-85; Beniger, J.R., Information and Communication: The New Convergence. *Communication Research* 15:2 (1988), pp. 198-218.

²³ Roberts, N., "Social Considerations Towards a Definition of Information Science", *Journal of Documentation* 32 (1976), pp. 249-257.

nombre de intervalo de confianza estimado. Es importante observar que el estimador no es un valor concreto sino una variable aleatoria, ya que aunque depende unívocamente de los valores de la muestra observados, la elección de la muestra es un proceso aleatorio en nuestro caso, pues se pensó en el desarrollo de un validador para uso indiscriminado y no especializado

En nuestro caso las variables aleatorias se obtuvieron del proceso de peso de los sinónimos que se validaron utilizando variaciones de fuzzy-logic. Las características deseables eran: consistencia, cuando el tamaño de la muestra crece arbitrariamente, el valor estimado se aproximaba al parámetro desconocido; carencia de sesgo: el valor medio que se obtenía de la estimación para diferentes muestras era el valor del parámetro; eficiencia: buscaba que la dispersión del estimador con respecto al valor central, es decir, la varianza, sea tan pequeña como sea posible; suficiencia: el estimador debería aprovechar toda la información existente en la muestra.

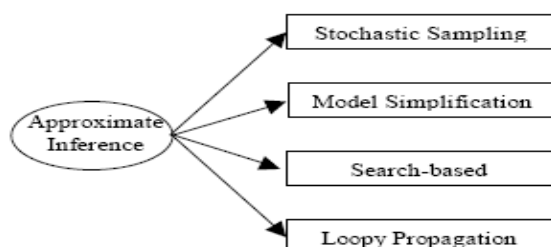
El sistema intuitivo está definido por la función:

$$P(C_i^{T+1}|S^T) = \sum_{A^{T+1}} P(A^{T+1}|S^T)P(C_i^{T+1}|A^{T+1}, S^T)$$

3.1. Posicionamiento y difusión.

La propagación de evidencia, es decir, el cálculo de las probabilidades a posteriori de todas o algunas de las variables (o incluso de las distribuciones conjuntas de grupos de variables), dada alguna evidencia o conocimiento relativo a los valores que han tomado determinadas variables (las cuales se dice que han sido instanciadas) alcanzará en nuestro caso en torno al 94-96%. La realidad de difusión óptima se encuentra entre el 75-86% por lo cual nuestra opción de page rank es extremadamente óptima. Existen diversos algoritmos que realizan este proceso de propagación de la forma más eficiente posible, que depende en gran medida de la topología del grafo²⁴.

²⁴ Marchionini, G. - G. Crane, G., "Evaluating Hypermedia and Learning: Methods and Results from the Perseus Project", *ACM Transactions on Information Systems* 12:1 (1994), pp. 5-34.



Categorías del algoritmo de aproximación según el método Markov

Existen modelos sobre el algoritmo ya empleados que se han utilizado como el enfoque *APNBC* (*Adjusted Probability Naïve-Bayesian Induction*) realiza un ajuste lineal del peso de la probabilidad de cada término. De esta forma, se asigna un factor de ajuste a cada término y documento, y la probabilidad estimada para cada atributo se multiplica por este factor. Los pesos relativos a cada término se buscan a través de un algoritmo voraz guiado por el tanto por ciento de bien clasificados, el cual se obtiene por el método de resustitución²⁵.

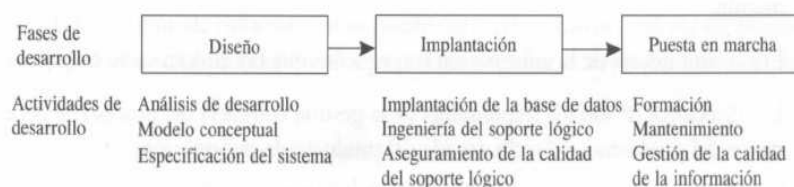
A través de *WenBay* (*Weighted naïve-Bayes*) se permite asociar pesos a particiones pero con una adecuada modificación permite asociar los pesos a los elementos instanciados en los documentos y en sus particiones, en lugar de a cada clase. Las instancias de cada variable predictora son particionadas de manera recursiva con el objetivo de maximizar la entropía

3.2. Control de información.

Puesto que la información generada por el sistema existente ha ido en función de todas las actividades completadas en las diversas fases del ciclo de desarrollo del sistema, es preciso prestar especial atención a cada fase en lo que respecta a cualquier posibilidad de mejora de la calidad de la información²⁶.

