

## Visualización y análisis de mapas temáticos de la actividad científica mexicana sobre cáncer (1977-2012)

HERRERA-MIRANDA, Israel\*† y HERRERA-MIRANDA, Miguel

*Cipes, UA Matemáticas, Universidad Autónoma de Guerrero, Chilpancingo, Gro..*

Recibido Enero 5, 2015; Aceptado Junio 3, 2015

### Resumen

Se obtuvieron los registros bibliográficos de artículos científicos sobre cáncer publicados en revistas de alto impacto a nivel internacional. Se consultaron para ello las bases de datos de Thomson Reuters: *Science Citation Index SCI*, *Social Science Citation Index SSCI* y *los Conference Proceedings* en Julio de 2013. El periodo de búsqueda abarcó de 1977 al 2012 de los trabajos sobre cáncer realizados por al menos un autor con institución de adscripción en México. Se utilizó el campo resumen de los registros mencionados para ser procesados por los programas de software Bibexcel y VOSviewer con la finalidad de generar mapas temáticos. Con la visualización de estos mapas por periodos (ventanas en el tiempo) se realizó un análisis cualitativo de las tendencias en este campo de dominio de la investigación científica mexicana.

**Indicadores bibliométricos, mapas de la ciencia, mapas bibliométricos, mapas disciplinares, investigación científica, cáncer, México, análisis de contenido, minería de datos.**

### Abstract

Bibliographic records were obtained from scientific articles on Mexican cancer research published in high impact journals worldwide. The bibliographic records were retrieved from the Thomson Reuters databases: *Science Citation Index SCI*, *SSCI Social Science Citation Index* and *Conference Proceedings* on July 2012. The search covered the period 1977 to 2012 of the published works on cancer research conducted by at least one author whose affiliation institution were from Mexico. It was used the field abstract of the bibliographic records to be processed by the software programs Bibexcel and VOSviewer in order to generate thematic maps. With these maps displayed by periods of time it was performed a qualitative analysis of the trends in this field of scientific mexican research.

**Bibliometric indicators, science maps, bibliometric maps, disciplinary maps, scientific research, cancer, Mexico, content analysis, data mining.**

**Citación:** HERRERA-MIRANDA, Israel y HERRERA-MIRANDA, Miguel. Visualización y análisis de mapas temáticos de la actividad científica mexicana sobre cáncer (1977-2012). *Revista de Aplicación Científica y Técnica* 2015, 1-1: 11-20

\* Correspondencia al Autor (Correo Electrónico: israel\_hm@hotmail.com)

† Investigador contribuyendo como primer autor.

**Introducción****Los Mapas temáticos y el análisis de contenido**

Los mapas temáticos son utilizados para proveer un entendimiento básico de cómo está organizada la ciencia. En este contexto, un tema, área de conocimiento o disciplina se conforma con un subconjunto o racimo (clusters) de términos extraídos de un cuerpo de texto (Sci Tech Strategies, 2012). El mapeo de las ciencias es un área importante de investigación que permite analizar la evolución del conocimiento en determinados campos de dominio (Börner, Chen y Boyack, 2003) (Boyack, Klavans y Börner, 2005).

Estos mapas son particularmente útiles en el análisis y la comparación de los perfiles disciplinarios de un país, que muchas veces es utilizado por agencias de apoyo a la ciencia u otras instituciones, como universidades y centros de investigación para orientar sus programas de desarrollo. Existen dos aspectos relevantes que se pueden distinguir dentro del mapeo de las ciencias: uno, es la construcción de los mapas (denominados bibliométricos en virtud de su origen a partir de registros bibliográficos de bases de datos especializadas) y el otro, es la representación gráfica de tales mapas.

Se ha prestado más atención a la construcción de estos mapas bibliométricos estudiando el efecto de diferentes medidas de similitud, y experimentado con diferentes técnicas de mapeo como las aplicadas por el software Pajek (2012).

La representación gráfica de los mapas bibliométricos, por otro lado, ha recibido una atención mucho menor. Sin embargo, parece que hay una tendencia a considerar mapas más grandes, y para tales mapas, simples representaciones gráficas no son muy adecuadas (Van Eck, Nees, Waltman y Ludo, 2010).

La representación gráfica de grandes mapas bibliométricos puede mejorarse mucho por medio de la funcionalidad del zoom, de algoritmos especiales de etiquetado y metáforas de densidad.

