

รายการอ้างอิง

ภาษาไทย:

- กฤษณี อริยชาญศิลป์. ระบบค้นคืนสารสนเทศภาษาไทย-อังกฤษ สำหรับคำทับศัพท์และแสดงผล
ลัพท์ด้วยวิธีจัดกลุ่มข้อมูล. วิทยานิพนธ์ปริญญาตรีบัณฑิต ภาควิชาวิศวกรรม
คอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย, 2545.
- กัลยา วานิชย์บัญชา. การวิเคราะห์สถิติ: สถิติสำหรับการบริหารและวิจัย. กรุงเทพมหานคร:
สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย, 2546.
- กัลยา วานิชย์บัญชา. การวิเคราะห์สถิติขั้นสูงด้วย SPSS for Windows. กรุงเทพมหานคร:
สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย, 2548.
- กิตติ ภักดีวัฒน์กุล อังศุมาลิน เวชนารายณ์ และกิตติพงษ์ ธีรวัฒน์เสถียร. PHP ฉบับ
โปรแกรมเมอร์. กรุงเทพมหานคร: KTP COMP & CONSULT, 2545.
- จินตนา บุญบงการ. สภาพแวดล้อมทางธุรกิจ. กรุงเทพมหานคร: สำนักพิมพ์แห่งจุฬาลงกรณ์
มหาวิทยาลัย, 2550.
- ชูชาติ หฤไชยะศักดิ์. "Information Retrieval and Search Engine." [ออนไลน์]. แหล่งที่มา:
<http://www.tcllab.org/events/uploads/IR.pdf> 2548 [7 พฤษภาคม 2550].
- ชูชาติ หฤไชยะศักดิ์. "Text Mining: Techniques and Applications." [ออนไลน์]. แหล่งที่มา:
<http://203.144.225.124/add/add2/lectures/choochart-TextMining.pdf> 2550
[20 พฤษภาคม 2550].
- ธีรภัทร มนตรีศาสตร์ "Apache และการติดตั้ง." [ออนไลน์]. แหล่งที่มา:
<http://www.itdestination.com/articles/idc-apache/> 2547 [2 พฤศจิกายน 2550].
- นิพนธ์ เจริญกิจการ. การจัดเก็บและค้นคืนสารสนเทศ. รังสิตสารสนเทศ 4, 2 (2541): 8-23.
- พิลาวัฒน์ พลับรู้อการ และกฤษณะ ไวยมัย. "การวัดความคล้ายคลึงระหว่างเอกสารโดยใช้แนวทาง
ด้านความหมาย." [ออนไลน์]. แหล่งที่มา:
https://pindex.ku.ac.th/file_research/paper_kufair41.doc 2546 [7 พฤษภาคม
2550].
- พิลาวัฒน์ พลับรู้อการ และกฤษณะ ไวยมัย. การปรับปรุงการแบ่งเอกสารโดยใช้ฐานความรู้เวิร์ด
เน็ต. ใน The 8th National Computer Science and Engineering Conference, 380-
388. Songkla: Songkla University, 2547.

ลัดดาวัลย์ เพชรโรจน์ และอัจฉรา ชำนิประศาสน์. ระเบียบวิธีวิจัย (Research Methodology).

กรุงเทพมหานคร: พิมพ์ดีการพิมพ์ จำกัด, 2547.

วิกิพีเดีย สารานุกรมเสรี. "การแบ่งกลุ่มข้อมูล." [ออนไลน์]. แหล่งที่มา:

<http://th.wikipedia.org/wiki/การแบ่งกลุ่มข้อมูล> 2550 [15 สิงหาคม 2550].

วิกิพีเดีย สารานุกรมเสรี. "ภาษาซี" [ออนไลน์]. แหล่งที่มา: <http://th.wikipedia.org/wiki/ภาษาซี> 2551 [18 มีนาคม 2551].

วันเพ็ญ ชูเมือง. "Information Retrieval on the WWW." [ออนไลน์]. แหล่งที่มา:

<http://www.cs.su.ac.th/~sirak/517632/pen/PRESENTir.doc> 2547 [11 มิถุนายน 2550].

ศิริรัตน์ ศิรนานนท์. การค้นคืนสารสนเทศโดยใช้กฎความสัมพันธ์ร่วมกลับผลสะท้อนกลับจากผู้ใช้.

วิทยานิพนธ์ปริญญาโทมหาบัณฑิต ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย, 2549.

สุชาติ กิระนันท์, คุณหญิง. คำถามคำตอบเกี่ยวกับการทำเหมืองข้อมูลเบื้องต้น. ใน สัมมนา
หนึ่งทศวรรษไอทีจุฬาฯ: การเพิ่มศักยภาพระบบการจัดการด้วยเทคโนโลยีสารสนเทศ,
1-10. [ม.ป.ท.], 2545.

สมพร จิวรสกุล. คู่มือการติดตั้งและใช้งาน Microsoft SQL Server 2000 ฉบับสมบูรณ์.

กรุงเทพมหานคร: สำนักพิมพ์ อินโฟเพรส, 2545.

ภาษาอังกฤษ:

Baeza-Yates, Ricardo, and Ribeiro-Neto, Berthier. Modern Information Retrieval.

New York: A Division of the Association for Computing Machinery (ACM), 1999.

Bass, L., Clements, P., and Kazman, R. Software Architecture in Practice. [n.p.]:

Addison Wesley. 1998.

C++Builder. "C++Builder home page" [Online]. Available from:

<http://www.borland.com/bcppbuilder/> [February 21, 2008].

Chowdhury, G. Introduction to Modern Information Retrieval. second edition. London:

Library Association Publishing, 2004.

Dimitrios, Z. and Gallopoulos, E. "TMG : A MATLAB Toolbox for generating term-document matrices from text collections." [Online]. Available from:

- <http://scgroup.hpclab.ceid.upatras.gr/scgroup/Projects/TMG/>. [October 24, 2007].
- Han ,Eui-Hong; Karypis, George; and Kumar, Vipin. Text Categorization Using Weight Adjusted k-Nearest Neighbor Classification. In Proceedings of the 5th Pacific-Asia Conference on Knowledge Discovery and Data Mining Lecture Notes In Computer Science, 53-65. London: Springer-Verleg, 2001.
- Goutte, Cyril, and Gaussier, Eric. A Probabilistic Interpretation of Precision, Recall and F-score, with Implication for Evaluation. In 27th European Conference on Information Retrieval (ECIR'05), 345-359. [n.p.]: LLNCS 3408(Springer), 2005.
- Greenwood, Mark A. "Implement a Vector Space Document Retrieval System." [Online]. Available from: <http://www.dcs.shef.ac.uk/~mark/> 2002. [April 20, 2007].
- Hand, D.J. et al.. Principles of Data Mining. Cambridge: The MIT Press, 2001.
- Hoch Riner. Using IR Techniques for Text Classification in Document Analysis. In Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval, 31-40. New York: Springer-Verleg, 1994.
- Karypis, George; and Kumar, Vipin. A Comparison of Document Clustering Techniques. In KDD Workshop on Text Mining. [n.p.], 2001.
- "K_means Clustering Algorithm" [Online]. Available from: <http://www.andrew.cmu.edu/user/smisra/Kmean.htm>. 2006. [March 12, 2008].
- Klabbankoh, Bangorn. Applied Genetic Algorithms in Information Retrieval. The Internet and Management 7, 3(1999): 30-35.
- Leuski, Anton. Evaluating Document Clustering for Interactive Information Retrieval. In Proceedings of the tenth international conference on Information and knowledge management CIKM '01, 33-40. Massachusetts: ACM, 2002.
- Porter, M. "The Porter Stemming Algorithm." [Online]. Available from: <http://www.tartarus.org/~martin/PorterStemmer/index.html>. 1980. [May 20, 2007].

- Qian, Gang et al.. Similarity between Euclidean and cosine angle distance for nearest neighbor queries. In Proceedings of the 2004 ACM symposium on Applied computing SAC, 1232–1237. Nicosia: ACM, 2004.
- Moody, Daniel. Empirical Research. In Research Methods class. [n.p.], 2002.
- Moulanont, Saroch. Information retrieval using vector space model concept. A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science Program in Computer Science Department of Mathematics. Faculty of Science. Chulalongkorn University, 1992.
- Salton, Gerard, and McGill, Michael J. Introduction to Modern information Retrieval. Auckland:[n.p.], 1987.
- Smart System. "Stop word list." [Online]. Available from:
<http://www.unine.ch/info/clef/englishST.txt>. 2005. [May 20, 2007].
- Udomchaiporn, Akadej. Use Case Retrieval using Terms and Use Case Structure Similarity Computation. A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science Program in Software Engineering Department of Computer Engineering. Faculty of Engineering. Chulalongkorn University. 2005.
- Weiss, Sholom M. et al.. Text Mining: Predictive Methods for Analyzing Unstructured Information. New York: Springer Science+Business Media, 2005.
- Zhang, Roger. "kmeans.c" [Online]. Available from:
http://cs.smu.ca/~r_zhang/code/kmeans.c. 2005. [February 19, 2008].

ภาคผนวก

ภาคผนวก ก

ตัวอย่างเอกสารและข้อสอบถาม

งานวิจัยนี้กำหนดให้ฐานข้อมูลนิตยสารไทม์ (TIME Collection) ปี 1963 มาทดสอบระบบค้นคืนเอกสารที่พัฒนาขึ้น ซึ่งประกอบด้วยเอกสารจำนวน 425 เอกสาร และข้อสอบถามจำนวน 83 ข้อสอบถาม ฐานข้อมูลนิตยสารไทม์ได้กำหนดข้อสอบถามไว้สำหรับทดสอบระบบค้นคืนเอกสาร (Smart Collection, 1963) จากข้อสอบถามทั้ง 83 ข้อสอบถาม โดยเอกสารจากฐานข้อมูลนิตยสารไทม์เป็นเอกสารที่เกี่ยวกับข่าวสารทั่วไป ฐานข้อมูลนิตยสารไทม์เป็นฐานข้อมูลมาตรฐานที่สร้างโดยมหาวิทยาลัยคอเนล (Cornell University) และเป็นฐานข้อมูลที่ใช้ในการทดสอบงานวิจัยทางการค้นคืนเอกสารมากมาย (Chowdhury 2004)

ตัวอย่างเอกสาร

*TEXT 017 01/04/63 PAGE 020

THE ALLIES AFTER NASSAU IN DECEMBER 1960, THE U.S. FIRST PROPOSED TO HELP NATO DEVELOP ITS OWN NUCLEAR STRIKE FORCE. BUT EUROPE MADE NO ATTEMPT TO DEVISE A PLAN. LAST WEEK, AS THEY STUDIED THE NASSAU ACCORD BETWEEN PRESIDENT KENNEDY AND PRIME MINISTER MACMILLAN, EUROPEANS SAW EMERGING THE FIRST OUTLINES OF THE NUCLEAR NATO THAT THE U.S. WANTS AND WILL SUPPORT. IT ALL SPRANG FROM THE ANGLO-U.S. CRISIS OVER CANCELLATION OF THE BUG-RIDDEN SKYBOLT MISSILE, AND THE U.S. OFFER TO SUPPLY BRITAIN AND FRANCE WITH THE PROVED POLARIS (TIME, DEC. 28). THE ONE ALLIED LEADER WHO UNRESERVEDLY WELCOMED THE POLARIS OFFER WAS HAROLD MACMILLAN, WHO BY THUS KEEPING A SEPARATE NUCLEAR DETERRENT FOR BRITAIN HAD SAVED HIS OWN NECK. BACK FROM NASSAU, THE PRIME MINISTER BEAMED THAT BRITAIN NOW HAD A WEAPON THAT "WILL LAST A GENERATION. THE TERMS ARE VERY GOOD." MANY OTHER BRITONS WERE NOT SO SURE. THOUGH THE GOVERNMENT WILL SHOULDERS NONE OF THE \$800 MILLION DEVELOPMENT COST OF POLARIS, IT HAS ALREADY POURED \$28 MILLION INTO SKYBOLT AND WILL HAVE TO SPEND PERHAPS \$1 BILLION MORE FOR A FLEET OF MISSILE-PACKING SUBMARINES. AT BEST, THE BRITISH WILL NOT BE ABLE TO DESIGN, BUILD AND PROVE ITS NUCLEAR FLEET BEFORE 1970, THREE YEARS AFTER BRITAIN'S BOMBER FORCE HAS PRESUMABLY BECOME OBSOLETE. THEN WHAT? TORY BACKBENCHERS ARE LOUDLY SKEPTICAL OF WHAT THEY CALL "THE SMALL TYPE" IN THE NASSAU PACT, WHICH STIPULATES THAT BRITAIN'S POLARIS SUBMARINE FLEET, EXCEPT WHEN "SUPREME NATIONAL INTERESTS" INTERVENE, MUST BE COMMITTED TO A TRULY MULTILATERAL NATO FORCE. DOES THAT MEAN THAT BRITAIN WILL EVENTUALLY HAVE NO STRIKE FORCE OF ITS OWN? WHO WILL DECIDE WHEN OR WHETHER NATIONAL INTERESTS JUSTIFY WITHDRAWAL OF SUBMARINES FROM NATO, PARTICULARLY IF THOSE NATIONAL INTERESTS CONFLICT WITH U.S. POLICY? THE BIGGEST QUESTION OF ALL IS WHETHER FRANCE'S INCLUSION IN THE OFFER WAS A DELIBERATE PLOY BY JACK KENNEDY TO END OR AT LEAST DOWNGRADE BRITAIN'S PRIZED "SPECIAL RELATIONSHIP" WITH THE U.S. THE CARTOONISTS WENT EVEN

FARTHER . THEY NOT ONLY SHOWED SUPERMAC JUMPING TO SUPERJACK'S COMMANDS, BUT DE GAULLE AND ADENAUER AS WELL . AS EDITH SAID . THE FRENCH, WHO GOT NO HELP FROM THE U.S . IN DEVELOPING THEIR FORCE DE FRAPPE, WERE QUICK TO CROW THAT BRITAIN'S VAUNTED TIES WITH THE U.S . HAD BROUGHT IT NOTHING BUT HUMILIATION . BY CONTRAST, BRAGGED FRENCH OFFICIALS, THE SKYBOLT FIASCO ONLY VINDICATED FRANCE'S DECISION TO DEVELOP ITS OWN BOMBS AND DELIVERY SYSTEMS . THUS, THOUGH CHARLES DE GAULLE PROMISED TO " REFLECT " ON THE POLARIS OFFER, THERE WAS LITTLE LIKELIHOOD THAT HE WOULD ACCEPT ANY OFFER THAT WOULD SUBJECT A FRENCH FORCE TO ALLIED CONTROL . IT IS DE GAULLE'S UNSWERVING CONVICTION THAT IF THE RUSSIANS WERE ACTUALLY TO INVADE WESTERN EUROPE, NO NATION THAT WAS NOT DIRECTLY ATTACKED MEANING THE U.S . WOULD INVITE NUCLEAR DEVASTATION BY HELPING ITS ALLIES . THUS UNLIKE BRITAIN'S BOMBER FORCE, WHICH ALL ALONG HAS BEEN PLEDGED TO "THE WESTERN STRATEGIC DETERRENT, " FRANCE'S FORCE DE FRAPPE WILL BE RESPONSIBLE ONLY FOR FRANCE'S DEFENSE . AT THE SAME TIME, DE GAULLE HAS LONG ARGUED THAT THE ATLANTIC ALLIANCE COULD BE RUN MOST EFFICIENTLY BY A TRIUMVIRATE THAT WOULD INCLUDE FRANCE AS AN EQUAL OF THE U.S . AND BRITAIN . THIS IS ONE OF HIS MAJOR, IF UNSPOKEN, CONDITIONS FOR BRITISH MEMBERSHIP IN THE COMMON MARKET ; AND DE GAULLE SUGGESTED POINTEDLY TO MACMILLAN THAT IT WOULD HELP IF BRITAIN WERE TO SHARE ITS ADVANCED MISSILE TECHNOLOGY WITH FRANCE . WHEN MACMILLAN REPLIED NONCOMMITTALLY THAT HE WOULD HAVE TO DISCUSS THIS WITH KENNEDY, DE GAULLE TOLD HIS GUEST WITH HAUTEUR THAT FRANCE IN THAT CASE COULD DO NOTHING TO EASE BRITAIN'S ENTRY INTO EUROPE . GO-IT-ALONE GRANDEUR . KONRAD ADENAUER, ON THE OTHER HAND, IS FEARFUL THAT DE GAULLE WILL SNAP UP THE POLARIS OFFER AND IN THIS WAY ACHIEVE HIS GOAL OF A THREE-NATION NATO DIRECTORATE . THOUGH HIS GOVERNMENT VOWED IN 1954 NOT TO MANUFACTURE NUCLEAR WEAPONS, ADENAUER HAS BECOME INCREASINGLY APPREHENSIVE THAT WITHOUT THEM, AND WITH NO SAY IN THEIR USE, WEST GERMANY WILL BE RELEGATED TO SECOND-CLASS CITIZENSHIP IN THE ALLIANCE . LAST WEEK AN OFFICIAL BULLETIN EVEN REVIVED THE OLD, BITTER CRY THAT U.S . PLEAS FOR GREATER RELIANCE ON CONVENTIONAL FORCES ARE AIMED AT RAISING GERMAN "CANNON FODDER " FOR U.S . " ATOMIC KNIGHTS . " A FROSTY LETTER FROM THE CHANCELLOR TO PRESIDENT KENNEDY SUGGESTED THAT GERMANY, WHICH ALREADY SUPPLIES ALMOST 50 PER CENT OF NATO GROUND STRENGTH, DOES NOT INTEND TO RAISE ANY MORE DIVISIONS FOR CONVENTIONAL WARFARE . YET U.S . STRATEGIC PLANNERS REASON THAT THE ONLY CREDIBLE DETERRENT TO SOVIET ATTACK IS A STRONG ARMY ON THE GROUND, BACKED BY THE VAST U.S . NUCLEAR ARSENAL . FACT IS, THE BRITISH AND FRENCH NUCLEAR WEAPONS COULD NEVER BE USED INDEPENDENTLY OF THE U.S . AGAINST RUSSIA WITHOUT INVITING DEVASTATING SOVIET RETALIATION . AFTER ALL THEIR EFFORTS, THE BRITISH AND FRENCH WILL HAVE MANAGED TO CREATE A NUCLEAR CAPACITY THAT REPRESENTS ONLY 4 PER CENT OF U.S . NUCLEAR POWER . " IT IS JUST A DAMNED NUISANCE, " SAID A STATE DEPARTMENT OFFICIAL LAST WEEK . " IT MEANS NOTHING MILITARILY EXCEPT THAT WE WILL BE EXPECTED TO BAIL OUT THE FIRST COUNTRY THAT THROWS THE FIRST PEA AT THE RUSSIANS OR ANYONE ELSE . " CHARLES DE GAULLE COULD HARDLY BE EXPECTED TO AGREE, AT LEAST UNTIL HIS FORCE DE FRAPPE BECOMES OBSOLETE . FOR BRITAIN AND GERMANY, THE

