

# StrokeMed: an integrated literature database for stroke and the differentiation of stroke syndrome

Young Uk Kim<sup>1</sup>, Jin Ho Kim<sup>1</sup>, Young Kyu Park<sup>1</sup> and Young Joo Kim<sup>1,\*</sup>

<sup>1</sup> Korea Research Institute of Bioscience and Biotechnology, 52 Eoeun-dong, Yuseong-gu, Daejeon, Korea, 305-333

**Subject areas:** Medical informatics, Bioinformatics/Computational biology/Molecular modeling

**Author contribution:** Kim YU designed and wrote this manuscript.

\***Correspondence** and requests for material should be addressed to Y.J. K. ([ykim8@kribb.re.kr](mailto:ykim8@kribb.re.kr))

**Reviewer:** Keun Woo Lee, Gyeongsang National University, Republic of Korea. Sun Shim Choi, Kangwon National University, Republic of Korea.

**Editor:** Sanghyuk Lee, Ewha Womans University, Republic of Korea

**Received** April 26, 2010;

**Accepted** May 01, 2010;

**Published** May 03, 2010

**Citation:** Kim, Y. U. StrokeMed: an Intergrated Literature database for stroke and the differntiation of stroke syndrome. IBC 2010, 2(2):1, 1-8. doi:10.4051/ibc.2010.2.2.0002

**Competing interests:** All authors declare no financial or personal conflict that could inappropriately bias their experiments or writing.

**Copyright:** This article is licensed under a Creative Commons Attribution License, which freely allows to download, reuse, reprint, modify, distribute, and/or copy articles as long as a proper citation is given to the original authors and sources.

## SYNOPSIS

Complex diseases, such as stroke and cancer, have two or more genetic influences and are affected by environmental factors, which complicate them. Due to the complex characteristics of these diseases, we must search and study comprehensive literature-based article resources. Some disease-related literature databases have been developed through specialized journal issues or major websites. Most of them, however, are scattered throughout a website, and users encounter difficulties in finding accurate and comprehensive information easily and quickly.

We developed StrokeMed, an integrated literature database for stroke and the differentiation of stroke syndrome. The system allows users to explore PubMed search results, categorized by MeSH (Medical Subject Headings), and the differentiation of stroke syndrome in Oriental medicine. StrokeMed collects data from important sites, such as PubMed, Scirus, and Scopus, automatically to maintain higher-quality and updated content. Currently, the system indexes more than 20,000 PubMed abstracts that are related to stroke, stroke etiology, and Oriental medicine. The system provides valuable literature information to the scientific and medical fields in stroke.



## StrokeMed : a literature database for stroke

The screenshot shows the StrokeMed search interface. At the top, there is a search bar with 'All MEDS' selected and 'for Atrial Fibrillation' entered. Below the search bar, there are options for 'Display: Summary', 'Show: 20', 'Items: 54', and 'Page 1 of 3'. The search results are listed as follows:

- 1: Effect of antihypertensive factor on Ca<sup>2+</sup> influx in arterial smooth muscle from normotensive and spontaneously hypertensive rats  
Wan Y.Y.  
Zhongguo yi xue ke x. 1991; 13(2):129-132.  
STKMID:1000020816 Category: Atrial Fibrillation [Google Scholar](#) [Scirus](#) [PubMed](#)
- 2: The structural and functional changes of carotid and brachial arteries in hypertensive stroke patients  
"Chen D., Zhang Z., Chen L."  
Chinese Journal of C. 1994; 22(2):83-86.  
STKMID:1000020744 Category: Atrial Fibrillation [Google Scholar](#) [Scirus](#) [PubMed](#)
- 3: Morphological changes in intracranial and extracranial arteries of autopsy cases of cerebrovascular diseases  
"Liu F., Zhang B., Tian Y."  
Zhonghua yi xue za z. 1996; 76(11):832-835.  
STKMID:1000020710 Category: Atrial Fibrillation [Google Scholar](#) [Scirus](#) [PubMed](#)
- 4: Doppler analysis of vertebral system in patients with sudden sensorineural hearing loss  
"Chao C.-K., Hsu C.-J., Lin K.-N."  
Journal of the Otol. 1997; 32(5):413-417.  
STKMID:1000020703 Category: Atrial Fibrillation [Google Scholar](#) [Scirus](#) [PubMed](#)
- 5: Effect of SNP on KCa of resistance vascular smooth muscle cell membrane in SHRp  
"Zhang Y., Wang G., Bao G., Qu J., Wang X."  
Zhongguo yi xue ke x. 1998; 20(1):8-16.  
STKMID:1000020680 Category: Atrial Fibrillation [Google Scholar](#) [Scirus](#) [PubMed](#)