En la década de 1960, Solla Price (1965) sugirió estudiar a la ciencia con los métodos científicos provenientes de la misma. La mayor parte de los primeros trabajos se centraron en la identificación de redes de grupos de autores, documentos o referencias (Garfield, 1995). Los métodos alternativos basados en el análisis de co-palabra se desarrollaron para identificar temas desde un enfoque semántico (Callon y Law, 1983). Por otro lado, los recientes avances en las técnicas de visualización de mapas temáticos han añadido la posibilidad de visualizar diversos campos de dominio del conocimiento (Börner, Chen, y Boyack, 2003).

En bibliometría, el mapeo temático es una herramienta cada vez más importante para la clasificación de documentos en grupos y subgrupos, así como para el análisis de otro tipo de patrones. La materia prima para los mapas bibliométricos han sido las palabras clave, las citas y los términos técnicos que aparecen en los títulos y los resúmenes de los registros de las bases de datos bibliográficas.

El mapeo en forma de redes ha proporcionado información sobre cuestiones tales como las relaciones entre los campos científicos, o las revistas, así como la colaboración científica entre investigadores, las instituciones y los países. Un mapa temático, es un mapa que muestra las relaciones entre palabras o expresiones de un campo de dominio del conocimiento. En estos mapas, entre más cerca se encuentren los términos, más fuerte es la relación entre ellos y se conocen como mapas de co-términos (Peters y Van Raan, 1993).

El análisis de contenido de estos mapas se orienta en una metodología cualitativa cuyo objetivo consiste en identificar el contenido semántico de los conceptos correspondiente a los términos efectivamente utilizados en determinados textos, cuyo análisis se considera interesante para determinada finalidad. Proviene, por una parte del estructuralismo norteamericano, y por otra, de la escuela francesa de análisis del discurso (Magariños, 2010).

El estudio del desenvolvimiento de la investigación sobre cáncer basado en el análisis de contenido (Atkins, 1988) (Zins, 2007), ofrece la oportunidad de identificar los temas de interés por parte de los investigadores en esta importante área del desarrollo científico en México

## Métodos

Se llevó a cabo en Julio de 2013 una búsqueda de información bibliográfica en las bases de datos *Science Citation Index SCI*, *Social Science Citation Index SSCI* y los *Conference Proceedings* tanto de ciencia como de ciencias sociales *CPCI-S*, *CPCI-SSH* respectivamente de Thomson Reuters (2013).

El periodo de búsqueda abarcó de 1977 a 2012 de los trabajos sobre cáncer realizados por al menos un autor con institución de adscripción en México, recuperándose 2688 registros. Los registros se descargaron en formato *plain text*, para a su vez, ser procesado por el software *Bibexcel* (2012) con el fin de generar datos tabulares del campo resumen.

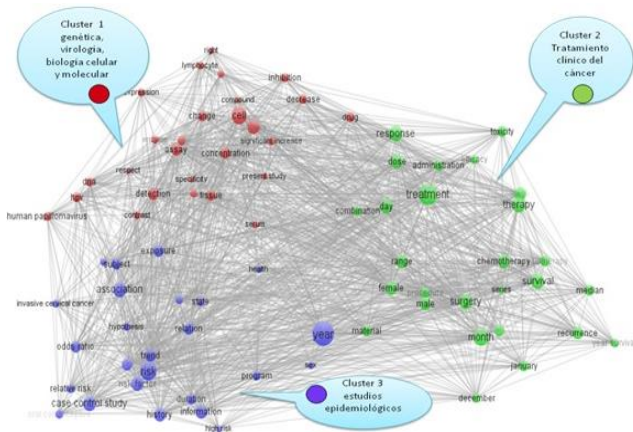
Posteriormente se utilizó el software *VOSviewer* (2012) para generar mapas temáticos a partir del campo resumen. Este software es una aplicación de minería de datos desarrollado en el *Centre for Science and Technology Studies* (CWTS) de Leiden University, The Netherlands, para la construcción y visualización de mapas que se denominan bibliométricos.

Los mapas bibliométricos generados con *VOSviewer* proporcionan una perspectiva dinámica en el ámbito científico, revelando tendencias, como la frecuente atención a un tema, o la convergencia de las corrientes de investigación diferentes. *VOSviewer* identifica secuencias de palabras que se componen exclusivamente de sustantivos y adjetivos y que terminan con un sustantivo (por ejemplo, *paper*, *visualization*, *interesting result*, *and text mining*, *pero no degrees of freedom and highly cited*). El programa dispone de funciones de acercamiento (zoom), desplazamiento (scroll), y búsqueda de términos o expresiones para facilitar el examen detallado de un mapa

## Resultados

### Mapas temáticos generado a partir de un corpus de texto (campo resumen) de los registros bibliográficos de la investigación mexicana sobre cáncer (1977-2012).