MULTILATERAL DETERRENT MAKES IMMEDIATE SENSE . EVENTUALLY, FRANCE, TOO, MAY WELL FIND A NATO-CONTROLLED POLARIS FLEET, OR ITS POSSIBLE SUCCESSOR, A EUROPEAN MINUTEMAN ARSENAL, THE ONLY ANSWER TO THE SPIRALING COST AND DIMINISHING VALUE OF GO-IT-ALONE GRANDEUR .

*TEXT 018 01/04/63 PAGE 021

RUSSIA WHO'S IN CHARGE HERE ? IT WAS IN 1954 THAT NIKITA KHRUSHCHEV LAUNCHED HIS GRANDIOSE " VIRGIN LANDS " GAMBLE . PART OF THE PLAN WAS TO PLOW UP 32 MILLION ACRES OF MARGINAL LAND IN KAZAKHSTAN, AND SETTLE IT WITH COMMUNIST " PIONEERS, " WHO WERE TO PLANT AND PRODUCE HUGE QUANTITIES OF DESPERATELY NEEDED GRAIN WITHIN TWO YEARS. NIKITA'S SCHEME FLOPPED . THERE WAS NOT ENOUGH RAINFALL, AND THE PIONEERS DID NOT TAKE TO TRACTOR LIFE ON THE BLEAK FRONTIER . EXCEPT FOR 1958, EACH HARVEST HAS BEEN LOWER THAN THE PREVIOUS YEAR'S. WORST YEAR OF ALL WAS 1962, WHEN THE VIRGIN LANDS DELIVERED ONLY HALF THEIR QUOTAS . NATURALLY, KHRUSHCHEV TAKES NONE OF THE BLAME FOR THE FIASCO . THREE YEARS AGO HE FOUND A SCAPEGOAT IN KAZAKHSTAN PARTY BOSS NIKOLAI BELYAEV, FIRED HIM FOR HIS " ERRORS . " LAST WEEK BELYAEV'S SUCCESSOR, DINMUKHAMED KUNAEV, WAS SIMILARLY BOUNCED FOR " LAPSES " IN HIS WORK . FOR GOOD MEASURE, MOSCOW ALSO PURGED THE FORMER PREMIER OF THE TERRITORY FROM THE LOCAL PARTY'S CENTRAL COMMITTEE . IT WAS PERHAPS NO COINCIDENCE THAT NIKOLAI IGNATOV, 61, A ONETIME KHRUSHCHEV CRONY, LAST WEEK ABRUPTLY LEFT HIS POST AS A SOVIET DEPUTY PREMIER AFTER ONLY NINE MONTHS ON THE JOB . FARM EXPERT IGNATOV HAD THE MISFORTUNE TO BE BOSS OF A SPECIAL COMMITTEE TO BOOST FOOD PRODUCTION .

*TEXT 019 01/04/63 PAGE 021

BERLIN ONE LAST RUN HANS WEIDNER HAD BEEN HOPING FOR MONTHS TO ESCAPE DRAB EAST GERMANY AND MAKE HIS WAY TO THE WEST . THE ODDS WERE AGAINST HIM, FOR WEIDNER, 40, WAS A CRIPPLE ON CRUTCHES WHO LIVED IN THE VILLAGE OF NEUGERSDORF, 115 MILES SOUTHEAST OF THE FRONTIER OF FREEDOM BUT HANS WEIDNER DID HAVE ONE MAJOR ASSET, THE BUS THAT HE OPERATED FOR THE LOCAL COMMUNIST REGIME . IT WAS AN UGLY THING, AND ANCIENT . ITS CHASSIS CREAKED, AND THE ENGINE COUGHED ; A CREAM-COLORED COAT OF PAINT COULD NOT DISGUISE THE WELTS AND BRUISES OF TWO DECADES OF CHUGGING SERVICE . IN FACT, THE BUS WAS READY FOR THE JUNK PILE WHEN WEIDNER DECIDED TO PRESS IT INTO SERVICE FOR ONE LAST RUN . SHARP BLADES . THE HAZARDS WOULD BE GREAT ON THE JOURNEY TO THE BORDER ; SO WEIDNER SIGNED UP A FELLOW VILLAGER, JURGEN WAGNER, 22, TO TAKE THE WHEEL . EIGHT DAYS BEFORE CHRISTMAS, THE PAIR BEGAN THE FEVERISH PREPARATIONS IN WEIDNER'S GARAGE . FIRST WEIDNER AND WAGNER ATTACHED A HEAVY SNOWPLOW TO THE FRONT OF THE BUS, NOT TO PLOW SNOW, BUT TO SCOOP AWAY THE HEAVY OBSTACLES THEY KNEW AWAITED THEM AT ROADBLOCKS AHEAD . TO ALL SIX LUGS ON EACH FRONT WHEEL THEY BOLTED SHARP BLADES OF THE TOUGHEST STEEL, AFFIXED SO THAT THE WHIRLING EDGES WOULD CHOP BARBED WIRE TO BITS . THEN THEY WEDGED ONE-QUARTER-INCH SECTIONS OF STEEL PLATE INSIDE THE BUS TO STOP BULLETS . AT LAST ALL WAS READY . ON CHRISTMAS EVE, WEIDNER AND WAGNER PILED THEIR WIVES AND FOUR CHILDREN ABOARD, NOT FORGETTING THREE TONS OF HOUSEHOLD BELONGINGS. FOR ADDED PROTECTION THE PLOTTERS SHOVELED A TON OF COAL AND POTATOES INTO THE BACK OF THE BUS . THEN THEY CHUGGED OFF NORTH TOWARD BERLIN ALONG BACK ROADS TO ESCAPE COMMUNIST PATROLS. JUST BEFORE THEY

REACHED THE WALL, THEY PLANNED TO SWING WEST IN ORDER TO ENTER THE EAST-WEST AUTOBAHN LEADING TO THE U.S. SECTOR OF THE CITY. EN ROUTE, THE RADIATOR FROZE IN THE SUBZERO WEATHER. THAT FIXED, THEY WERE ONLY A FEW MILES FARTHER WHEN A TIRE BLEW OUT. THE KIDS WERE CRYING AND THE WIVES SHIVERING WITH COLD AND PANIC WHEN, AT LAST, THEY ARRIVED AT DREWITZ, THE MOST HEAVILY GUARDED CHECKPOINT ON THE ENTIRE AUTOBAHN TO BERLIN. IT WAS NO TIME TO STOP AND RECONSIDER. FLYING POTATOES. "WAH-AH, WAH-AH," SHRIEKED THE POLICE-TYPE KLAXONS THAT WEIDNER HAD THOUGHTFULLY INSTALLED IN ADVANCE. THE COMMUNIST GUARDS OBEDIENTLY RAISED THE FIRST OF THREE BARRIERS. BUT WHAT WAS A BUS DOING ON EMERGENCY DUTY? SUDDENLY THE SHOOTING BEGAN TOO LATE. WAGNER, AT 40 M.P.H., WAS ALREADY CRASHING THROUGH THE SECOND BARRIER 100 YARDS AHEAD, THEN THE THIRD, ONLY 20 YARDS AWAY. ITS WINDSHIELD SMASHED, ITS PASSENGERS SHAKEN, ITS CARGO OF COAL AND POTATOES IN EVERY CORNER OF THE CAB, THE OLD BUS FINALLY LURCHED TO A STOP A FEW MILES DOWN THE ROAD WHERE THE COMMUNISTS NO LONGER MATTERED AT THE U.S. CHECKPOINT, A FOOT OR TWO INSIDE WEST BERLIN.

ตัวอย่างข้อสอบถาม

*FIND 1

KENNEDY ADMINISTRATION PRESSURE ON NGO DINH DIEM TO STOP SUPPRESSING THE BUDDHISTS.

*FIND 2

EFFORTS OF AMBASSADOR HENRY CABOT LODGE TO GET VIET NAM'S PRESIDENT DIEM TO CHANGE HIS POLICIES OF POLITICAL REPRESSION.

*FIND 3

NUMBER OF TROOPS THE UNITED STATES HAS STATIONED IN SOUTH VIET NAM AS COMPARED WITH THE NUMBER OF TROOPS IT HAS STATIONED IN WEST GERMANY.

*FIND 4

U.S. POLICY TOWARD THE NEW REGIME IN SOUTH VIET NAM WHICH OVERTHREW PRESIDENT DIEM.

*FIND 5

PERSONS INVOLVED IN THE VIET NAM COUP.

*FIND 6

CEREMONIAL SUICIDES COMMITTED BY SOME BUDDHIST MONKS IN SOUTH VIET NAM AND WHAT THEY ARE SEEKING TO GAIN BY SUCH ACTS.

*FIND 7

REJECTION BY PRINCE NORODOM SIHANOUK, AN ASIAN NEUTRALIST LEADER, OF ALL FURTHER U.S. AID TO HIS NATION.

*FIND 8

U.N. TEAM SURVEY OF PUBLIC OPINION IN NORTH BORNEO AND SARAWAK ON THE QUESTION OF JOINING THE FEDERATION OF MALAYSIA.

ภาคผนวก ข
รายการคำยกเว้น (Stop words list)

รายการคำที่เป็นคำยกเว้นที่ใช้ในงานวิจัยนี้ นำมาจากระบบค้นคืนเอกสารสมาร์ท (SMART) ที่พัฒนาโดยมหาวิทยาลัยคอแนล (Cornell University) (SMART System, 2005)
รายการคำยกเว้นดังต่อไปนี้

a	everywhere	N	Thanx
a's	Ex	Name	That
able	Exactly	Namely	that's
about	Example	Nd	thats
above	Except	near	the
according	F	nearly	their
accordingly	Far	necessary	theirs
across	Few	need	them
actually	Fifth	needs	themselves
after	First	neither	then
afterwards	Five	never	thence
again	Followed	nevertheless	there
against	Following	New	there's
ain't	Follows	Next	thereafter
all	For	Nine	thereby
allow	Former	No	therefore
allows	Formerly	nobody	therein
almost	Forth	Non	theres
alone	Four	none	thereupon
along	From	noone	these
already	Further	Nor	they
also	Furthermore	normally	they'd

although	G	Not	they'll
always	Get	nothing	they're
am	Gets	novel	they've
among	Getting	Now	think
amongst	Given	nowhere	third
an	Gives	O	this
and	Go	obviously	thorough
another	Goes	Of	thoroughly
any	Going	Off	those
anybody	Gone	often	though
anyhow	Got	Oh	three
anyone	Gotten	Ok	through
anything	Greetings	okay	throughout
anyway	H	Old	thru
anyways	Had	On	thus
anywhere	hadn't	once	to
apart	Happens	One	together
appear	Hardly	ones	too
appreciate	Has	Only	took
appropriate	hasn't	Onto	toward
are	Have	Or	towards
aren't	haven't	other	tried
around	Having	others	tries
as	He	otherwise	truly
aside	he's	ought	try
ask	Hello	Our	trying
asking	Help	Ours	twice
associated	Hence	ourselves	two
at	Her	Out	u

available	Here	outside	un
away	here's	over	under
awfully	Hereafter	overall	unfortunately
b	Hereby	Own	unless
be	Herein	P	unlikely
became	Hereupon	particular	until
because	Hers	particularly	unto
become	Herself	Per	up
becomes	Hi	perhaps	upon
becoming	Him	placed	us
been	Himself	please	use
before	His	Plus	used
beforehand	Hither	possible	useful
behind	Hopefully	presumably	uses
being	How	probably	using
believe	Howbeit	provides	usually
below	However	Q	uucp
beside	I	Que	v
besides	i'd	quite	value
best	i'll	Qv	various
better	i'm	R	very
between	i've	rather	via
beyond	le	Rd	viz
both	If	Re	vs
brief	Ignored	really	w
but	Immediate	reasonably	want
by	In	regarding	wants
c	Inasmuch	regardless	was
c'mon	Inc	regards	wasn't

c's	Indeed	relatively	way
came	Indicate	respectively	we
can	Indicated	right	we'd
can't	Indicates	S	we'll
cannot	Inner	Said	we're
cant	Insofar	same	we've
cause	Instead	Saw	welcome
causes	Into	Say	well
certain	Inward	saying	went
certainly	Is	says	were
changes	isn't	second	weren't
clearly	It	secondly	what
co	it'd	See	what's
com	it'll	seeing	whatever
come	it's	seem	when
comes	Its	seemed	whence
concerning	Itself	seeming	whenever
consequently	J	seems	where
consider	Just	seen	where's
considering	K	Self	whereafter
contain	Keep	selves	whereas
containing	Keeps	sensible	whereby
contains	Kept	Sent	wherein
corresponding	Know	serious	whereupon
could	Knows	seriously	wherever
couldn't	Known	seven	whether
course	L	several	which
currently	Last	shall	while
d	Lately	She	whither

definitely	Later	should	who
described	Latter	shouldn't	who's
despite	Latterly	since	whoever
did	Least	Six	whole
didn't	Less	So	whom
different	Lest	some	whose
do	Let	somebody	why
does	let's	somehow	will
doesn't	Like	someone	willing
doing	Liked	something	wish
don't	Likely	sometime	with
done	Little	sometimes	within
down	Look	somewhat	without
downwards	Looking	somewhere	won't
during	Looks	soon	wonder
e	Ltd	sorry	would
each	M	specified	wouldn
edu	Mainly	specify	wouldn't
eg	Many	specifying	x
eight	May	Still	y
either	Maybe	Sub	yes
else	Me	such	yet
elsewhere	Mean	Sup	you
enough	Meanwhile	Sure	you'd
entirely	Merely	T	you'll
especially	Might	t's	you're
et	More	Take	you've
etc	Moreover	taken	your
even	Most	Tell	yours

ever	Mostly	tends	yourself
every	Much	Th	yourselves
everybody	Must	Than	z
everyone	My	thank	zero
everything	Myself	thanks	

ภาคผนวก ค

ขั้นตอนวิธีของพอร์ทเตอร์ (Porter's Algorithm)

ขั้นตอนวิธีของพอร์ทเตอร์วิธีนี้มีความคิดที่จะประยุกต์ใช้ชุดของกฎที่กำหนดขึ้นมาเพื่อเติมหรือปรับเปลี่ยนส่วนที่อยู่ท้ายของคำต่าง ๆ ในเอกสาร เช่น การเปลี่ยนคำที่อยู่ในรูปพหูพจน์เป็นเอกพจน์จะตัด s ที่อยู่ท้ายคำออกแสดงด้วยกฎ $s \rightarrow \phi$ (Baeza-Yates and Ribeiro-Neto 1999)

กฎ (Rule) ของขั้นตอนวิธีของพอร์ทเตอร์กำหนดข้อตกลงเริ่มแรกดังนี้ (Baeza-Yates and Ribeiro-Neto 1999)