On the left side of the interface, there is a navigation menu with categories like HOME, Stroke Category, Basic Search, Advanced Search, STATISTICS, LINKS, and CONTACT. Under 'Stroke Category', there is a tree view showing 'Stroke Class', 'Synonyms of Stroke', 'Ischemic Strokes', 'Throbotic Strokes', 'Embolic Strokes', 'Arterial Fibrillation', 'Abnormal Heart Rhytha', 'Heorrhagic Strokes', 'CVD Risks', '변종', and 'MeSH Tree'.

**Keywords:** Stroke, Stroke syndrome differentiation, MeSH, Text Mining, Information extraction, Information retrieval

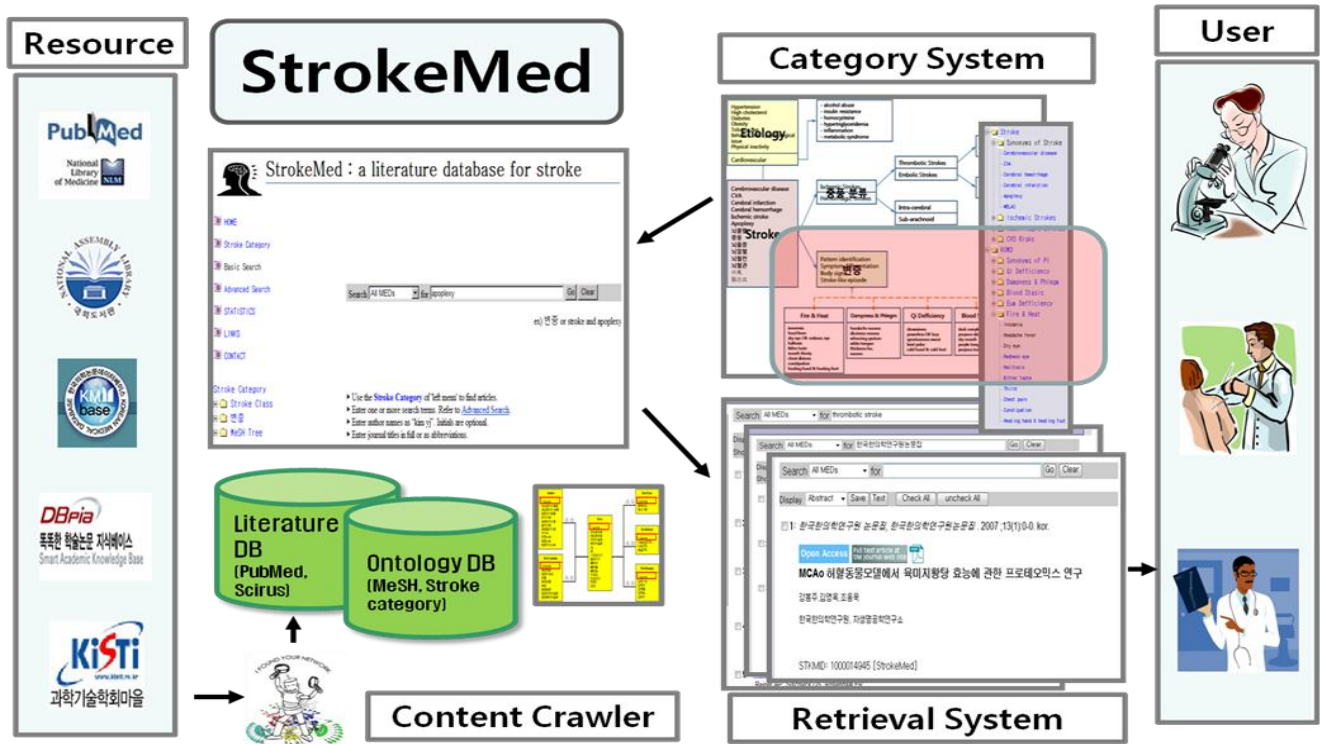


Figure 1. StrokeMed system work flow.