Para observar el desenvolvimiento (diversificación y especialización) de los temas de interés asociados a la investigación sobre el cáncer, se generaron mapas temáticos a partir del campo resumen para los siguientes periodos o ventanas en el tiempo: 1977-1999; 2000-2004; 2005-2008; 2009-2012. Por razones de espacio solo se presenta la información de los temas más relevantes extraídos de los mapas de dichos periodos. Al final se presenta un mapa temático de todo el periodo estudiado (1977-2012). En el mapa de la Fig. 1 se observan tres subconjuntos (racimos o clusters) de los términos relevantes obtenidos del corpus de texto del campo resumen de los registros bibliográficos sobre cáncer en el periodo 1977-1999.



**Figura 1** Visualización del mapa temático (generado por VOSviewer) basado en el campo resumen de los registros bibliográficos de la investigación mexicana sobre cáncer (1977-1999).

**(Fig. 1, Cluster 1)** cell, activity, assay, detection, concentration, tissue, change, hpv, decrease, drug, dna, gene, human, papillomavirus, inhibition, lymphocyte, significant increase, sensitivity, expression, enzyme, contrast, serum, human, specificity, present study, right, compound, respect.

**(Fig. 1, Cluster 2)** treatment, month, response, survival, surgery, therapy, dose, day, female, range, chemotherapy, week, male, recurrence, toxicity, January, radiotherapy, mean age, median, material, administration, procedure, december, combination, efficacy, year survival, series

**(Fig. 1, Cluster 3)** year, risk, association, case control study, risk factor, information, history, trend, exposure, relation, confidence interval, odds ratio, subject, duration, program, relative risk, state, oral contraceptive, gastric cancer, hypothesis, invasive cervical, cancer, sex, diet, health, high risk

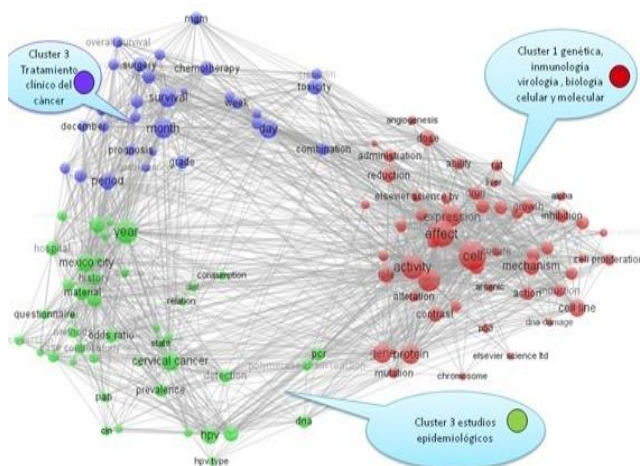
En la Fig. 1, el cluster 1 indica estudios relacionados con la genética, la inmunología, la virología y las investigaciones sobre las bases moleculares del comportamiento de las células cancerígenas. El cluster 2 muestra los aspectos relacionados con el tratamiento clínico de la enfermedad. El cluster 3 presenta por su parte un conjunto de términos que indican estudios epidemiológicos y de investigaciones especializadas donde se aplican los métodos estadísticos.

En el mapa temático de la Fig. 2 se observan tres racimos o clusters de los términos relevantes obtenidos del corpus de texto del campo resumen de los registros bibliográficos referentes a la investigación mexicana sobre cáncer en el periodo 2000-2004.

**(Fig. 2, Cluster 1)** cell, effect, activity, right, expression, protein, gene, cell line, mechanism, dose, role, contrast, concentration, inhibition, receptor, growth, reduction, compound, vitro, drug, action, animal, lymphocyte, proliferation, apoptosis, mouse, administration, induction, cell proliferation, decrease, rat, present study, ability, Elsevier science bv, mutation, human, alteration, beta, regulation, absence, culture, alpha, immunohistochemistry, angiogenesis, vivo, elsevier science ireland ltd, arsenic, cytotoxicity, liver, dna damage, copyright, elsevier science ltd, chromosome, Elsevier science, significant increase.