- ตัวแปรกลมกลืนถูกแสดงโดย C แต่ไม่ใช่ตัวอักษร a,e,i,o,u และไม่ใช่ตัวอักษร y ที่ตามหลังตัวแปรกลมกลืน
- ตัวแปรที่เป็นเสียงสระถูกแสดงโดย V แสดงถึงทุกตัวอักษรที่ไม่ใช่ตัวแปรกลมกลืน
- ตัวอักษรโดยทั่วไปถูกแสดงโดย L
- สัญลักษณ์ ϕ ถูกใช้เพื่อแสดงถึงสายอักขระว่าง
- การรวมกันของ C V และ L จะเรียกว่า pattern
- สัญลักษณ์ * ถูกใช้เพื่อแสดงถึงการเกิดซ้ำของ pattern ตั้งแต่ศูนย์หรือมากกว่าของ
- สัญลักษณ์ + ถูกใช้เพื่อแสดงถึงการเกิดซ้ำของ pattern ตั้งแต่หนึ่งหรือมากกว่าของ
- วงเล็บจะถูกใช้ในส่วนของขยายของลำดับตัวแปรที่ใช้ตัวดำเนินการ * และ +
- Pattern ธรรมดาถูกรวมตัวโดยสัญลักษณ์ วงเล็บและตัวดำเนินการ * และ +
- กฎการแทนที่จะถูกปฏิบัติเช่นเดียวกับคำสั่งที่แยกโดย ;
- กฎการแทนที่ถูกประยุกต์ไปเป็นส่วนต่อท้ายคำ (suffix) ในคำปัจจุบัน
- เงื่อนไขคำสั่ง if จะแสดงเป็น "if (pattern) rule" และ rule นั้นจะถูกดำเนินการเมื่อ pattern นั้นอยู่ในเงื่อนไขที่ตรงกับ current word
- บรรทัดที่เริ่มจาก "%" จะแสดงถึงข้อคิดเห็น (comment)
- "{}" จะใช้เพื่อสร้างคำสั่งที่ประกบกัน
- จะเลือกกฎเดียวที่จะดำเนินการจากทุกกฎในคำสั่งที่ประกบกัน โดยที่การเลือกกฎนั้นจะต้องเหมือนกับส่วนต่อท้ายคำ (suffix) ที่ยาวที่สุด

ตัวอย่างคำสั่ง

$if(*V * L) then ed \rightarrow \phi$

แสดงว่าจะมี การแทนที่ suffix "ed" ที่เป็นว่าง ถ้าคำในปัจจุบันนั้นมีเสียงสระและตัวอักษรอย่างน้อยหนึ่งตัว

ขั้นตอนวิธีของพอร์ทเตอร์จะประยุกต์ใช้กับคำทุกคำในเอกสาร ซึ่งจะมีกระบวนการคำสั่งดังต่อไปนี้

% Phase 1: Plural and past participles.

Select rule with longest suffix {

sses → ss;

ies → I;

ss → ss;

s → ϕ ;

}

select rule with longest suffix{

If $((C)^*((V)+(C)^+)+(V)^*eed)$ then $eed \rightarrow ee$;

If $(*V^*ed$ or $*V^*ing)$ then{

select rule with longest suffix {

$ed \rightarrow \phi$;

$ing \rightarrow \phi$;

select rule with longest suffix {

$at \rightarrow ate$;

$bl \rightarrow ble$;

$iz \rightarrow ize$;

if $((*C_1C_2)$ and $(C_1 = C_2)$ and $(C_1 \notin \{l,s,z\}))$

then $C_1C_2 \rightarrow C_1$;

if $((((C)^*((V)+(C)^+)C_1V_1C_2)$ and $(C_2 \notin \{w,x,y\}))$

then $C_1V_1C_2 \rightarrow C_1V_1C_2e$;

}

}

If $(*V^*y)$ then $y \rightarrow I$;

If $(C)^*((V)+(C)^+)+V)$ then

select rule with longest suffix {

ational → ate;

tional → tion;

enci → ence;

anci → ance;

izer → ize;

abli → able;

alli → al;

entli → ent;

eli → e;

ousli → ous;

ization → ize;

ation → ate;

ator → ate;

alism → al;

iveness → ive;

fulness → ful;

ousness → ous;

aliti → al;

iviti → ive;

biliti → ble; }

if((C)*((V)+(C)+)+(V)*) then

select rule with longest suffix {

icate → ic;

ative → ϕ ;

alize → al;

iciti → ic;

ical → ic;

ful → ϕ ;

ness $\rightarrow \phi$;

if $((C)^*(V)+(C)+((V)+(C)+)+V^*)$ then

select rule with longest suffix {

al $\rightarrow \phi$;

ance $\rightarrow \phi$;

ence $\rightarrow \phi$;

er $\rightarrow \phi$

ic $\rightarrow \phi$;

able $\rightarrow \phi$;

ible $\rightarrow \phi$;

ant $\rightarrow \phi$;

ement $\rightarrow \phi$;

ment $\rightarrow \phi$;

ent $\rightarrow \phi$;

ou $\rightarrow \phi$;

ism $\rightarrow \phi$;

ate $\rightarrow \phi$;

iti $\rightarrow \phi$;

ous $\rightarrow \phi$;

ive $\rightarrow \phi$;

ize $\rightarrow \phi$;

if (*s or *t) then ion $\rightarrow \phi$;

select rule with longest suffix {

if $((C)^*(V)+(C)+((V)+(C)+)+V^*)$ then e $\rightarrow \phi$;

if $((C)^*(V)+(C)+V^*)$ and not $((C_1V_1C_2)$ and $(C_2 \notin \{w,x,y\}))$

then e $\rightarrow \text{nil}$; }

if $((C)^*(V)+(C)+((V)+(C)+)+V^*ll)$ then LL $\rightarrow l$

ภาคผนวก ง การออกแบบการทำงานของเครื่องมือทดสอบ

การออกแบบการทำงานของเครื่องมือทดสอบการค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ 2 รูปแบบ คือ

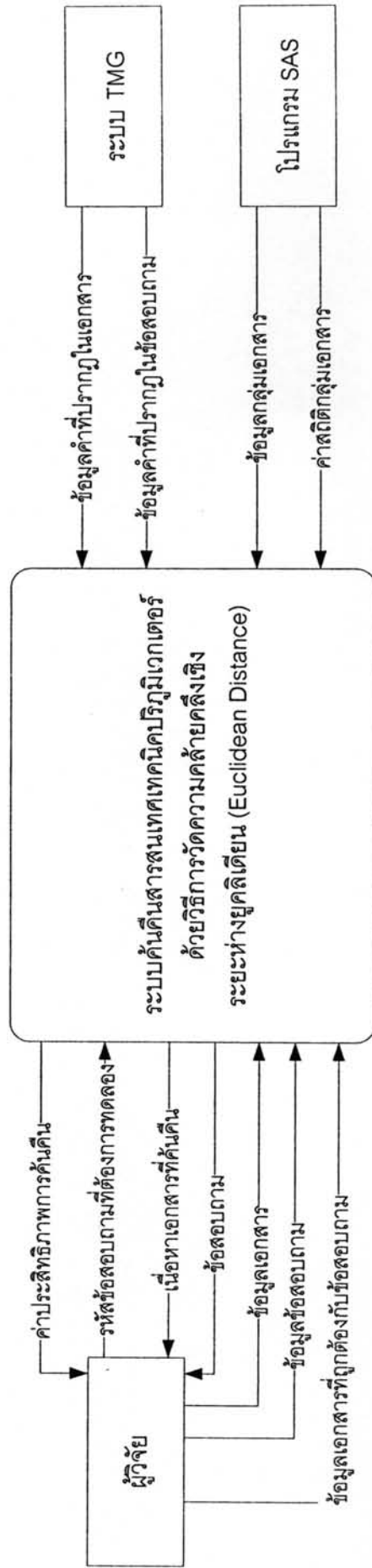
- 1) การค้นคืนเอกสารด้วยวิธีการวัดความคล้ายคลึงเชิงระยะห่างยูคลิดีเนียน (Euclidean Distance) ภายในกรอบค่าความคล้ายคลึงที่กำหนดด้วยผลลัพธ์ที่ได้จากเทคนิคการจัดกลุ่มข้อมูล (Clustering) (การค้นคืนเอกสารรูปแบบที่ 1)
- 2) การค้นคืนเอกสารด้วยวิธีการวัดความคล้ายคลึงเชิงมุม (Cosine Angle) (การค้นคืนเอกสารรูปแบบที่ 2)

ซึ่งได้ออกแบบเครื่องมือทดสอบ แผนภาพการไหลข้อมูล (Data Flow Diagram) ผังงานการทำงานของเครื่องมือ (Flowchart Diagram) การออกแบบฐานข้อมูล (Database Design) และการออกแบบหน้าจอ (Interface Design) โดยมีรายละเอียดดังต่อไปนี้

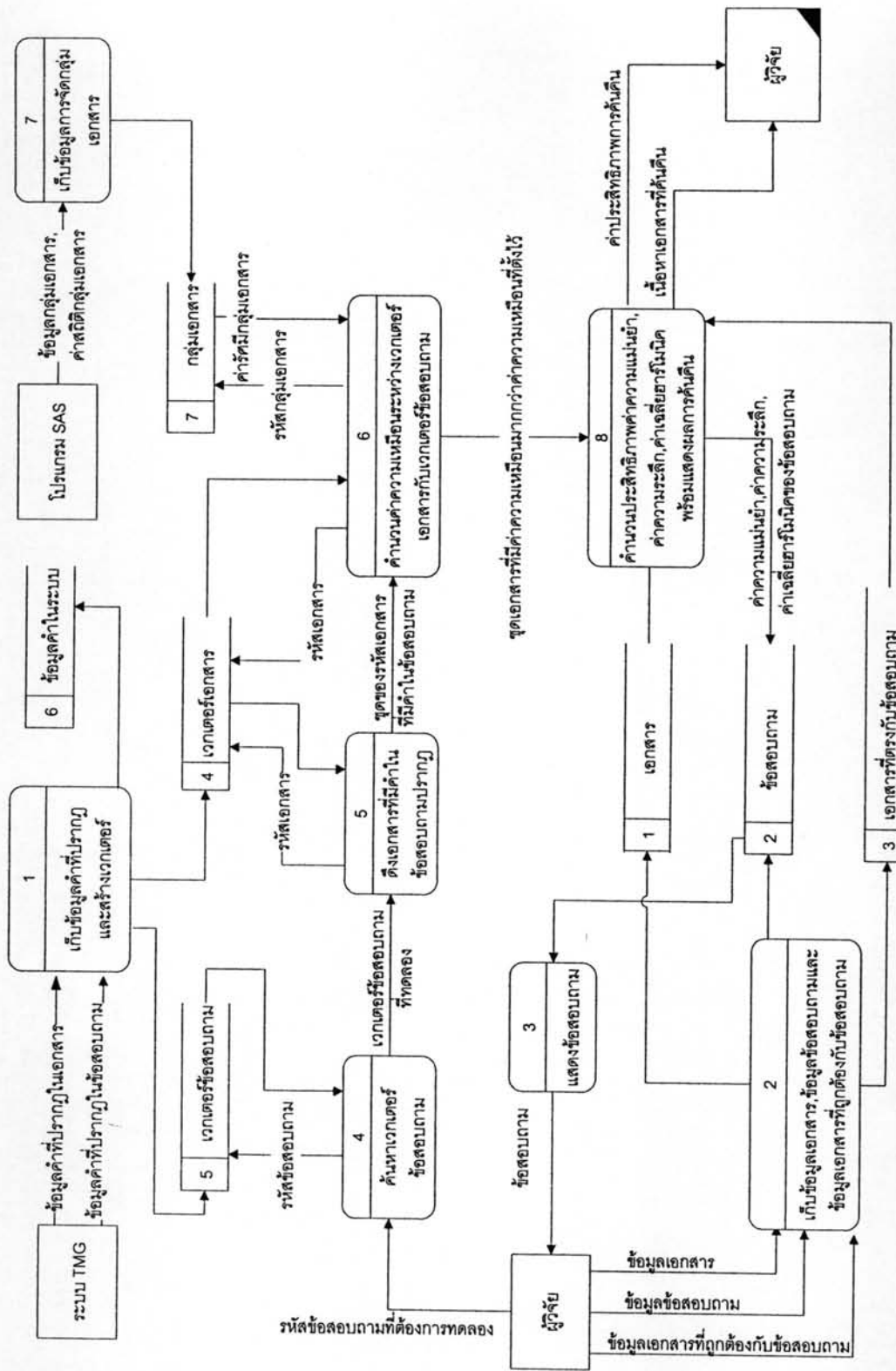
แผนภาพการไหลข้อมูล (Data Flow Diagram)

แผนภาพการไหลของข้อมูล (Data Flow Diagram) เป็นเครื่องมือที่ใช้ในการเขียนภาพที่ได้จากการวิเคราะห์ระบบ และเป็นเครื่องมือแสดงถึงทิศทางการส่งผ่านข้อมูลในระบบ แสดงความสัมพันธ์ระหว่างกระบวนการกับข้อมูลที่เกี่ยวข้องภายในระบบ แสดงการส่งผ่านข้อมูลนำเข้าและข้อมูลนำออก และขั้นตอนการทำงานของระบบ

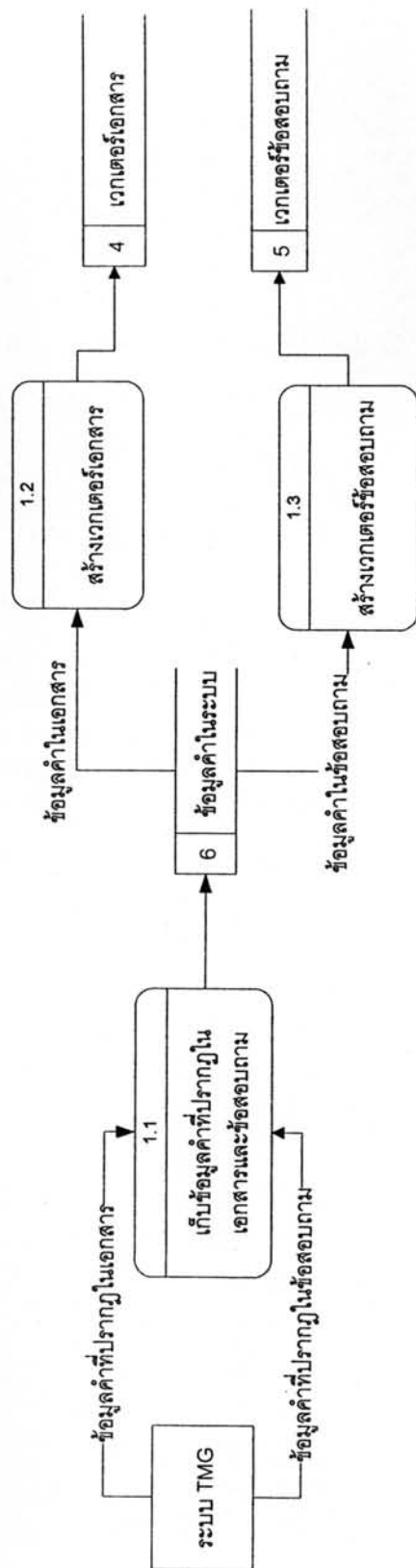
การออกแบบในส่วนของการไหลของข้อมูลในเครื่องมือทดสอบการค้นคืนเอกสารทั้ง 2 รูปแบบ สามารถแสดงได้ดังรูปที่ ง.1 ถึง ง.7