Introduction

We developed the web-based StrokeMed, which is a literature database of Korean, Chinese, and international stroke studies that are focused on the differentiation of Oriental medicine. In the system, a tree-structured GUI was developed for easy and convenient access of literature. The integrated stroke network can serve as a framework of systematic genomic research for stroke, as well as other complex diseases, such as cancers (Kim et al. 2008).

To allow users to explore PubMed search results with respect to MeSH (Medical Subject Headings) and the Oriental medical system, we categorized stroke by MeSH and by stroke syndrome category in Oriental medicine (Rhee et al. 2007; Hearst et al. 2007). Currently, the system indexes more than 20,000 PubMed abstracts that are related to stroke, stroke etiology and Oriental medicine (Nakazato et al. 2009). StrokeMed collects data from important sites, such as PubMed, Scirus and Scopus automatically to maintain higher-quality and updated content (Doms et al. 2005).

Table 1. Category/resource with article numbers.

Category \ Resources	Pubmed	Kmbase	Kisti	Dbpia	Nanet	Scopus	Total
Synonyms of Stroke	3015	354	346	368	362	904	5349
Ischemic Strokes	7054	233	174	186	255	1162	9064
Hemorrhagic Strokes	120	93	85	76	95	1012	1481
CVD Risks	2101	318	283	306	286	813	4107
Synonyms of PI	115	19	23	26	27	0	210
Qi Deficiency	3	19	23	38	35	0	118
Dampness & Phlegm	4	33	39	56	35	0	167
Blood Stasis	1	15	13	11	21	0	61
Eum Deficiency	3	9	2	6	1	0	21
Fire & Heat	2	89	93	86	85	0	355
Total	12418	1182	1081	1159	1202	3891	20933



# StrokeMed : a literature database for stroke

The screenshot shows the StrokeMed search interface. At the top, there is a search bar with 'All MEDs' selected and 'Artrial Fibrillation' entered. Below the search bar, there are navigation links: HOME, Stroke Category, Basic Search, Advanced Search, STATISTICS, LINKS, and CONTACT. On the left, a 'Stroke Category' tree is visible, with 'Artrial Fibrillation' selected under 'Ischemic Strokes'. The main area displays five search results, each with a title, author, journal information, and links to Google Scholar, Scirus, and HubMed. The results are: 1. Effect of antihypertensive factor on Ca2+ influx in arterial smooth muscle from normotensive and spontaneously hypertensive rats; 2. The structural and functional changes of carotid and brachial arteries in hypertensive stroke patients; 3. Morphological changes in intracranial and extracranial arteries of autopsy cases of cerebrovascular diseases; 4. Doppler analysis of vertebral system in patients with sudden sensorineural hearing loss; 5. Effect of SNP on KCa of resistance vascular smooth muscle cell membrane in SHRsp.

**Figure 2.** User interface of StrokeMed. A screenshot of StrokeMed displays the results for 'Artrial Fibrillation,' classified by stroke category. On the left, part of the stroke category that is relevant to the query is shown. On the right, the article is shown for the selected stroke category term.

navigate through abstracts by category quickly (Fig. 2).

## Results

Here, we describe the development of the stroke literature information database (StrokeMed; <http://sysbio.kribb.re.kr:8080/strokeMed/>). StrokeMed consists of three systems: a content crawler, a category system, and a retrieval system (Fig. 1). We developed a crawler to obtain literature information. The crawler is based on the Entrez Programming Utilities (eUtils) web service API, which allows external developers to take advantage of the NCBI databases and processing power to provide alternative representations of the biomedical literature (Yamamoto et al. 2007).

Over 20,000 articles that have been published in the USA, Korea, Chinese, and Japan were collected from databases, including PubMed, SCOPUS, KoreaMed and GoogleScholar (Table. 1). We extracted the literature thoroughly from 15,000 MEDLINE abstracts and 3000 SCOPUS abstracts through a pattern-matching method using regular expression and eUtils (Alfred et al. 2006). We classified the literature into several categories using MeSH terms and constructed the search system based on this categorical system (Table 2). The Korean versions of Tables 1 and 2 are supplied as Supplementary materials.