(Fig. 2, Cluster 2) year, cervical cancer, mexico city, hpv, material, human papillomavirus, confidence interval, risk factor, hospital, pcr, prevalence, dna, history, odds ratio, trend, case control study, methods, paper, polymerase chain reaction, questionnaire, english version, mexican woman, mortality rate, pap, consumption, hpv infection, state, diet, specificity, age group, cin, hpv type, prevention, relation, sex, statistical analysis, cervical intraepithelial, neoplasia, percent, imss, may, high risk, mexican population.

(Fig. 2, cluster 3) month, survival, day, period, toxicity, chemotherapy, surgery, grade, median, , combination, recurrence, mortality, week, radiotherapy, cisplatin, efficacy, december, january, prognosis, morbidity, overall survival, course, mean age, metastasis, head, gastric cancer, multivariate analysis, symptom, most patien.

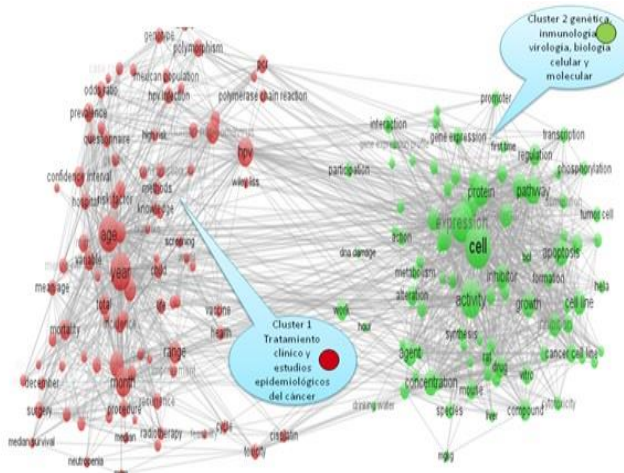


**Figura 2** Visualización del mapa temático (generado por VOSviewer) basado en el campo resumen de los registros bibliográficos de la investigación mexicana sobre cáncer (2000-2004).

En el cluster 1 de la Fig. 2, los términos indican estudios relacionados con la genética, la inmunología, la virología y las investigaciones sobre las bases moleculares del comportamiento de las células cancerígenas. También destacan la identificación específica de genes y la presencia de la editorial científica *Elsevier*. En ese periodo se presenta por primera vez el término apoptosis y el gen *p53* asociado a la misma que indica el inicio de una nueva línea de investigación sobre la muerte celular. Se observa además términos que relacionan la investigación utilizando animales de laboratorio.

En lo referente al cluster 2, el conjunto de términos indican estudios epidemiológicos y de investigaciones donde se aplican los métodos estadísticos. También aparecen las pruebas de laboratorio aplicadas al diagnóstico de la enfermedad. Por su parte, El cluster 3 contiene términos que indican los aspectos relacionados con el tratamiento clínico de la enfermedad, además de estudios donde se aplican los métodos estadísticos. Nótese la aparición del término metástasis y otros que indican nuevos fármacos (ej: cisplatin).

En el mapa temático de la Figura 3 se observan dos racimos o clusters de los términos relevantes obtenidos del corpus de texto del campo resumen de los registros bibliográficos referentes a la investigación mexicana sobre cáncer en el periodo 2005-2008.



**Figura 3** Visualización del mapa temático (generado por VOSviewer) basado en el campo resumen de los registros bibliográficos de la investigación mexicana sobre cáncer (2005-2008).

**(Fig. 3, Cluster 1)** Year, age, hpv, month, diagnosis, human papillomavirus, range, material, country, variable, total, confidence interval, mortality, risk factor, incidence, prevalence, man, polymorphism, pcr, surgery, hpv infection, gastric cancer, procedure, history, odds ratio, child, hospital, recurrence, questionnaire, genotype, mexico city, mean age, impact, toxicity, life, introduction, neoplasm, hpv type, morbidity, complication, methods, management, mexican population, consumption, health, trend, knowledge, vaccine, radiotherapy, december, polymerase chain, reaction, cycle, gender, allele, overall survival, improvement, problem, case control study, cisplatin, vaccination, self, person, pregnancy, safety, participant, quality, significant association, January, high risk, screening, prognostic factor), median survival, median, multivariate analysis, mexican woman, wiley liss, statistical analysis, world, intervention, neutropenia, h pylori, institution, mgm, histology, clinical stage, effectiveness, sex, advanced stage, high rate, feasibility, june.