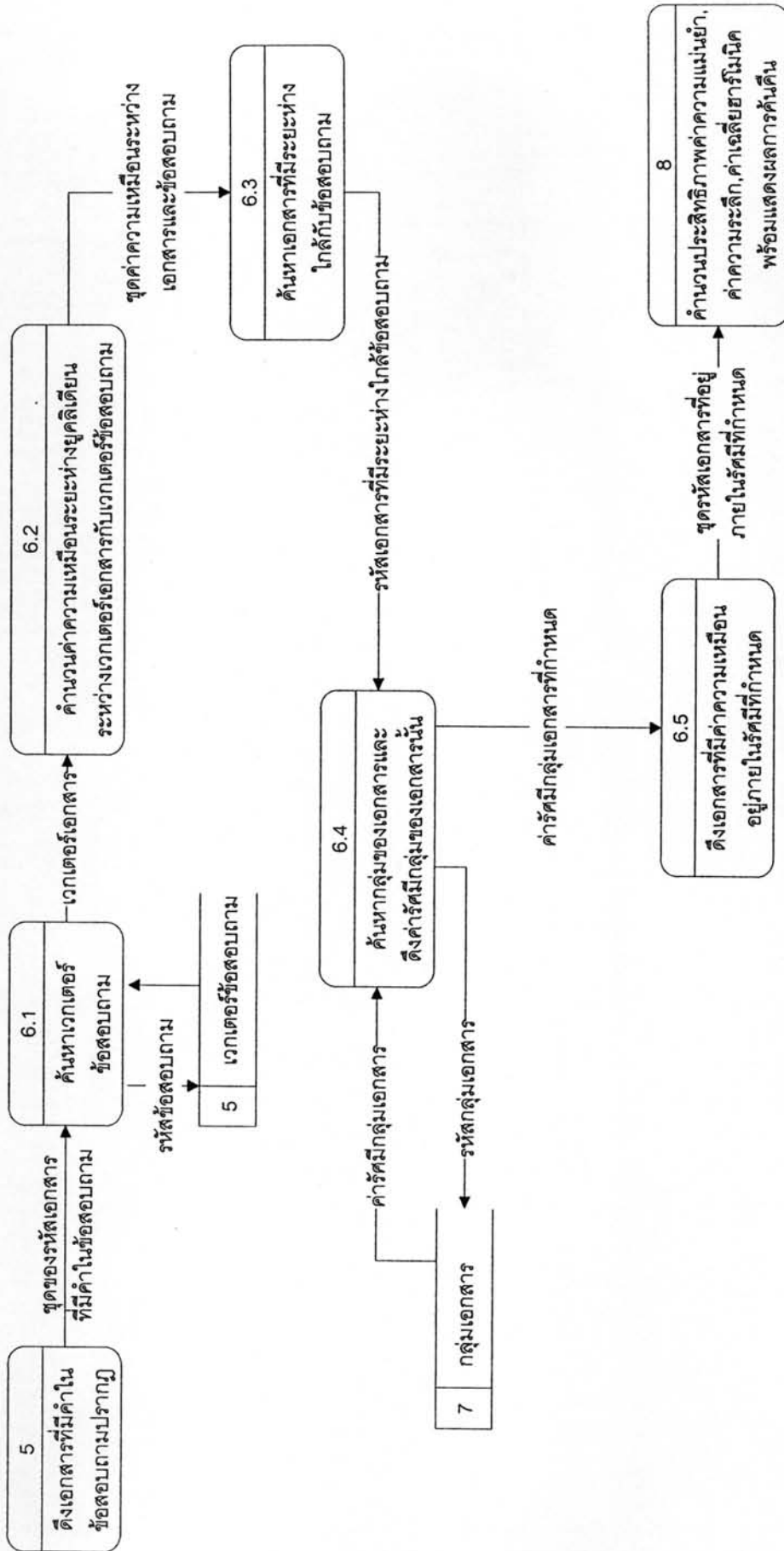
รูปที่ ง.1 รูปแบบแผนภาพการไหลของข้อมูลบริบท (Context Diagram) การค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ด้วยวิธีการวัดความคล้ายคลึงเชิง ระยะห่างยูคลิดีเนียน (Euclidean Distance) ภายในกรอบข้อความคล้ายคลึงที่กำหนดด้วยผลลัพธ์ที่ได้จากเทคนิคการจัดกลุ่มข้อมูล (Clustering)



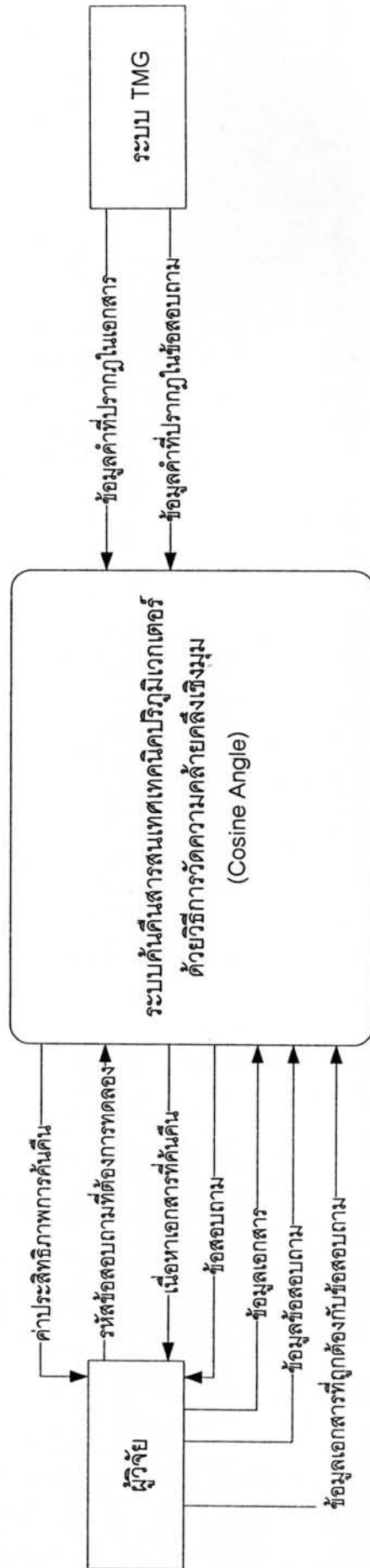
รูปที่ 4.2 รูปแสดงแผนภาพการไหลของข้อมูลระดับที่ 1 (Data Flow Diagram Level 1) ของการค้นคืนเอกสารรูปแบบที่ 1



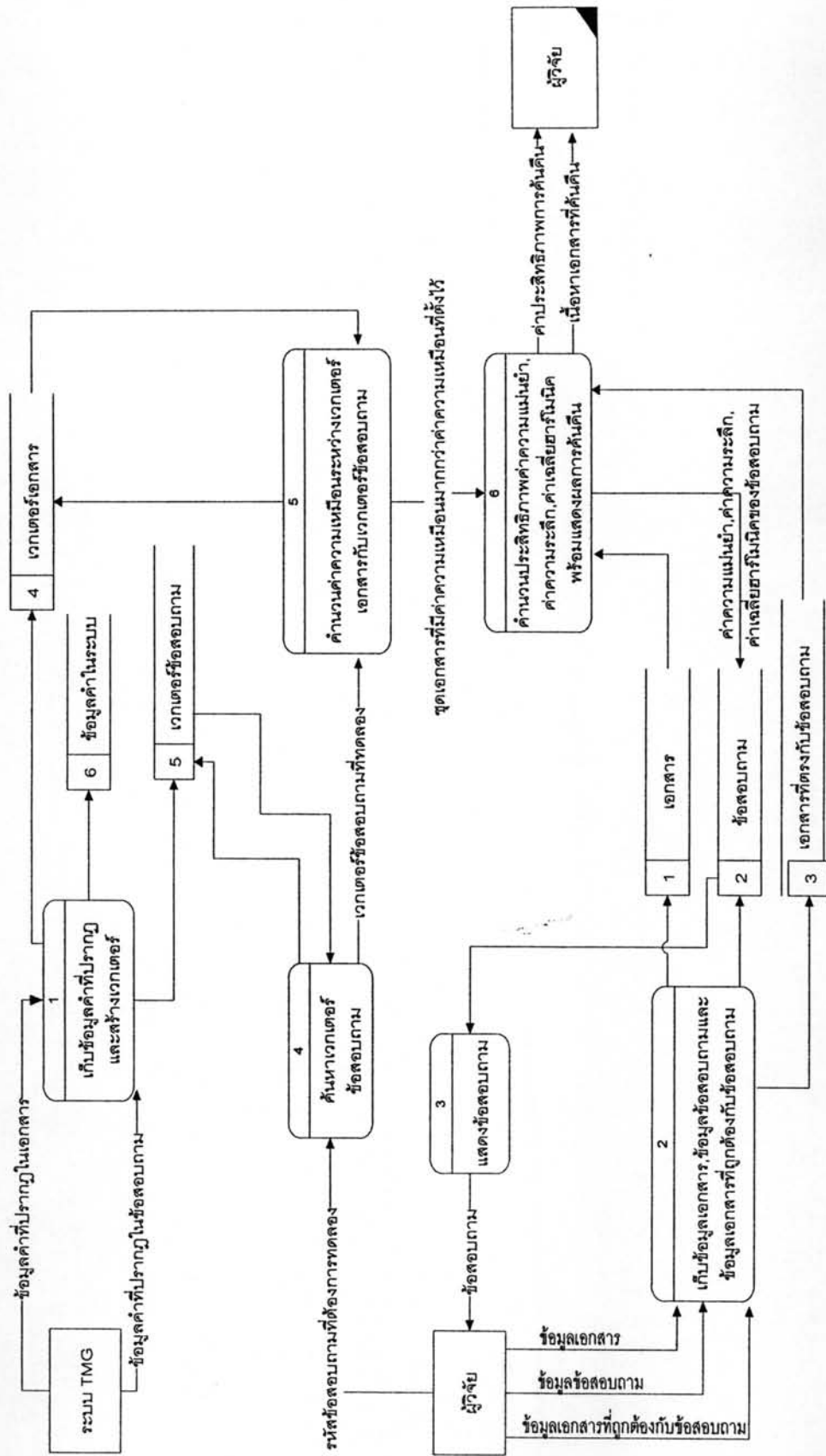
รูปที่ ๓.3 แสดงแผนภาพการไหลของข้อมูลระดับที่ 2 (Data Flow Diagram Level 2) ของ Process 1 เก็บข้อมูลค่าที่ปรากฏในเอกสาร, ข้อสอบถาม



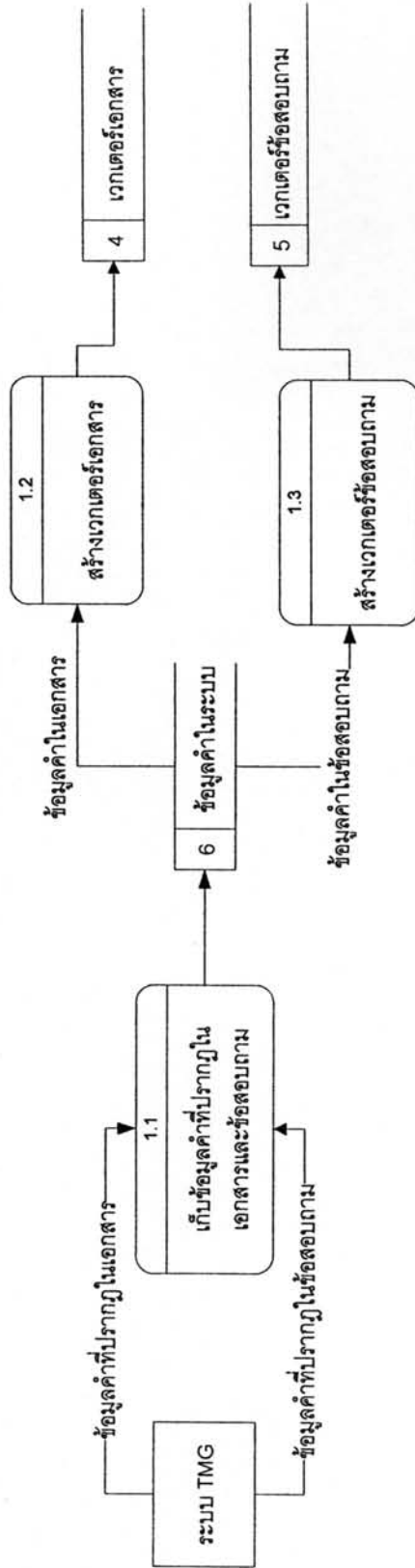
รูปที่ ๓.4 รูปแสดงแผนภาพการไหลของข้อมูลระดับที่ 2 (Data Flow Diagram Level 2) ของ Process 6 คำนวณค่าความเหมือนระหว่างเวกเตอร์เอกสารและเวกเตอร์ข้อสอบถาม



รูปที่ ๓.5 แสดงแผนภาพการไหลของข้อมูลบริบท (Context Diagram) การค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ ด้วยวิธีการวัดความคล้ายคลึงเชิงมุม (Cosine Angle)



รูปที่ ๖.๖ แสดงแผนภาพการไหลของข้อมูลระดับที่ 1 (Data Flow Diagram Level 1) ของการค้นคืนเอกสารรูปแบบที่ 2



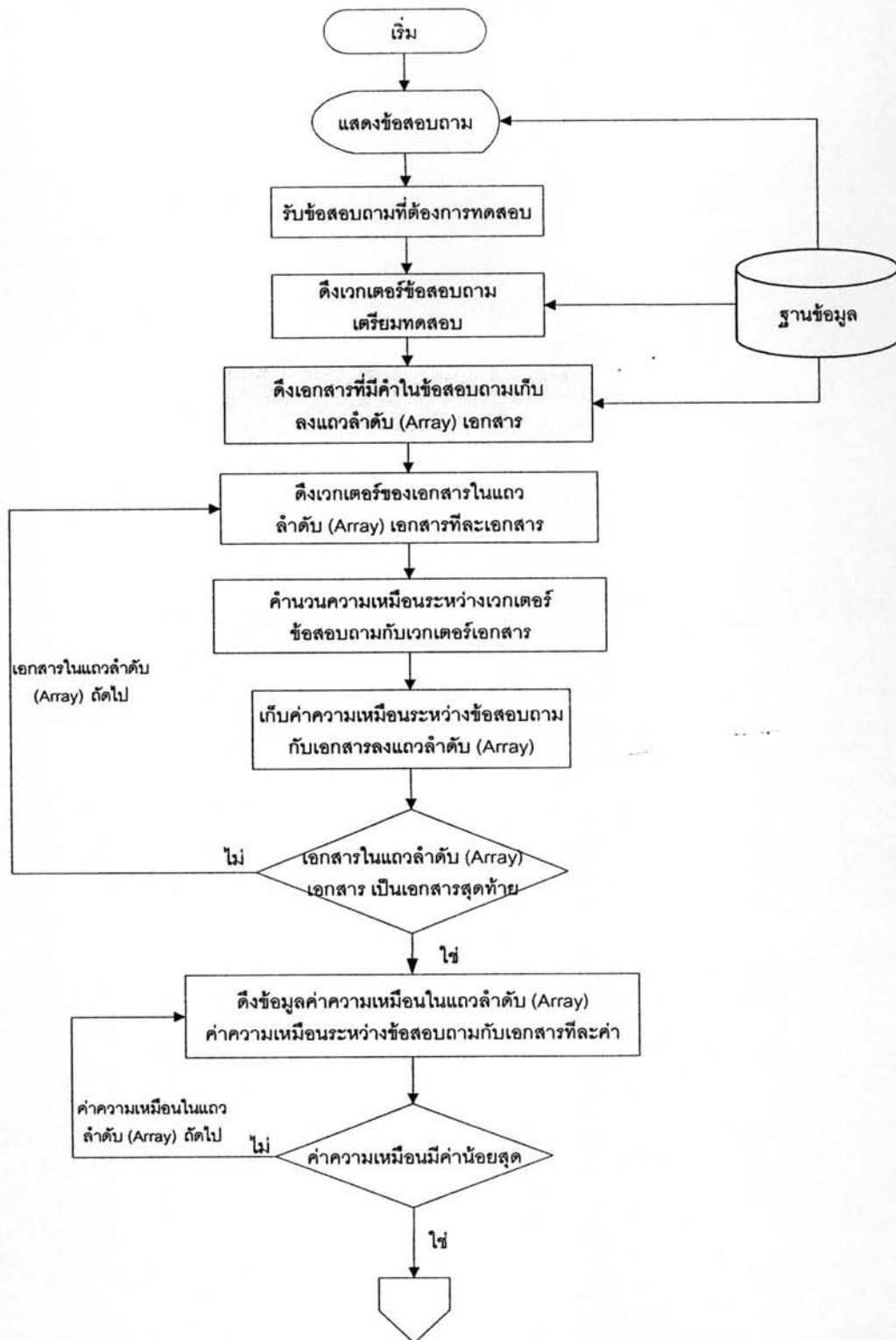
รูปที่ ง.7 แสดงแผนภาพการไหลของข้อมูลระดับที่ 2 (Data Flow Diagram Level 2) ของ Process 1 เก็บข้อมูลค่าที่ปรากฏและสร้างเวกเตอร์

การออกแบบผังงานการทำงานจากระบบ (Flowchart Diagram)