²⁵ Bellman, R.E. - R. Kalaba - L.A. Zadeh, Abstraction and pattern classification. RAND Memo, RM-4307-PR. [Online] RAND. Consulta realizada el 27/02/2010.

²⁶ Xie, H., "Shifts of Interactive Intentions and Information-Seeking Strategies in Interactive Information Retrieval", *Journal of the American Society for Information Science* 51:9 (2000), pp. 841-857.



Para lograr un control de autoridades en los procesos de recomendación hemos sistematizado según Souza todos los elementos descriptores, para ello se presentan de forma sistemática en nueve facetas y otra alfabética. Para analizar otros elementos relacionados con la Gestión de Calidad Total de los Datos, teniendo como norte que partíamos de un control terminológico exhaustivo con fuentes de reconocida fiabilidad como hemos referido anteriormente, empleamos la gestión transversal de los datos en consulta con *Integrity Analyzer* (IA). Es uno de los productos resultado de los estudios realizados por MIT en materia de gestión de la calidad total de los datos (TDQM), que presenta una metodología TDQM en la que se combinan los principios del ciclo aplicable a este tipo de gestión con los correspondientes a las restricciones de integridad en las bases de datos relacionales:

Filiaciones: habitualmente, las herramientas de contextualización no identifican la filiación de todos y cada uno de los términos de un artículo. Esta es la información que lógicamente puede ofrecer esta herramienta. No obstante es posible identificar la filiación de la mayor parte de términos haciendo un barrido de la base de datos simultáneo a los vocabularios maestros de materias, autores, centros y países, etc.

Para cada período se pueden establecer propiedades particulares del vocabulario, tales como número de descriptores y las características de la lista: libre, truncada o redondeada. Además debe elegirse la ocurrencia mínima de los descriptores y la co-ocurrencia también mínima que deben poseer los enlaces candidatos para formar parte de la contextualización. Jugando con estos valores es posible obtener resultados verdaderamente interesantes y explicativos

El *Integrity Analyzer* funcionó tanto con las restricciones formales de integridad de Codd como con las restricciones definidas por el usuario. La evaluación de la integridad de dominio, de columna, de entidad y referencial permanecieron sin cambios en todas las aplicaciones. Todos los sistemas de gestión de base de datos relacionales, de forma idónea, deberían seguir estas normas. Asimismo, todas las bases de datos, al margen del tema que

recojan, deberían adherirse a las normas que imponen este tipo de restricciones.

Referencias bibliográficas

- ARTANDI, S. (1973): "Information Concepts and Their Utility", *Journal of the American Society for Information Science*, 24, 242-245.
- BAEZA-YATES, R. A.; RIBEIRO-NETO, B. (2010): *Modern Information Retrieval*. Reading, Mass: Addison-Wesley.
- BAIRD, J.C. (1984): Information Theory and Information Processing. *Information Processing & Management* 20:3, 373-381.
- BEESTON, A.F.L. (1973): "The Inscription Jausse-Savignac 71", *Proceedings of the Seminar for Arabian Studies* 3, 69-72.
- BELKIN, N.J. (1978): "Information Concepts for Information Science", *Journal of Documentation* 34, 55-85.
- _____ (1975): "Some Soviet Concepts of Information for Information Science", *Journal of the American Society for Information Science* 26, 56-64.
- BELLAMY, J. (1985): "A New Reading of the Namārah Inscription", *Journal of the American Oriental Society* 105/1, 31-74.
- BELLMAN, R.E.; KALABA R.; ZADEH, L. A.; Abstraction and pattern classification. RAND Memo, RM-4307-PR. [Online] RAND. COnsulta realizada el 27/02/2010.
- BENIGER, J.R. (1988): Information and Communication: The New Convergence. *Communication Research* 15:2, 198-218.
- CANTINEAU, J. (1978²): *Le nabatéen*. 2 vols., Osnabrück: Otto Zeller. (= Paris, 1930-32).
- CASKEL, W. (1969): "Die Inschrift von en-Nemāra- Neu gesehen", *Mélanges de l'Université St. Joseph* 45, 367-379.
- COBO, A. (2007): *PHP y MySQL: tecnologías para el desarrollo de aplicaciones web*, Madrid: Ediciones Díaz de Santos.
- CODD, E.F. (1970): *A Relational Model of Data for Shared Data Banks*, USA: IBM Research Laboratory.
- CONVERSE, T. (2004): *PHP 5 and MySQL Bible*, USA: Wiley Publishing.
- CORRIENTE, F. (2005): "The Phonemic System of Semitic from the Advantage Point of Arabic and its Dialectology", *Aula Orientalis* 23, 169-173.
- CROFT, W. B. (2010): Metzler, D. - T. Strohman, *Search Engines: Information Retrieval in Practice*. London: Pearson.
- DONAHOO, M.J. (2005): *SQL. Practical Guide for Developers*, USA: Morgan Kaufmann Elsevier Inc.