**(Fig. 3, Cluster 2)** cell, expression, activity, protein, pathway, mechanism, concentration, apoptosis, receptor, activation, agent, growth, inhibitor, proliferation, cell line, contrast, inhibition, regulation, drug, compound, alteration, cell proliferation, work, cancer cell line, mouse, vitro, target, interaction, alpha, cancer cell, animal, gene expression, induction, production, capacity, important role, molecule, promoter, ability, variety, lymphocyte, transcription, tumor cell, formation, action, mrna, hela cell, species, metabolism, structure, cytotoxicity, phosphorylation, synthesis, participation, enzyme, organ, stimulation, rt per, mcf, cytokine, mu m, cell cycle, hela, caspase, human, cervical carcinoma, flow cytometry, breast cancer cell, liver, application, bcl, first time, member, present work, elsevier ireland ltd, western blot, estradiol, estrogen receptor, hypothesis, immune system, angiogenesis, prostate, mgkg, lipid peroxidation, hour, gene expression profile, dna damage, drinking water

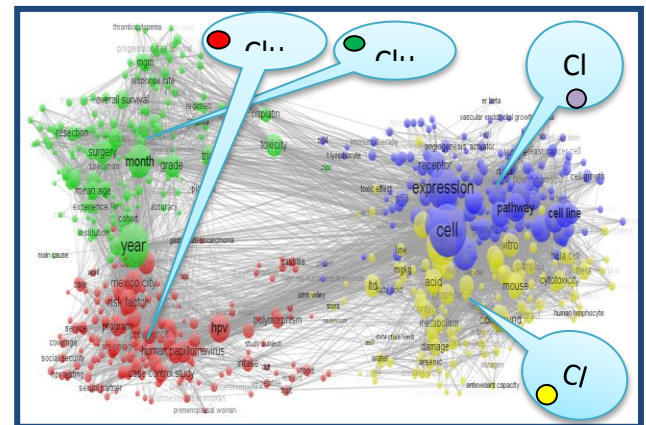
En el cluster 1 de la Figura 3 destacan algunos términos los cuales indican estudios relacionados a la atención clínica de la enfermedad, los estudios de laboratorio, así como las investigaciones estadísticas. Nótese la aparición de los términos relacionados con el cáncer gástrico asociado a la bacteria *H. pylori*. En lo referente al cluster 2 de la Figura 3, el conjunto de términos indican estudios relacionados con la genética, la inmunología, la virología y las investigaciones sobre las bases moleculares del comportamiento de las células cancerígenas. También aparecen términos que indican una amplia diversificación y especialización en estas disciplinas, reflejando también la aplicación de nuevas técnicas de laboratorio y el descubrimiento e identificación de genes y líneas de células.



(Fig. 4, Cluster 2) Year, woman, mexico, age, diagnosis, month, hpv, status, risk factor, country, confidential interval, man, prevalence, total, stage, trial, grade, human papillomavirus, surgery, genotype, program, questionnaire, overall survival, polymorphism, efficacy, week, health, variable, life, history, quality, hospital, mexico city, hpv infection, introduction, center, odds ratio, median, January, need, screening, case control study, mexican woman, cycle, gender, methods, male, susceptibility, trend, safety, multivariate analysis, child, mean age, management, radiotherapy, complication, point, cohort, sex, female, progression free survival, intervention, mexican population, participant, median age, allele, pain, symptom, self, cisplatin, hpv type, institution, mgm obesity, recurrence, statistical analysis, significant association, clinical stage, smoking, series, body mass index, world, survey, december, bmi, education, histology, retrospective study, cross sectional study, tumor size, gastric cancer, ngml, cost, latin America, physician, breast cancer risk, epidemiology, mexican patient, colorectal cancer, access, placebo, progesterone receptor, national cancer, institute, general population, brazil, snp Europe, baseline, early detection, statistical significance, inclusion, performance, diabetes.

En el cluster 1 de la Figura 4 destacan algunos términos los cuales indican estudios relacionados con la genética, la inmunología, la virología y las investigaciones sobre las bases moleculares del comportamiento de las células cancerígenas. También aparecen términos que indican una amplia diversificación y especialización en estas disciplinas, reflejando también la aplicación de nuevas técnicas de laboratorio.

En cluster 2, por otra parte se refleja una amplia diversificación y especialización de los temas relacionados con los estudios epidemiológicos de la enfermedad.



**Figura 5** Mapa temático (generado por VOSviewer) basado en el campo resumen de los registros bibliográficos de la investigación mexicana sobre cáncer (1977-2012).

En la Figura 5 se presenta el mapa disciplinar de los términos obtenidos del corpus de texto del campo resumen de los registros bibliográficos en el periodo 1977-2012 referentes a la investigación mexicana sobre cáncer. La Tabla 1 presenta los términos de mayor relevancia en cada uno de los 4 clusters de la Figura 5.