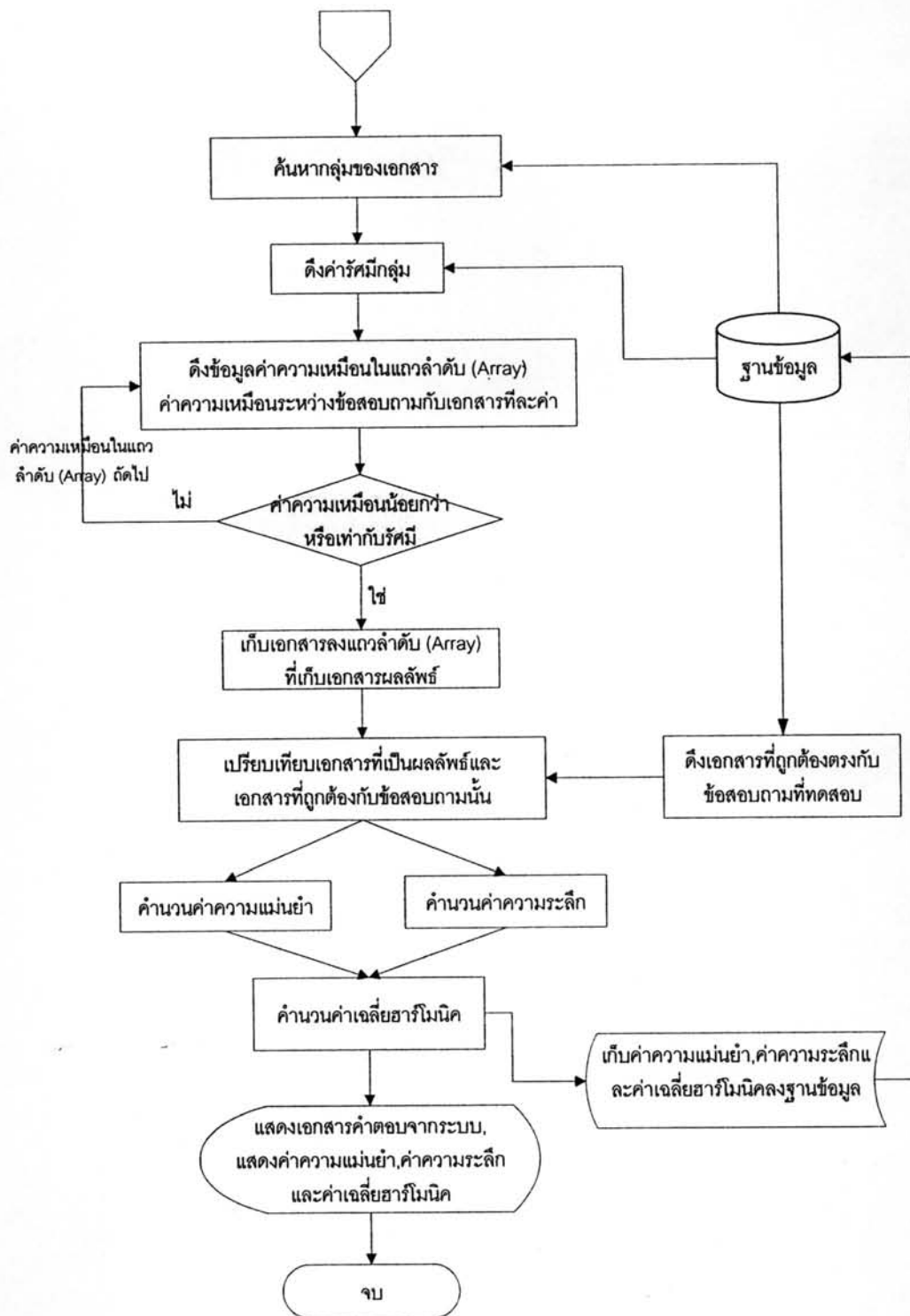
Flowchart Diagram เป็นแผนผังที่แสดงการเชื่อมโยงในกระบวนการทำงานหรือกิจกรรมที่เป็นไปตามลำดับขั้นตอน เป็นรูปสัญลักษณ์ที่บ่งบอกถึงขั้นตอนการดำเนินงานในแต่ละขั้น การออกแบบการทำงานหลักของระบบคั่นเค้นเอกสารรูปแบบต่าง ๆ มีขั้นตอนการทำงานตั้งแต่เริ่มจนถึงสิ้นสุดโปรแกรม ดังต่อไปนี้

1) ระบบคั่นเค้นเอกสารเทคนิคปริภูมิเวกเตอร์ด้วยวิธีการวัดความคล้ายคลึงเชิงระยะทางยูคลิเดียน (Euclidean Distance) ภายในกรอบค่าความคล้ายคลึงที่กำหนดด้วยผลลัพธ์ที่ได้จากเทคนิคการจัดกลุ่มข้อมูล (Clustering)

ขั้นตอนการทำงานของระบบคั่นเค้นเอกสารรูปแบบที่ 1 สามารถแสดงได้ดังรูปที่ ๘.8 และ ๘.9 โดยเริ่มการทำงานจากเครื่องมือจะดึงข้อสอบถามจากฐานข้อมูลออกมาแสดงทางหน้าจอ จากนั้นผู้วิจัยจะกำหนดข้อสอบถามที่ต้องการทดสอบเข้ามายังระบบ เครื่องมือจะดึงเวกเตอร์ข้อสอบถามที่ต้องการทดสอบจากฐานข้อมูลเก็บไว้ก่อน ต่อมาเครื่องมือไปดึงเอกสารที่มีค่าในข้อสอบถามปรากฏเก็บลงแถวลำดับ (Array) เอกสาร เครื่องมือไปดึงเวกเตอร์ของเอกสารที่เก็บไว้ในแถวลำดับ (Array) เอกสารทีละเอกสาร จากนั้นไปคำนวณหาค่าความเหมือนระหว่างเวกเตอร์เอกสารกับเวกเตอร์ข้อสอบถามนั้น ๆ ที่ได้เก็บไว้แล้ว และเก็บค่าความเหมือนลงแถวลำดับ (Array) จนกว่าเอกสารในแถวลำดับ (Array) เก็บเอกสารจะหมด จากนั้นเครื่องมือทดสอบหาค่าความเหมือนในแถวลำดับ (Array) ที่ได้เก็บไว้แล้วมาหาค่าความเหมือนน้อยสุด ถ้าค่าความเหมือนระหว่างเอกสารและข้อสอบถามใดมีค่าน้อยสุด เครื่องมือจะนำเอกสารที่มีค่าความเหมือนกับข้อสอบถามน้อยสุดนั้นไปคั่นหากกลุ่มจากฐานข้อมูล แล้วไปดึงคำรัศมีของกลุ่มนั้น ๆ เก็บไว้ ต่อมาเครื่องมือทดสอบจะเปรียบเทียบค่าความเหมือนในแถวลำดับ (Array) กับคำรัศมีที่ได้เก็บไว้แล้ว ถ้าค่าความเหมือนของเวกเตอร์เอกสารกับข้อสอบถามใดมีค่าน้อยกว่าหรือกับคำรัศมี จะเก็บเอกสารนั้นลงแถวลำดับ (Array) ที่เก็บเอกสารผลลัพธ์ เมื่อเครื่องมือได้เอกสารที่เป็นผลลัพธ์ทั้งหมดแล้ว เครื่องมือจะดึงผลลัพธ์ที่ถูกต้องของข้อสอบถามนั้นตามที่กำหนดไว้แล้วจากฐานข้อมูลนิตยสารไทม์ (TIME Collection) ในฐานข้อมูล จากนั้นนำรายการเอกสารที่เป็นผลลัพธ์มาเปรียบเทียบหาจำนวนเอกสารที่ถูกต้อง จากนั้นคำนวณหาค่าความแม่นยำ (Precision) ค่าระลึก (Recall) และค่าเฉลี่ยฮาร์โมนิก (Harmonic mean) แล้วเก็บลงฐานข้อมูล และทำการแสดงผลทางหน้าจอแก่ผู้วิจัยด้วย เป็นอันเสร็จสิ้นการคั่นเค้นเอกสารของข้อสอบถามหนึ่ง ๆ



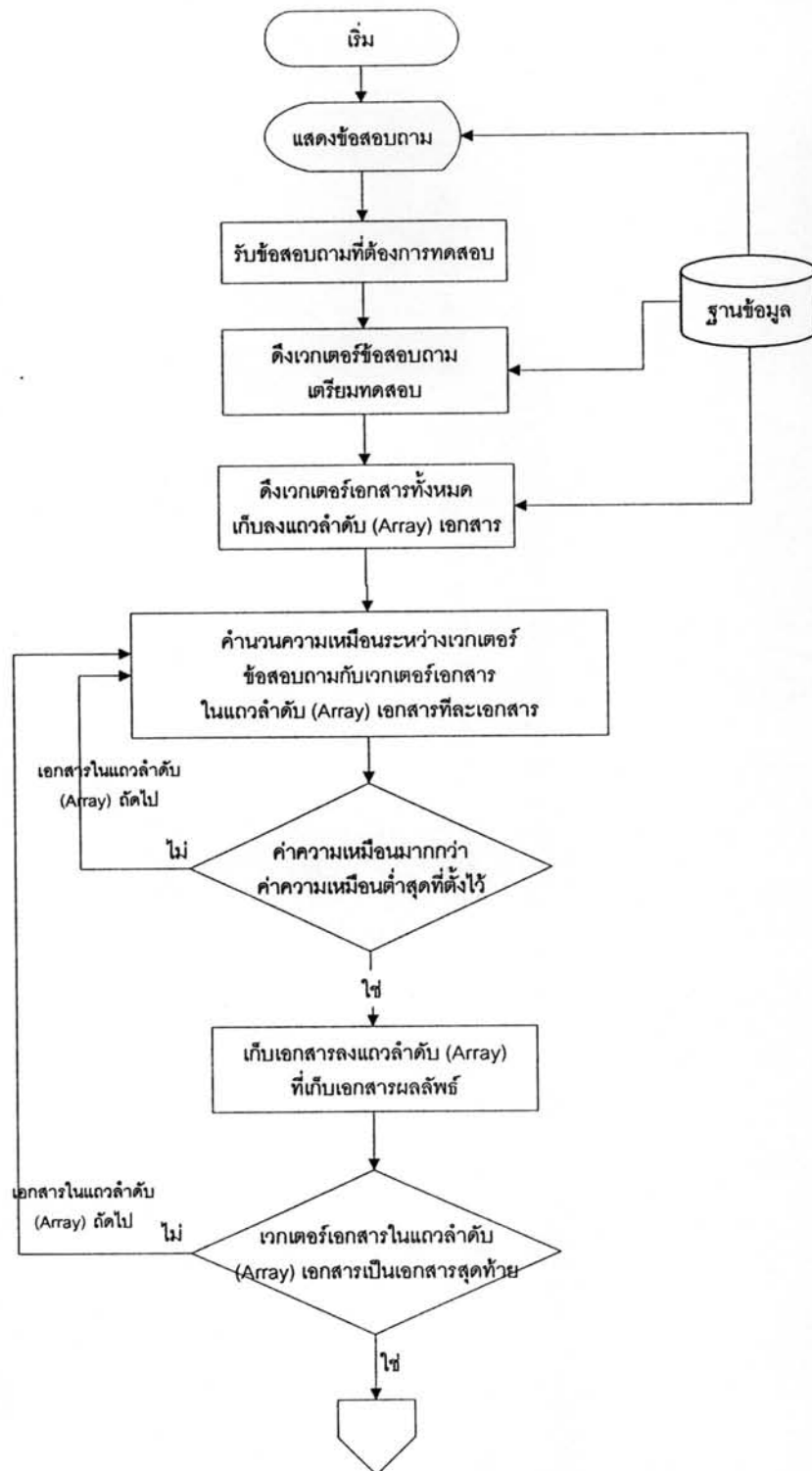
รูปที่ ๘.๘ รูปแสดงขั้นตอนการทำงานหน้า ๑ ของการค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ด้วยวิธีการวัดความคล้ายคลึงเชิงระยะห่างยูคลิเดียน (Euclidean Distance) ภายในกรอบค่าความคล้ายคลึงที่กำหนดด้วยผลลัพธ์ที่ได้จากเทคนิคการจัดกลุ่มข้อมูล (Clustering)



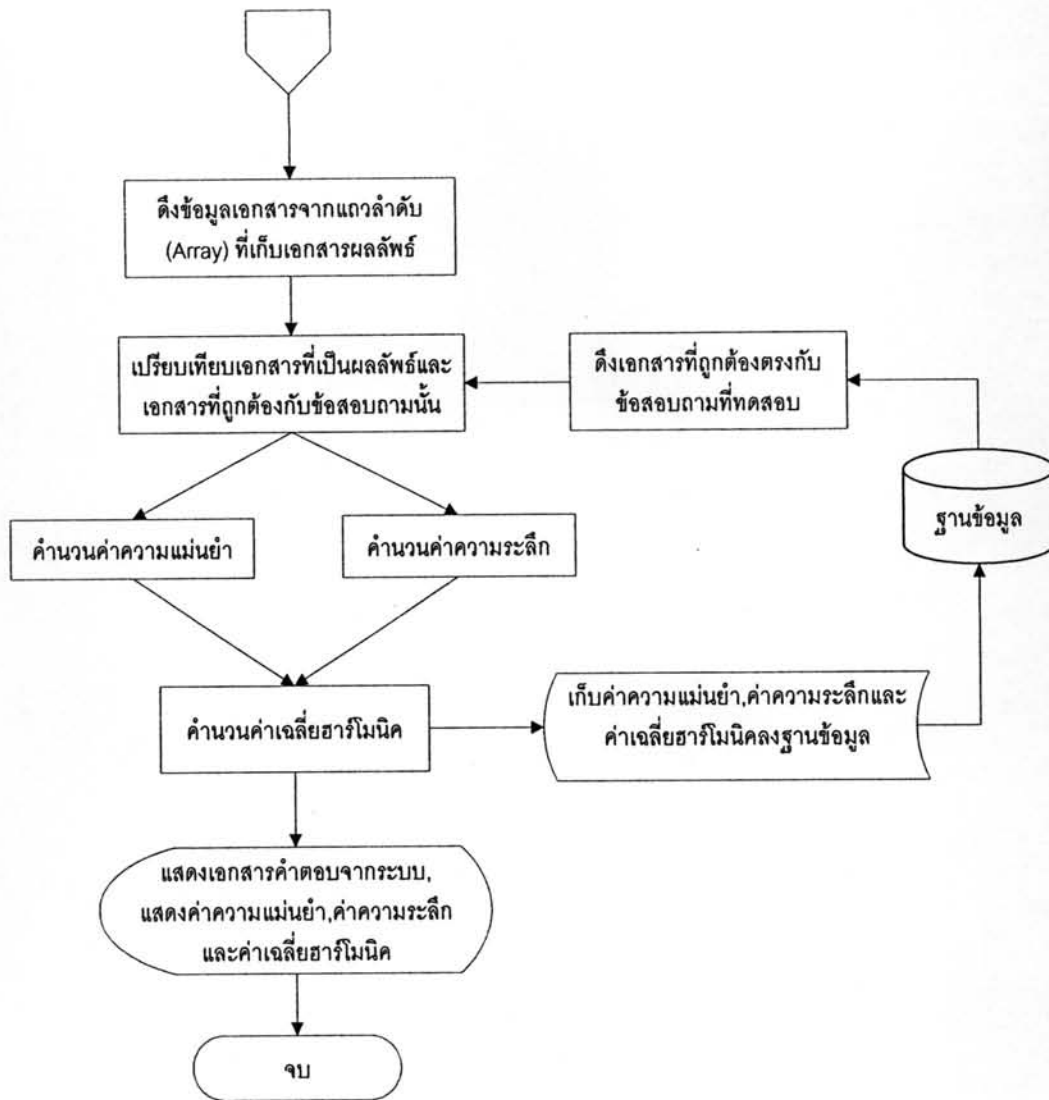
รูปที่ ง.9 รูปแสดงขั้นตอนการทำงานหน้าที่ 2 ของการค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ด้วยวิธีการวัดความคล้ายคลึงเชิงระยะห่างยูคลิเดียน (Euclidean Distance) ภายในกรอบค่าความคล้ายคลึงที่กำหนดด้วยผลลัพธ์ที่ได้จากเทคนิคการจัดกลุ่มข้อมูล (Clustering)

2) ระบบค้นคืนเอกสารเทคนิคปริภูมิเวกเตอร์ด้วยวิธีการวัดความคล้ายคลึงเชิงมุม (Cosine Angle)

ขั้นตอนการทำงานของระบบค้นคืนเอกสารรูปแบบที่ 2 สามารถแสดงได้ดังรูปที่ ง.10 และ ง.11 การทำงานเริ่มจากเครื่องมือจะดึงข้อสอบถามจากฐานข้อมูลออกมาแสดงทางหน้าจอ จากนั้นผู้วิจัยจะกำหนดข้อสอบถามที่ต้องการทดสอบเข้ามายังระบบ ต่อมาเครื่องมือจะดึงเวกเตอร์ข้อสอบถามที่ต้องการทดสอบจากฐานข้อมูลเก็บไว้ก่อน จากนั้นเครื่องมือจะไปดึงเวกเตอร์เอกสารทั้งหมดเก็บลงแถวลำดับ (Array) เอกสาร แล้วคำนวณหาค่าความเหมือนระหว่างเวกเตอร์เอกสารในแถวลำดับ (Array) เก็บเอกสารทีละเอกสารกับเวกเตอร์ข้อสอบถามนั้น ๆ ที่ได้เก็บไว้แล้ว ถ้าค่าความเหมือนของเวกเตอร์เอกสารกับข้อสอบถามใดมีค่ามากกว่าค่าความเหมือนต่ำที่สุดที่ตั้งไว้จะเก็บเอกสารนั้นลงแถวลำดับ (Array) ที่เก็บเอกสารผลลัพธ์ จนกว่าจะเอกสารในแถวลำดับ (Array) เก็บเอกสารจะหมด เมื่อเครื่องมือได้เอกสารที่เป็นผลลัพธ์ทั้งหมดแล้ว เครื่องมือจะดึงผลลัพธ์ที่ถูกต้องของข้อสอบถามนั้นตามที่กำหนดไว้แล้วจากฐานข้อมูลนิตยสารไทม์ (TIME Collection) ในฐานข้อมูล จากนั้นนำรายการเอกสารที่เป็นผลลัพธ์มาเปรียบเทียบหาจำนวนเอกสารที่ถูกต้อง จากนั้นคำนวณหาค่าความแม่นยำ (Precision) ค่าระลึก (Recall) และค่าเฉลี่ยฮาร์โมนิก (Harmonic mean) แล้วเก็บลงฐานข้อมูล และทำการแสดงผลทางหน้าจอแก่ผู้วิจัยด้วยเป็นอันเสร็จสิ้นการค้นคืนเอกสารของข้อสอบถามหนึ่ง ๆ