StrokeMed provides a dynamic and intuitive interface that integrates it with data from other sources to improve researchers' ability to find and manage the biomedical literature that is related to their field. It incorporates external web services (Google Scholar, Scirus, and HubMed) to explore connections between related papers. It gives an overview of the literature abstracts by categorizing them according to stroke category, allowing users to

## Discussion

StrokeMed, a comprehensive literature database system for stroke, was constructed based on refined and systemically categorized stroke literature and provides stroke-related data to researchers through an efficient embedded retrieval system. For future development, StrokeMed will continue to gather articles that are related to StrokeMed and the differentiation of stroke syndrome and improve search efficiency and user-friendliness (Hur et al. 2009). Personalization of searches and recommendations, based on the patterns of user attention and implied interests, might also improve the accuracy of the search results (Tang et al. 2009). The Stroke Category for semantic stroke data management will continue to adapt to new developments and the needs of researchers in this area.

## Acknowledgment

This work was supported by a grant from the Korea Institute of Oriental Medicine (KIOM, K09200), in part by a grant (KGS2210911) from the KRIBB Research Initiative Program, and in part by a grant from the Korea Research Council of Fundamental Science & Technology (NTM1300711) of the Republic of Korea.

**Table 2.** Stroke category classification of stroke and differentiation of stroke syndrome.

Stroke	WHO standard term
Synonyms of Stroke	Synonyms of PI
Cerebrovascular disease	Pattern Identification
CVA	Symptom Differentiation
Cerebral hemorrhage	Body Sign
Cerebral infarction	Stroke-like Episode
Apoplexy	
MELAS	
Ischemic Strokes	Qi Deficiency
Thrombotic Strokes	Drowsiness
Atherosclerosis	Powerless OR lazy
TIA	Spontaneous sweat
Embolic Strokes	Faint pulse
Arterial Fibrillation	Cold hand & cold foot
Abnormal Heart Rhythm	
Hemorrhagic Strokes	Dampness & Phlegm
Intra-cerebral Hemorrhage	Headache nausea
Sub-arachnoid Hemorrhage	Dizziness nausea
	Wheezing sputum
	White tongue
	Thickness fur
	Nausea
	Blood Stasis
	Dark complexion
	Purpura skin
	Dry mouth
	Purple tongue
	Purpura tongue
	Eum Defficiency
	Mouth thrush
	Palpitation
	Night sweat
	Fire & Heat
	Insomnia
	Headache fever
	Dry eye
	Redness eye
	Halitosis
	Bitter taste
	Thirst
	Chest pain
	Constipation
	Heating hand & heating foot

**References**

Alfred, E. (2006). HubMed: a web-based biomedical literature search interface. *Nucleic Acids Res* 34,745-747.

Doms, A., and Schroeder, M. (2005). GoPubMed: exploring PubMed with the Gene Ontology. *Nucleic Acids Res* 7,783–786.

Hearst, M.A., Divoli, A., Guturu, H., Ksikes, A., Nakov, P., Wooldridge, M.A., and Ye, J. (2007). BioText Search Engine: beyond abstract search. *Bioinformatics* 23,2196-2197.

Hur, J., Schuyler, A.D., States, D.J., and Feldman, E.L. (2009). SciMiner: web-based literature mining tool for target identification and functional enrichment analysis. *Bioinformatics* 25,838-840.

Kim, Y.U., Kim, I.H., Bang, O.S., and Kim, Y.J. (2008). StrokeBase: A Database of Cerebrovascular Disease-related Candidate Gene. *Genomics & informatics* 6, 153-156.

Nakazato, T., Bono, H., Matsuda, H., and Takagi T. (2009). Gendoo: functional profiling of gene and disease features using

MeSH vocabulary. *Nucleic Acids Res* 37,W166-W169.

Rhee, H., and Lee, J.S., (2007). PADB: published association database. *BMC Bioinformatics* 9, 348-355.

Tang, S., Zhang, Z., Kavitha, G., Tan, E.K., and Ng, S.K. (2009). MDPD: an integrated genetic information resource for Parkinson's disease. *Nucleic Acids Res* 1,858–862.

Yamamoto, Y., and Takagi, T. (2007). OReFiL: an online resource finder for life sciences. *BMC Bioinformatics* 8, 287-294.

Entrez Programming Utilities. [http://eutils.ncbi.nlm.nih.gov/corehtml/query/static/eutils\\_help.html](http://eutils.ncbi.nlm.nih.gov/corehtml/query/static/eutils_help.html)

KoreaMed. <http://koreamed.org/SearchBasic.php>

MeSH. <http://www.nlm.nih.gov/mesh/meshhome.html>

PubMed. <http://www.pubmed.gov/>