- DRIJVERS, H.J.W.; HEALEY, J. F. (1998): *The Old Syriac Inscriptions of Edessa and Osrhoene. Texts, Translations, and Commentary*, Leiden: E. J. Brill.
- DRIJVERS, H.J.W. (1972): *Old Syriac (Edessan) Inscriptions*, Leiden: E. J. Brill, 1972
- GRAF, D. F. (1990): "The Origin of the Nabateans", *Aram* 2, 59-68.
- GROSSMAN, D. A.; FRIEDER, O. (2004): *Information Retrieval: Algorithms and Heuristics*, Berlin: Springer. 2.^a ed.
- HEARST, M.A. (2009): *Search User Interfaces*. Cambridge: Cambridge University Press.
- LITTMANN, E. (1949): *Semitic Inscriptions*, Leyden: E. J. Brill, IV/A, 37-40, n^o 41.
- MACDONALD, M.C.A., "Ancient North Arabian", en Roger D. Woodard (ed.), *The Ancient Languages of Syria-Palestine and Arabia*, Cambridge: Cambridge University Press, 179-224.
- MANNING, C. D.; RAGHAVAN, P.; SCHÜTZE, H. (2008): *Introduction to Information Retrieval*, Cambridge: Cambridge University Press.
- MANNING, C.D.; SCHÜTZE, H. (1999): *Foundations of Statistical Natural Language Processing*, Cambridge, Mass.: MIT Press.
- MARCHIONINI, G.; CRANE, G. (1994): "Evaluating Hypermedia and Learning: Methods and Results from the Perseus Project", *ACM Transactions on Information Systems* 12:1, 5-34.
- MÜLLER, W. (1986): "Das Altarabische und das Klassische Arabisch", en FISCHER, W.– GÄTJE, H. (eds.), *Grundriss der arabischen Philologie. I. Sprachwissenschaft*, Wiesbaden: Ludwig Reichert, 1982, pp. 26-28.
- O'Connor, M., "The Arabic Loanwords in Nabatean Aramaic", *Journal of Near Eastern Studies* 45, 213-229.
- ÖZSU, M. T.; LIU, E. (2009, eds.): *Encyclopedia of Database Systems*, Berlin: Springer.
- RABIN, Ch. (1951): *Ancient West-Arabian*, London: Taylor's Foreign Press.
- RÍO SÁNCHEZ, F. (2006): del, *Textos epigráficos en arameo palmireno, hatreo y nabateo*, «Estudios de Filología Semítica» 6, Barcelona: Universitat de Barcelona.
- ROBERTS, N. (1976): "Social Considerations Towards a Definition of Information Science", *Journal of Documentation* 32, 249-257.
- ROBIN, Ch. (1991-93): *L'arabie antique de Karib'il à Mahomet: nouvelles données sur l'histoire des Arabes grâce aux inscriptions*, Aix-en-Provence: Edisud.
- ROSENTHAL, F. (1967, ed.): *An Aramaic Handbook*. 4 vols., Wiesbaden: Otto Harrassowitz.
- SHAHÍD, I. (1979): "Philological Observations on the Namāra Inscription", *Journal of Semitic Studies* 24, 33-42.

- STARCKY, J. (1955): "The Nabataeans: a historical sketch", *Biblical Archaeologist* 18, 924-926.
- XIE, H. (2000): "Shifts of Interactive Intentions and Information-Seeking Strategies in Interactive Information Retrieval", *Journal of the American Society for Information Science* 51:9, 841-857.