รูปที่ ง.10 รูปแสดงขั้นตอนการทำงานหน้าที่ 1 ของการค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ ด้วยวิธีวัดความคล้ายคลึงเชิงมุม (Cosine Angle)

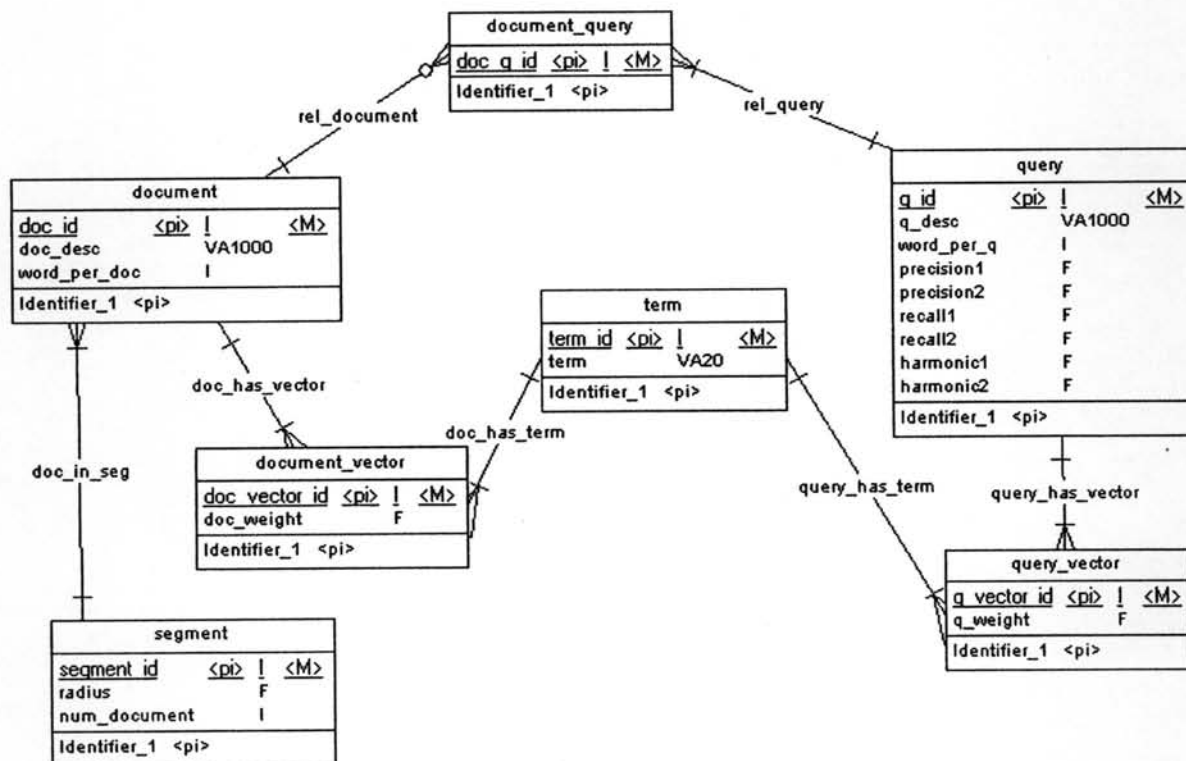


รูปที่ ง.11 รูปแสดงขั้นตอนการทำงานหน้าที่ 2 ของการค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ ด้วยวิธีวัดความคล้ายคลึงเชิงมุม (Cosine Angle)

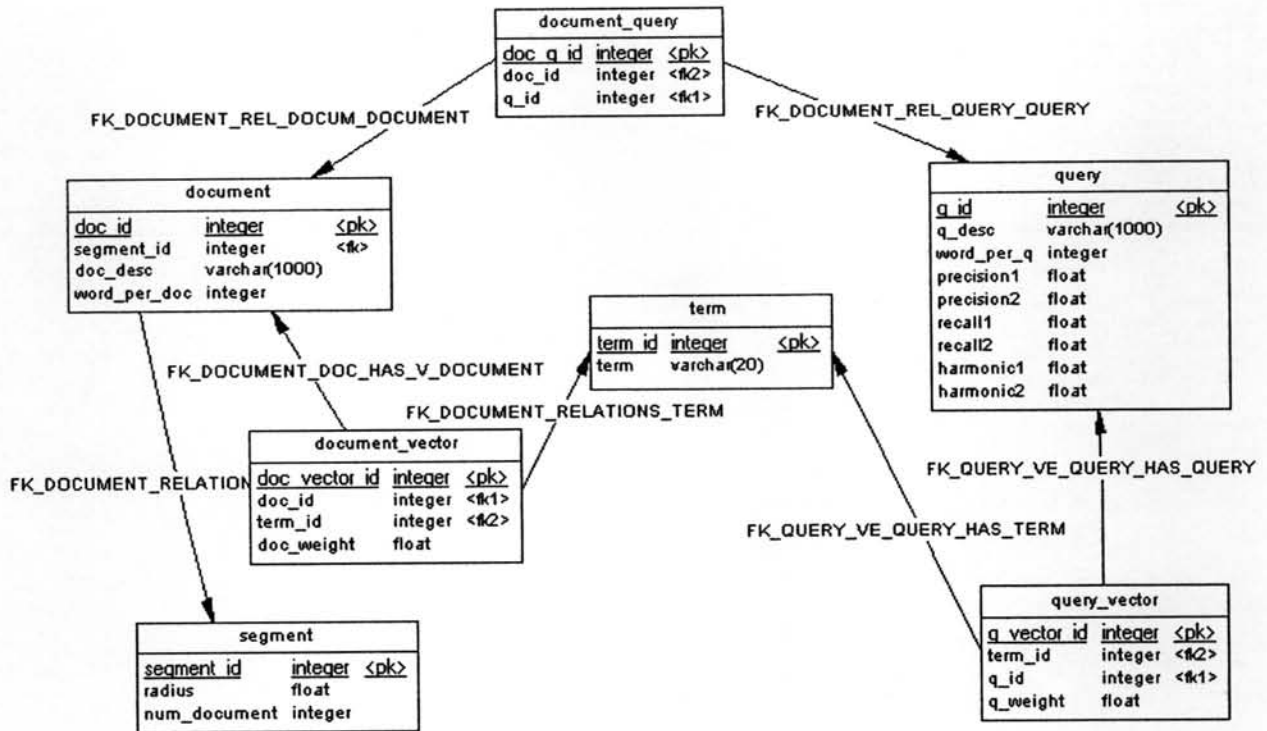
การออกแบบฐานข้อมูล (Database Design)

การออกแบบในส่วฐานข้อมูลของการค้นคืนเอกสาร เพื่อเก็บข้อมูลของการทดสอบการค้นคืนเอกสารทั้งหมดในการทดลอง โดยจะแสดง ER Diagram เป็นส่วนการออกแบบความสัมพันธ์ของตารางทั้งหมดในฐานข้อมูลของการค้นคืนเอกสาร

1) ER Diagram



รูปที่ ง.12 รูปแสดงแผนภาพเชิงแนวคิด (Conceptual Diagram)



รูปที่ ง.13 รูปแสดงแผนภาพเชิงกายภาพ (Physical Diagram)

2) พจนานุกรมข้อมูล (Data Dictionary)

ชื่อตาราง :	document_quiry			
คำอธิบาย :	ตารางเก็บข้อมูลเอกสารที่เป็นคำตอบของข้อสอบถามแต่ละข้อสอบถาม			
ชื่อคอลัมน์	คีย์	ชนิด	ความยาว	คำอธิบาย
doc_q_id	PK	integer	4	คีย์หลักของตารางเก็บข้อมูลเอกสารคำตอบของข้อสอบถามแต่ละข้อสอบถาม
doc_id	FK	integer	4	รหัสเอกสารซึ่งเป็นคีย์อ้างอิงมาจากตารางเอกสาร โดยจะจับคู่กับข้อสอบถามที่เกี่ยวข้องเนื่องกัน
q_id	FK	integer	4	รหัสข้อสอบถามซึ่งเป็นคีย์อ้างอิงมาจากตารางข้อสอบถาม โดยจะจับคู่กับเอกสารที่เกี่ยวข้องเนื่องกัน

ชื่อตาราง :	document			
คำอธิบาย :	ตารางเก็บข้อมูลเอกสารนิตยสารไทม์			
ชื่อคอลัมน์	คีย์	ชนิด	ความยาว	คำอธิบาย
doc_id	PK	integer	4	รหัสเอกสาร ซึ่งใช้เป็นคีย์หลักตารางเอกสาร
segment_id	FK	integer	4	รหัสกลุ่มเอกสาร ซึ่งเป็นคีย์อ้างอิงมาจากตารางกลุ่มเอกสาร
doc_desc		varchar	1000	บทความของเอกสารที่ใช้ในการทดสอบการค้นคืนเอกสาร โดยจะเก็บเพียงส่วนหนึ่งของบทความในเอกสารทั้งหมด
word_per_doc		integer	4	จำนวนคำของเอกสารนั้น ๆ

ชื่อตาราง :	query			
คำอธิบาย :	ตารางเก็บข้อมูลข้อสอบถามจากนิตยสารไทม์			
ชื่อคอลัมน์	คีย์	ชนิด	ความยาว	คำอธิบาย
q_id	PK	integer	4	รหัสข้อสอบถาม ซึ่งใช้เป็นคีย์หลักตารางข้อสอบถาม
q_desc		varchar	1000	ข้อสอบถามที่ใช้ในการทดลอง มีรูปแบบเป็นข้อความ
word_per_q		integer	4	จำนวนคำของข้อสอบถาม
precision1		integer	4	ค่าความแม่นยำ (precision) ของการค้นคืนเอกสารวิธีการวัดความคล้ายคลึงเชิงระยะห่างยูคลิเดียน ภายในกรอบค่าความคล้ายคลึงที่กำหนดด้วยผลลัพธ์ที่ได้จากเทคนิคการจัดกลุ่มข้อมูล (Clustering)
precision2		integer	4	ค่าความถูกต้อง (precision) ของการค้นคืนเอกสารวิธีการวัดความคล้ายคลึงเชิงมุม
recall1		integer	4	ค่าความระลึก (recall) ของการค้นคืนเอกสารวิธีการวัดความคล้ายคลึงเชิงระยะห่างยูคลิเดียน ภายในกรอบค่าความคล้ายคลึงที่กำหนดด้วยผลลัพธ์ที่ได้จากเทคนิคการจัดกลุ่มข้อมูล (Clustering)

recall2		integer	4	ค่าความระลึก (recall) ของการค้นคืนเอกสารวิธีการวัดความคล้ายคลึงเชิงมุม
harmonic1		integer	4	ค่าเฉลี่ยฮาร์โมนิก (Harmonic mean) ของการค้นคืนเอกสารวิธีการวัดความคล้ายคลึงเชิงระยะห่างยูคลิเดียน ภายในกรอบค่าความคล้ายคลึงที่กำหนดด้วยผลลัพธ์ที่ได้จากเทคนิคการจัดกลุ่มข้อมูล (Clustering)
harmonic2		integer	4	ค่าเฉลี่ยฮาร์โมนิก (Harmonic mean) ของการค้นคืนเอกสารวิธีการวัดความคล้ายคลึงเชิงมุม

ชื่อตาราง :	document_vector			
คำอธิบาย :	ตารางเก็บเวกเตอร์เอกสารโดยจะเก็บรหัสเอกสารและรหัสคำที่อยู่ในเอกสารนั้น ๆ			
ชื่อคอลัมน์	คีย์	ชนิด	ความยาว	คำอธิบาย
doc_vector_id	PK	integer	4	คีย์หลักของตารางเวกเตอร์เอกสาร
doc_id	FK	integer	4	รหัสเอกสารซึ่งเป็นคีย์อ้างอิงมาจากตารางเอกสาร โดยจะจับคู่กับรหัสคำที่ปรากฏในเอกสารนั้น ๆ
term_id	FK	integer	4	รหัสคำซึ่งเป็นคีย์อ้างอิงมาจากตารางคำ โดยจะจับคู่กับเอกสารที่มีค่านั้นๆปรากฏ
doc_weight		float	8	ค่าน้ำหนักของค่านั้น ๆ ในเอกสารหนึ่ง ๆ

ชื่อตาราง :	query_vector			
คำอธิบาย :	ตารางเก็บเวกเตอร์ข้อสอบถามโดยจะเก็บรหัสข้อสอบถามและรหัสคำที่อยู่ในข้อสอบถามนั้น ๆ			
ชื่อคอลัมน์	คีย์	ชนิด	ความยาว	คำอธิบาย
q_vector_id	PK	integer	4	คีย์หลักของเวกเตอร์ข้อสอบถาม
q_id	FK	integer	4	รหัสข้อสอบถามซึ่งเป็นคีย์อ้างอิงมาจากตารางข้อสอบถาม โดยจะจับคู่กับรหัสคำที่

				ปรากฏในข้อสอบถามนั้น ๆ
term_id	FK	integer	4	รหัสคำซึ่งเป็นคีย์อ้างอิงมาจากตารางคำ โดยจะจับคู่กับข้อสอบถามที่มีค่านั้นๆปรากฏ
q_weight		float	8	ค่าน้ำหนักของค่านั้น ๆ ในข้อสอบถามหนึ่ง ๆ

ชื่อตาราง :	term			
คำอธิบาย :	ตารางเก็บคำทั้งหมดที่ปรากฏในเอกสารและข้อสอบถามที่นำมาทดสอบระบบการค้นคืนเอกสาร			
ชื่อคอลัมน์	คีย์	ชนิด	ความยาว	คำอธิบาย
term_id	PK	integer	4	รหัสคำ ซึ่งเป็นคีย์หลักของตารางคำ
term		varchar	20	รายการคำต่าง ๆ

ชื่อตาราง :	segment			
คำอธิบาย :	ตารางกลุ่ม เป็นตารางเก็บคำรัศมีในแต่ละกลุ่มเอกสาร			
ชื่อคอลัมน์	คีย์	ชนิด	ความยาว	คำอธิบาย
segment_id	PK	integer	4	รหัสกลุ่ม ซึ่งเป็นคีย์หลักของตารางกลุ่ม
radius		float	8	คำรัศมีในแต่ละกลุ่ม
num_document		integer	4	จำนวนเอกสารทั้งหมดในกลุ่ม

การออกแบบหน้าจอ (Interface Design)

การออกแบบหน้าจอของเครื่องมือทดสอบการค้นคืนเอกสารทั้ง 2 รูปแบบนั้น จะมีทั้งหมด 3 หน้าจอ ซึ่งทั้ง 3 หน้าจอจะปรากฏอยู่ในการค้นคืนเอกสารทั้ง 2 รูปแบบ ดังนี้

หน้าจอ 1 : หน้าจอเลือกข้อสอบถาม

หน้าจอนี้จะแสดง Radio Button ของข้อสอบถามทั้งหมด เพื่อให้ผู้ใช้เลือกข้อสอบถามที่ต้องการจะทดสอบ ผู้ใช้สามารถเลือกข้อสอบถามที่ต้องการจะทดสอบ จากนั้นกดปุ่ม OK เพื่อเข้าสู่หน้าจอถัดไป ดังรูปที่ ง.14

ลำดับ	รายการข้อสอบถาม
○	1. KENNEDY ADMINISTRATION PRESSURE ON NGO DINH DIEM TO STOP SUPPRESSING THE BUDDHISTS .
○	2. EFFORTS OF AMBASSADOR HENRY CABOT LODGE TO GET VIET NAM'S PRESIDENT DIEM TO CHANGE HIS POLICIES OF POLITICAL REPRESSION .
○	3. NUMBER OF TROOPS THE UNITED STATES HAS STATIONED IN SOUTH VIET NAM AS COMPARED WITH THE NUMBER OF TROOPS IT HAS STATIONED IN WEST GERMANY .
○	4. U.S . POLICY TOWARD THE NEW REGIME IN SOUTH VIET NAM WHICH OVERTHREW PRESIDENT DIEM .
○	5. PERSONS INVOLVED IN THE VIET NAM COUP .
○	6. CEREMONIAL SUICIDES COMMITTED BY SOME BUDDHIST MONKS IN SOUTH VIET NAM AND WHAT THEY ARE SEEKING TO GAIN BY SUCH ACTS .
○	7. REJECTION BY PRINCE NORODOM SIHANOUK, AN ASIAN NEUTRALIST LEADER, OF ALL FURTHER U.S . AID TO HIS NATION .
○	8. U.N . TEAM SURVEY OF PUBLIC OPINION IN NORTH BORNEO AND SARAWAK ON THE QUESTION OF JOINING THE FEDERATION OF MALAYSIA .
○	9. OPPOSITION OF INDONESIA TO THE NEWLY-CREATED MALAYSIA .
	GROWING CONTROVERSY IN SOUTHEAST ASIA OVER THE PROPOSED CREATION OF A FEDERATION OF

รูปที่ ง.14 รูปแสดงหน้าจอแรก สำหรับเลือกข้อสอบถามที่ต้องการทดสอบ

หน้าจอ 2 : หน้าจอยืนยันข้อสอบถามที่ต้องการทดสอบ

หน้าจอนี้จะแสดงข้อสอบถามที่ได้เลือกไปแล้วจากหน้าจอแรก เพื่อยืนยันให้ผู้ใช้ทราบอีกครั้งหนึ่งของข้อสอบถามที่ต้องการทดสอบ จากนั้นกดปุ่ม OK เพื่อเข้าสู่หน้าจอถัดไป ดังรูปที่ ง.15

ข้อสอบถามที่คุณเลือกคือ
ข้อสอบถามที่ 61: NATIONS WORKING ON NUCLEAR WEAPONS DEVELOPMENT .