## Conclusiones

Se considera que el tratamiento dado a la información de los registros bibliográficos de los artículos científicos sobre cáncer con Bibexcel y VOSviewer permite hacer un primer acercamiento en el esfuerzo de entender las tendencias temáticas y la evolución del conocimiento científico en esta área del conocimiento.

Es factible, que con estos mapas, en una etapa posterior, se pueda trabajar con equipos interdisciplinarios para analizar y evaluar con mayor precisión los cambios que permitan detectar la aparición de nuevas disciplinas, su especialización y la diversificación de las mismas, así como tratamientos médicos más eficaces y el nacimiento de nuevos campos de conocimiento en el quehacer de la investigación científica.



Mediante la inspección y el análisis de estos conjuntos de términos los especialistas pueden descubrir las tendencias temáticas de las investigaciones que se muestran estrechamente relacionadas entre sí. Es importante que un mapa, sobre todo de un campo de dominio de la ciencia sea tan preciso como sea posible. Esto se debe a que en muchos casos, estos estudios sirven para respaldar la toma de decisiones dentro del contexto de una planificación nacional de la ciencia y la tecnología. Desde luego, que la utilización de mapas de la ciencia deben considerarse como un elemento más, que puede asociarse con otros procesos para evaluar la actividad científica y la asignación de recursos a la misma.

## Referencias

- Atkins, S.E. (1988) Subject trends in library and Information Science Research, 1975-1984, *Library Trends*, vol. 36 (Spring) pp. 633-658.
- BIBEXCEL [en línea] disponible en <http://www8.umu.se/inforsk/Bibexcel/> [Consultado 23 de marzo 2012]
- Börner, K., Chen, C., Boyack, K. W.(2003) Visualizing knowledge domains, *Annual Review of Information Science and Technology*, vol. 37, pp. 179–255.
- Boyack, Kevin W., Klavans, Richard, Börner, Katy.(2005) Mapping the backbone of science, *Scientometrics*, vol. 64, num. 3, pp. 351-374.
- Callon, M., Law, J.(1983) From translations to problematic networks: an introduction to co-word analysis, *Social Science Information*, vol. 22, pp. 191-235
- Garfield, E.(1995) Citation indexes for science: A new dimension in documentation through association of ideas, *Science*, vol. 122, pp. 108-111.
- Magariños de Morentin, J., Fundamentos para un análisis del discurso vigente en determinada comunidad [en línea] Buenos Aires: Instituto on-line de semiótica Disponible en: <http://www.semiotica-online.com.ar/1Sintesisanálisisoinion.html> [consultado el 5 de marzo de 2010]
- Pajak [en línea] disponible en <http://www8.umu.se/inforsk/Bibexcel/> [Consultado el 23 de marzo 2012]
- Peters, H. P. F., Van Raan, A. F. J.(1993) Co-word-based science maps of chemical engineering, Part I: Representations by direct multidimensional scaling, *Research Policy*, vol. 22, num. 1, pp. 23–45.
- Sci Tech Strategies, consultado en: <http://mapofscience.com> (Consultado el 24 de junio de 2012)
- Solla Price, D. de (1965) Networks of scientific papers, *Science* num. 149, pp.510-515.
- Thomson Reuters. Web of Knowledge, Consultado en: [http://apps.webofknowledge.com.pbidi.unam.mx:8080/RAMore.do?product=WOS&search\\_mode=GeneralSearch&SID=1D5fPnNDF90fhCAm@6F&qid=1224&ra\\_mode=more&ra\\_name=JCRCategories&colName=WOS&viewType=raMore](http://apps.webofknowledge.com.pbidi.unam.mx:8080/RAMore.do?product=WOS&search_mode=GeneralSearch&SID=1D5fPnNDF90fhCAm@6F&qid=1224&ra_mode=more&ra_name=JCRCategories&colName=WOS&viewType=raMore) [Consultado el 1de Julio de 2013]
- Van Eck, Nees Jan, Waltman, Ludo (2010) Software survey: VOSviewer, a computer program for bibliometric mapping, *Scientometrics*, vol 84, pp. 523–538.
- VOSviewer [en línea] Disponible en <http://www.socialsciences.leiden.edu/cwts/products-services/VOSviewer-cwts.html> [Consultado el 15 de octubre de 2012]

ZINS, Chaim (2007) Conceptions in information science, *Journal of the American Society for Information Science and Technology*, vol 58, num. 3, pp. 335-35.