รูปที่ ง.15 รูปแสดงหน้าจอยืนยันการเลือกข้อสอบถามที่ต้องการทดสอบ

หน้าจอ 3 : หน้าจอแสดงผลการค้นคืนเอกสาร .

หน้าจอนี้เป็นหน้าจอสุดท้าย เป็นการแสดงผลลัพธ์หลังจากที่ผ่านกระบวนการที่ออกแบบไว้ในแต่ละเครื่องมือทดสอบการค้นคืนเอกสารทั้ง 2 รูปแบบ โดยจะแสดงรายการเอกสารที่ถูกค้นคืน และค่าประสิทธิภาพทั้ง 3 ค่า คือ ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าเฉลี่ยฮาร์โมนิค (Harmonic mean) ดังรูปที่ ง.16

รายการเอกสารที่ค้นคืน	
รหัสเอกสาร	รายละเอียดเอกสาร
1	THE ALLIES AFTER NASSAU IN DECEMBER 1960, THE U.S . FIRST PROPOSED TO HELP NATO DEVELOP ITS OWN NUCLEAR STRIKE FORCE . BUT EUROPE MADE NO ATTEMPT TO DEVISE A PLAN . LAST WEEK, AS THEY STUDIED THE NASSAU ACCORD BETWEEN PRESIDENT KENNEDY AND PRIME MINISTER
38	COLD WAR OF BASES & BOMBS AT FIRST GLIMPSE, IT CERTAINLY BORE ALL THE SIGNS OF A U.S.-SOVIET DEAL . FIRST THE NEWS BROKE THAT PREMIER KHRUSHCHEV, IN A SECRET LETTER TO PRESIDENT KENNEDY, HAD AGREED TO REMOVE THE BIGGEST OBSTACLE TO A NUCLEAR TEST-BAN TREA
47	EUROPE A NEW & OBSCURE DESTINATION IN AN ALLIANCE IN WHICH PARTNERS HAD BECOME INCREASINGLY MINDFUL OF ONE ANOTHER'S SENSITIVITIES, IN WHICH VICTORIES WERE TACTFULLY NOT CROWED OVER, AND TOGETHERNESS IN ITSELF WAS REGARDED AS A GOOD THING, CHARLES DE GAUL
82	WEST GERMANY TANKS, BUT NO TANKS AMONG THE MANY HIGH-SOUNDING BUT VAGUE PROVISIONS SPRINKLED THROUGH THE FRANCO-GERMAN RECONCILIATION TREATY CEREMONIALLY SIGNED IN PARIS SIX WEEKS AGO WAS ONE CALLING FOR JOINT RESEARCH AND DEVELOPMENT OF CONVENTIONAL WEAP
89	THE ALLIES THE NATO DETERRENT ALMOST AS OLD AS NATO ITSELF IS THE " DISARRAY " THAT SEEMS TO HAUNT ITS COUNCILS . YET ALLIED DIFFERENCES ARE NOT SO MUCH SYMPTOMS OF ANY DEEP-ROOTED DISUNITY AS THE RESULT OF NATO'S MILITARY EFFECTIVENESS . BECAUSE THEY KNO
	MULTI-BAFFLEMENT THE U.S . TODAY IS ENGAGED IN NOT ONE, BUT TWO NUCLEAR ARMS RACES .

รูปที่ ง.16 รูปแสดงหน้าจอผลการค้นคืนเอกสาร

ภาคผนวก จ

สรุปผลค่าประสิทธิภาพความแม่นยำ (Precision), ความระลึก (Recall) และค่าเฉลี่ยฮาร์โมนิก (Harmonic mean) ของเครื่องมือทดสอบการค้นคืน

ตารางสรุปผลค่าความแม่นยำ (Precision), ค่าความระลึก (Recall) และค่าเฉลี่ยฮาร์โมนิก (Harmonic mean) ของเครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 1 และเครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 3 เมื่อกำหนดกรอบค่าความคล้ายคลึงที่ต่างกัน ดังนี้

- 1) เครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 1 เมื่อกำหนดกรอบค่าความคล้ายคลึงด้วยค่ารัศมีกลุ่ม (R) และค่ารัศมีกลุ่มหารสอง (R/2)
- 2) เครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 1 ที่ค้นคืนเฉพาะค่าที่มีในข้อสอบถามปรากฏ (เครื่องมือทดสอบการค้นคืนรูปแบบที่ 1*) เมื่อกำหนดกรอบค่าความคล้ายคลึงด้วยค่ารัศมีกลุ่ม (R) และค่ารัศมีกลุ่มหารสอง (R/2)
- 3) เครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 3 เมื่อกำหนดกรอบค่าความคล้ายคลึงด้วยค่ารัศมีความคล้าย (RS) และค่ารัศมีความคล้ายหารสอง (RS/2)

จากการค้นคืนเอกสารทั้ง 2 รูปแบบ ด้วยการทดลองกำหนดกรอบค่าความคล้ายคลึงการค้นคืนที่ต่างกัน ให้ผลการทดลองค่าความแม่นยำ (Precision), ค่าความระลึก (Recall) และค่าเฉลี่ยฮาร์โมนิก (Harmonic mean) ดังตารางที่ จ.1, จ.2 และ จ.3 ตามลำดับ

ตารางที่ จ.1 ตารางแสดงค่าความแม่นยำ (Precision), ค่าความระลึก (Recall) และค่าเฉลี่ยฮาร์โมนิก (Harmonic mean) เครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 1 เมื่อกรอบค่าความคล้ายคลึงที่ต่างกัน

ลำดับข้อ สอบถาม	ค่าความแม่นยำ (Precision)		ค่าความระลึก (Recall)		ค่าเฉลี่ยฮาร์โมนิก (Harmonic mean)	
	รัศมีกลุ่ม (R)	รัศมีกลุ่มหาร สอง (R/2)	รัศมีกลุ่ม (R)	รัศมีกลุ่มหาร สอง (R/2)	รัศมีกลุ่ม (R)	รัศมีกลุ่มหาร สอง (R/2)
1	0.0101	0.0000	0.5714	0.0000	0.0198	0.0000
2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	0.0075	0.0143	0.7500	0.7500	0.0149	0.0280
4	0.0101	0.0142	0.8000	0.6000	0.0199	0.0278

ลำดับข้อ สอบถาม	ค่าความแม่นยำ (Precision)		ค่าความระลึก (Recall)		ค่าเฉลี่ยฮาร์โมนิก (Harmonic mean)	
	รัศมีกลุ่ม (R)	รัศมีกลุ่มหาร สอง (R/2)	รัศมีกลุ่ม (R)	รัศมีกลุ่มหาร สอง (R/2)	รัศมีกลุ่ม (R)	รัศมีกลุ่มหาร สอง (R/2)
5	0.0075	0.0095	0.6000	0.4000	0.0149	0.0185
6	0.0126	0.0000	0.5556	0.0000	0.0246	0.0000
7	0.0050	0.0047	1.0000	0.5000	0.0100	0.0094
8	0.0050	0.0047	1.0000	0.5000	0.0100	0.0094
9	0.0176	0.0237	0.8750	0.6250	0.0345	0.0457
10	0.0126	0.0190	0.8333	0.6667	0.0248	0.0369
11	0.0050	0.0047	1.0000	0.5000	0.0100	0.0094
12	0.0151	0.0190	0.8571	0.5714	0.0296	0.0367
13	0.0075	0.0000	1.0000	0.0000	0.0150	0.0000
14	0.0025	0.0047	1.0000	1.0000	0.0050	0.0094
15	0.0101	0.0142	0.8000	0.6000	0.0199	0.0278
16	0.0075	0.0047	1.0000	0.3333	0.0150	0.0093
17	0.0050	0.0047	1.0000	0.5000	0.0100	0.0094
18	0.0025	0.0047	1.0000	1.0000	0.0050	0.0094
19	0.0101	0.0000	0.8000	0.0000	0.0199	0.0000
20	0.0025	0.0047	1.0000	1.0000	0.0050	0.0094
21	0.0050	0.0095	1.0000	1.0000	0.0100	0.0188
22	0.0050	0.0095	1.0000	1.0000	0.0100	0.0188
23	0.0025	0.0047	1.0000	1.0000	0.0050	0.0094
24	0.0025	0.0047	1.0000	1.0000	0.0050	0.0094
25	0.0025	0.0047	1.0000	1.0000	0.0050	0.0094
26	0.0050	0.0142	1.0000	1.5000	0.0100	0.0282
27	0.0075	0.0000	1.0000	0.0000	0.0150	0.0000
28	0.0126	0.0094	1.0000	0.4000	0.0248	0.0184
29	0.0050	0.0047	1.0000	0.5000	0.0100	0.0093

ลำดับข้อ สอบถาม	ค่าความแม่นยำ (Precision)		ค่าความระลึก (Recall)		ค่าเฉลี่ยฮาร์โมนิค (Harmonic mean)	
	รัศมีกลุ่ม (R)	รัศมีกลุ่มหาร สอง (R/2)	รัศมีกลุ่ม (R)	รัศมีกลุ่มหาร สอง (R/2)	รัศมีกลุ่ม (R)	รัศมีกลุ่มหาร สอง (R/2)
30	0.0101	0.0047	0.8000	0.2000	0.0199	0.0092
31	0.0151	0.0236	0.8571	0.7143	0.0296	0.0457
32	0.0025	0.0000	1.0000	0.0000	0.0050	0.0000
33	0.0050	0.0047	1.0000	0.5000	0.0100	0.0093
34	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
35	0.0025	0.0047	1.0000	1.0000	0.0050	0.0094
36	0.0025	0.0047	1.0000	1.0000	0.0050	0.0094
37	0.0050	0.0047	1.0000	0.5000	0.0100	0.0093
38	0.0025	0.0047	1.0000	1.0000	0.0050	0.0094
39	0.0226	0.0189	1.0000	0.4444	0.0442	0.0362
40	0.0226	0.0189	1.0000	0.4444	0.0442	0.0362
41	0.0126	0.0236	0.8333	0.8333	0.0248	0.0459
42	0.0025	0.0000	1.0000	0.0000	0.0050	0.0000
43	0.0050	0.0047	1.0000	0.5000	0.0100	0.0093
44	0.0050	0.0000	1.0000	0.0000	0.0100	0.0000
45	0.0126	0.0142	1.0000	0.6000	0.0248	0.0276
46	0.0402	0.0425	0.8889	0.5000	0.0769	0.0783
47	0.0126	0.0142	0.8333	0.5000	0.0248	0.0275
48	0.0025	0.0000	1.0000	0.0000	0.0050	0.0000
49	0.0176	0.0283	0.8750	0.7500	0.0345	0.0545
50	0.0025	0.0047	1.0000	1.0000	0.0050	0.0094
51	0.0075	0.0142	1.0000	1.0000	0.0150	0.0279
52	0.0050	0.0094	1.0000	1.0000	0.0100	0.0187
53	0.0050	0.0000	1.0000	0.0000	0.0100	0.0000
54	0.0050	0.0094	1.0000	1.0000	0.0100	0.0187

ลำดับข้อ สอบถาม	ค่าความแม่นยำ (Precision)		ค่าความระลึก (Recall)		ค่าเฉลี่ยฮาร์โมนิค (Harmonic mean)	
	รัศมีกลุ่ม (R)	รัศมีกลุ่มหาร สอง (R/2)	รัศมีกลุ่ม (R)	รัศมีกลุ่มหาร สอง (R/2)	รัศมีกลุ่ม (R)	รัศมีกลุ่มหาร สอง (R/2)
55	0.0276	0.0283	0.9167	0.5000	0.0537	0.0536
56	0.0025	0.0047	1.0000	1.0000	0.0050	0.0094
57	0.0050	0.0000	1.0000	0.0000	0.0100	0.0000
58	0.0176	0.0236	0.8750	0.6250	0.0345	0.0455
59	0.0050	0.0000	1.0000	0.0000	0.0100	0.0000
60	0.0050	0.0047	1.0000	0.5000	0.0100	0.0093
61	0.0327	0.0377	0.8667	0.5333	0.0630	0.0705
62	0.0025	0.0000	1.0000	0.0000	0.0050	0.0000
63	0.0226	0.0236	0.8182	0.4545	0.0440	0.0448
64	0.0050	0.0047	1.0000	0.5000	0.0100	0.0093
65	0.0025	0.0047	1.0000	1.0000	0.0050	0.0094
66	0.0050	0.0000	1.0000	0.0000	0.0100	0.0000
67	0.0050	0.0094	0.6667	0.6667	0.0100	0.0186
68	0.0176	0.0236	0.8750	0.6250	0.0345	0.0455
69	0.0276	0.0283	0.8462	0.4615	0.0535	0.0533
70	0.0025	0.0000	1.0000	0.0000	0.0050	0.0000
71	0.0075	0.0047	1.0000	0.3333	0.0150	0.0093
72	0.0025	0.0000	1.0000	0.0000	0.0050	0.0000
73	0.0025	0.0047	1.0000	1.0000	0.0050	0.0094
74	0.0050	0.0047	1.0000	0.5000	0.0100	0.0093
75	0.0025	0.0000	1.0000	0.0000	0.0050	0.0000
76	0.0126	0.0094	1.0000	0.4000	0.0248	0.0184
77	0.0025	0.0047	1.0000	1.0000	0.0050	0.0094
78	0.0050	0.0047	1.0000	0.5000	0.0100	0.0093
79	0.0025	0.0047	1.0000	1.0000	0.0050	0.0094

ลำดับข้อ สอบถาม	ค่าความแม่นยำ (Precision)		ค่าความระลึก (Recall)		ค่าเฉลี่ยฮาร์โมนิค (Harmonic mean)	
	รัศมีกลุ่ม (R)	รัศมีกลุ่มหาร สอง (R/2)	รัศมีกลุ่ม (R)	รัศมีกลุ่มหาร สอง (R/2)	รัศมีกลุ่ม (R)	รัศมีกลุ่มหาร สอง (R/2)
80	0.0377	0.0283	0.8824	0.3529	0.0723	0.0524
81	0.0050	0.0094	1.0000	1.0000	0.0100	0.0187
82	0.0101	0.0094	0.8000	0.4000	0.0199	0.0184
83	0.0025	0.0047	0.5000	0.5000	0.0050	0.0093

ตารางที่ ๑.2 ตารางแสดงค่าความแม่นยำ (Precision), ค่าความระลึก (Recall) และค่าเฉลี่ยฮาร์โมนิค (Harmonic mean) เครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 1* เมื่อกรอค่าความคล้ายคลึงที่ต่างกัน

ลำดับข้อ สอบถาม	ค่าความแม่นยำ (Precision)		ค่าความระลึก (Recall)		ค่าเฉลี่ยฮาร์โมนิค (Harmonic mean)	
	รัศมีกลุ่ม (R)	รัศมีกลุ่มหาร สอง (R/2)	รัศมีกลุ่ม (R)	รัศมีกลุ่มหาร สอง (R/2)	รัศมีกลุ่ม (R)	รัศมีกลุ่มหาร สอง (R/2)
1	0.0385	0.0000	0.5714	0.0000	0.0721	0.0000
2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	0.0103	0.0221	0.7500	0.7500	0.0203	0.0429
4	0.0255	0.0423	0.8000	0.6000	0.0494	0.0789
5	0.0259	0.0417	0.6000	0.4000	0.0496	0.0755
6	0.0272	0.0000	0.5556	0.0000	0.0518	0.0000
7	0.0072	0.0076	1.0000	0.5000	0.0144	0.0149
8	0.0109	0.0132	1.0000	0.5000	0.0215	0.0256
9	0.0648	0.1111	0.8750	0.6250	0.1207	0.1887
10	0.0413	0.0678	0.8333	0.6667	0.0787	0.1231
11	0.0066	0.0072	1.0000	0.5000	0.0130	0.0143
12	0.0451	0.0615	0.8571	0.5714	0.0857	0.1111
13	0.0084	0.0000	1.0000	0.0000	0.0167	0.0000

ลำดับข้อ สอบถาม	ค่าความแม่นยำ (Precision)		ค่าความระลึก (Recall)		ค่าเฉลี่ยฮาร์โมนิก (Harmonic mean)	
	รัศมีกลุ่ม (R)	รัศมีกลุ่มหาร สอง (R/2)	รัศมีกลุ่ม (R)	รัศมีกลุ่มหาร สอง (R/2)	รัศมีกลุ่ม (R)	รัศมีกลุ่มหาร สอง (R/2)
14	0.0031	0.0068	1.0000	1.0000	0.0062	0.0135
15	0.0173	0.0270	0.8000	0.6000	0.0339	0.0517
16	0.0101	0.0069	1.0000	0.3333	0.0200	0.0136
17	0.0230	0.0208	1.0000	0.5000	0.0449	0.0400
18	0.0031	0.0064	1.0000	1.0000	0.0063	0.0127
19	0.0129	0.0000	0.8000	0.0000	0.0253	0.0000
20	0.0033	0.0070	1.0000	1.0000	0.0065	0.0140
21	0.0073	0.0155	1.0000	1.0000	0.0144	0.0305
22	0.0105	0.0260	1.0000	1.0000	0.0207	0.0506
23	0.0035	0.0078	1.0000	1.0000	0.0070	0.0154
24	0.0041	0.0088	1.0000	1.0000	0.0082	0.0175
25	0.0115	0.0286	1.0000	1.0000	0.0227	0.0556
26	0.0082	0.0179	1.0000	1.0000	0.0162	0.0351
27	0.0105	0.0000	1.0000	0.0000	0.0208	0.0000
28	0.0187	0.0159	1.0000	0.4000	0.0368	0.0305
29	0.0061	0.0062	1.0000	0.5000	0.0121	0.0123
30	0.0131	0.0070	0.8000	0.2000	0.0258	0.0135
31	0.0234	0.0420	0.8571	0.7143	0.0456	0.0794
32	0.0030	0.0000	1.0000	0.0000	0.0060	0.0000
33	0.0079	0.0088	1.0000	0.5000	0.0156	0.0172
34	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
35	0.0053	0.0120	1.0000	1.0000	0.0106	0.0238
36	0.0045	0.0108	1.0000	1.0000	0.0090	0.0213
37	0.0085	0.0088	1.0000	0.5000	0.0169	0.0172
38	0.0048	0.0115	1.0000	1.0000	0.0095	0.0227

ลำดับข้อ สอบถาม	ค่าความแม่นยำ (Precision)		ค่าความระลึก (Recall)		ค่าเฉลี่ยฮาร์โมนิค (Harmonic mean)	
	รัศมีกลุ่ม (R)	รัศมีกลุ่มหาร สอง (R/2)	รัศมีกลุ่ม (R)	รัศมีกลุ่มหาร สอง (R/2)	รัศมีกลุ่ม (R)	รัศมีกลุ่มหาร สอง (R/2)
39	0.0300	0.0296	1.0000	0.4444	0.0583	0.0556
40	0.0301	0.0278	1.0000	0.4444	0.0584	0.0523
41	0.0161	0.0331	0.8333	0.8333	0.0316	0.0637
42	0.0034	0.0000	1.0000	0.0000	0.0067	0.0000
43	0.0187	0.0185	1.0000	0.5000	0.0367	0.0357
44	0.0096	0.0000	1.0000	0.0000	0.0190	0.0000
45	0.0223	0.0275	1.0000	0.6000	0.0437	0.0526
46	0.0588	0.0726	0.8889	0.5000	0.1103	0.1268
47	0.0350	0.0508	0.8333	0.5000	0.0671	0.0923
48	0.0087	0.0000	1.0000	0.0000	0.0172	0.0000
49	0.0526	0.1053	0.8750	0.7500	0.0993	0.1846
50	0.0032	0.0065	1.0000	1.0000	0.0064	0.0130
51	0.0195	0.0429	1.0000	1.0000	0.0382	0.0822
52	0.0062	0.0123	1.0000	1.0000	0.0122	0.0244
53	0.0068	0.0000	1.0000	0.0000	0.0135	0.0000
54	0.0067	0.0137	1.0000	1.0000	0.0134	0.0270
55	0.0379	0.0435	0.9167	0.5000	0.0728	0.0800
56	0.0038	0.0078	1.0000	1.0000	0.0075	0.0154
57	0.0400	0.0000	1.0000	0.0000	0.0769	0.0000
58	0.0303	0.0463	0.8750	0.6250	0.0586	0.0862
59	0.0100	0.0000	1.0000	0.0000	0.0197	0.0000
60	0.0230	0.0286	1.0000	0.5000	0.0449	0.0541
61	0.0492	0.0656	0.8667	0.5333	0.0932	0.1168
62	0.0064	0.0000	1.0000	0.0000	0.0127	0.0000
63	0.0251	0.0279	0.8182	0.4545	0.0488	0.0526

ลำดับข้อ สอบถาม	ค่าความแม่นยำ (Precision)		ค่าความระลึก (Recall)		ค่าเฉลี่ยฮาร์โมนิค (Harmonic mean)	
	รัศมีกลุ่ม (R)	รัศมีกลุ่มหาร สอง (R/2)	รัศมีกลุ่ม (R)	รัศมีกลุ่มหาร สอง (R/2)	รัศมีกลุ่ม (R)	รัศมีกลุ่มหาร สอง (R/2)
64	0.0051	0.0049	1.0000	0.5000	0.0102	0.0097
65	0.0119	0.0286	1.0000	1.0000	0.0235	0.0556
66	0.0112	0.0000	1.0000	0.0000	0.0221	0.0000
67	0.0202	0.0741	0.6667	0.6667	0.0392	0.1333
68	0.0337	0.0510	0.8750	0.6250	0.0648	0.0943
69	0.0358	0.0403	0.8462	0.4615	0.0688	0.0741
70	0.0031	0.0000	1.0000	0.0000	0.0062	0.0000
71	0.0126	0.0088	1.0000	0.3333	0.0248	0.0171
72	0.0032	0.0000	1.0000	0.0000	0.0065	0.0000
73	0.0037	0.0081	1.0000	1.0000	0.0074	0.0160
74	0.0154	0.0161	1.0000	0.5000	0.0303	0.0313
75	0.0032	0.0000	1.0000	0.0000	0.0064	0.0000
76	0.0151	0.0127	1.0000	0.4000	0.0298	0.0247
77	0.0062	0.0132	1.0000	1.0000	0.0123	0.0260
78	0.0081	0.0089	1.0000	0.5000	0.0160	0.0175
79	0.0031	0.0065	1.0000	1.0000	0.0063	0.0130
80	0.0497	0.0408	0.8824	0.3529	0.0940	0.0732
81	0.0068	0.0139	1.0000	1.0000	0.0136	0.0274
82	0.0138	0.0141	0.8000	0.4000	0.0271	0.0272
83	0.0033	0.0066	0.5000	0.5000	0.0066	0.0131

ตารางที่ ๑.3 ตารางแสดงค่าความแม่นยำ (Precision), ค่าความระลึก (Recall) และค่าเฉลี่ยฮาร์โมนิก (Harmonic mean) เครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 3 เมื่อรอบค่าความคล้ายคลึงที่ต่างกัน

ลำดับข้อ สอบถาม	ค่าความแม่นยำ (Precision)		ค่าความระลึก (Recall)		ค่าเฉลี่ยฮาร์โมนิก (Harmonic mean)	
	รัศมีความคล้าย (RS)	รัศมีความคล้าย หารสอง (RS/2)	รัศมีความคล้าย (RS)	รัศมีความคล้าย หารสอง (RS/2)	รัศมีความคล้าย (RS)	รัศมีความคล้าย หารสอง (RS/2)
1	0.3182	0.2059	1.0000	1.0000	0.4828	0.3415
2	0.0800	0.0488	1.0000	1.0000	0.1481	0.0930
3	0.1034	0.0606	0.7500	1.0000	0.1818	0.1143
4	0.1389	0.1000	1.0000	1.0000	0.2439	0.1818
5	0.1563	0.1190	1.0000	1.0000	0.2703	0.2128
6	0.2903	0.1957	1.0000	1.0000	0.4500	0.3273
7	0.1667	0.0870	1.0000	1.0000	0.2857	0.1600
8	0.2222	0.2000	1.0000	1.0000	0.3636	0.3333
9	0.7500	0.7000	0.7500	0.8750	0.7500	0.7778
10	0.8333	0.6667	0.8333	1.0000	0.8333	0.8000
11	1.0000	0.2857	0.5000	1.0000	0.6667	0.4444
12	0.7778	0.6364	1.0000	1.0000	0.8750	0.7778
13	0.4286	0.2727	1.0000	1.0000	0.6000	0.4286
14	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000
15	0.6667	0.2667	0.8000	0.8000	0.7273	0.4000
16	0.4000	0.3000	0.6667	1.0000	0.5000	0.4615
17	0.2857	0.1667	1.0000	1.0000	0.4444	0.2857
18	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000
19	0.3846	0.2632	1.0000	1.0000	0.5556	0.4167
20	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
21	0.1538	0.0645	1.0000	1.0000	0.2667	0.1212
22	0.0000	0.2500	0.0000	1.0000	0.0000	0.4000

ลำดับข้อ สอบถาม	ค่าความแม่นยำ (Precision)		ค่าความระลึก (Recall)		ค่าเฉลี่ยฮาร์โมนิค (Harmonic mean)	
	รัศมีความ คล้าย (RS)	รัศมีความคล้าย หารสอง (RS/2)	รัศมีความ คล้าย (RS)	รัศมีความคล้าย หารสอง (RS/2)	รัศมีความ คล้าย (RS)	รัศมีความคล้าย หารสอง (RS/2)
23	1.0000	0.1250	1.0000	1.0000	1.0000	0.2222
24	0.1429	0.0435	1.0000	1.0000	0.2500	0.0833
25	0.0714	0.0417	1.0000	1.0000	0.1333	0.0800
26	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
27	0.0000	0.2500	0.0000	1.0000	0.0000	0.4000
28	0.3333	0.1579	0.2000	0.6000	0.2500	0.2500
29	0.4000	0.2857	1.0000	1.0000	0.5714	0.4444
30	0.4000	0.2083	0.8000	1.0000	0.5333	0.3448
31	0.0000	0.3000	0.0000	0.4286	0.0000	0.3529
32	0.0556	0.0172	1.0000	1.0000	0.1053	0.0339
33	1.0000	0.2857	1.0000	1.0000	1.0000	0.4444
34	1.0000	0.5000	1.0000	1.0000	1.0000	0.6667
35	0.2500	0.1000	1.0000	1.0000	0.4000	0.1818
36	1.0000	0.3333	1.0000	1.0000	1.0000	0.5000
37	0.5000	0.1000	0.5000	0.5000	0.5000	0.1667
38	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
39	0.7500	0.4286	1.0000	1.0000	0.8571	0.6000
40	0.7500	0.5000	0.3333	1.0000	0.4615	0.6667
41	0.6000	0.2273	0.5000	0.8333	0.5455	0.3571
42	0.0000	0.0435	0.0000	1.0000	0.0000	0.0833
43	0.3333	0.0870	1.0000	1.0000	0.5000	0.1600
44	0.3333	0.1429	1.0000	1.0000	0.5000	0.2500
45	0.0000	0.1429	0.0000	0.2000	0.0000	0.1667
46	0.7333	0.5000	0.6111	0.8889	0.6667	0.6400
47	0.5000	0.2667	0.1667	0.6667	0.2500	0.3810

ลำดับข้อ สอบถาม	ค่าความแม่นยำ (Precision)		ค่าความระลึก (Recall)		ค่าเฉลี่ยฮาร์โมนิก (Harmonic mean)	
	รัศมีความ คล้าย (RS)	รัศมีความคล้าย หารสอง (RS/2)	รัศมีความ คล้าย (RS)	รัศมีความคล้าย หารสอง (RS/2)	รัศมีความ คล้าย (RS)	รัศมีความคล้าย หารสอง (RS/2)
48	0.0000	0.0714	0.0000	1.0000	0.0000	0.1333
49	0.8571	0.4444	0.7500	1.0000	0.8000	0.6154
50	0.3333	0.0909	1.0000	1.0000	0.5000	0.1667
51	0.5000	0.1579	0.3333	1.0000	0.4000	0.2727
52	1.0000	0.0714	0.5000	0.5000	0.6667	0.1250
53	0.5000	0.0909	1.0000	1.0000	0.6667	0.1667
54	0.1111	0.0351	1.0000	1.0000	0.2000	0.0678
55	0.6667	0.3704	0.6667	0.8333	0.6667	0.5128
56	0.2000	0.0400	1.0000	1.0000	0.3333	0.0769
57	0.1429	0.1176	0.5000	1.0000	0.2222	0.2105
58	0.5455	0.3333	0.7500	0.8750	0.6316	0.4828
59	0.4000	0.0870	1.0000	1.0000	0.5714	0.1600
60	0.2857	0.0952	1.0000	1.0000	0.4444	0.1739
61	0.6923	0.7059	0.6000	0.8000	0.6429	0.7500
62	0.3333	0.0909	1.0000	1.0000	0.5000	0.1667
63	0.5238	0.2558	1.0000	1.0000	0.6875	0.4074
64	0.2857	0.2222	1.0000	1.0000	0.4444	0.3636
65	0.0667	0.0435	1.0000	1.0000	0.1250	0.0833
66	0.1250	0.0645	1.0000	1.0000	0.2222	0.1212
67	0.4286	0.2143	1.0000	1.0000	0.6000	0.3529
68	0.3333	0.2143	0.5000	0.7500	0.4000	0.3333
69	0.9286	0.5652	1.0000	1.0000	0.9630	0.7222
70	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000
71	0.5000	0.5000	0.3333	0.6667	0.4000	0.5714
72	0.0909	0.0303	1.0000	1.0000	0.1667	0.0588

ลำดับข้อ สอบถาม	ค่าความแม่นยำ (Precision)		ค่าความระลึก (Recall)		ค่าเฉลี่ยฮาร์โมนิค (Harmonic mean)	
	รัศมีความ คล้าย (RS)	รัศมีความคล้าย หารสอง (RS/2)	รัศมีความ คล้าย (RS)	รัศมีความคล้าย หารสอง (RS/2)	รัศมีความ คล้าย (RS)	รัศมีความคล้าย หารสอง (RS/2)
73	0.2500	0.2000	1.0000	1.0000	0.4000	0.3333
74	0.1429	0.1000	1.0000	1.0000	0.2500	0.1818
75	0.0000	0.1429	0.0000	1.0000	0.0000	0.2500
76	0.3636	0.1667	0.8000	1.0000	0.5000	0.2857
77	0.0000	0.0400	0.0000	1.0000	0.0000	0.0769
78	0.1818	0.0800	1.0000	1.0000	0.3077	0.1481
79	0.3333	0.2000	1.0000	1.0000	0.5000	0.3333
80	0.5714	0.3636	0.2353	0.4706	0.3333	0.4103
81	0.0667	0.0172	0.5000	0.5000	0.1176	0.0333
82	0.5000	0.2083	0.6000	1.0000	0.5455	0.3448
83	0.2000	0.0556	0.5000	1.0000	0.2857	0.1053

ภาคผนวก จ
สรุปผลคำเฉลี่ยที่ปรากฏในแต่ละกลุ่ม และ
ผลการทดลองกำหนดกลุ่มให้กับข้อสอบถาม

ผลการจัดกลุ่มเอกสารนิตยสารไหม้ ด้วยการกำหนดจำนวนกลุ่มที่ต้องการแบ่งเท่ากับ 20 กลุ่ม เมื่อพิจารณาจากจำนวนคำเฉลี่ยที่ปรากฏของเอกสารในแต่ละกลุ่ม แสดงได้ดังในตารางที่ จ. 1 ซึ่งจากตารางที่ จ.1 พบว่ากลุ่มเอกสารที่มีจำนวนเอกสารในกลุ่มมากที่สุด หรือกลุ่มใหญ่ที่สุด (กลุ่มที่ 11) จำนวนคำเฉลี่ยที่ปรากฏในกลุ่มมีจำนวนน้อยที่สุด

ตารางที่ จ.1 ตารางแสดงจำนวนคำเฉลี่ยและค่าเบี่ยงเบนมาตรฐานที่ปรากฏในแต่ละกลุ่ม

กลุ่ม	จำนวนเอกสาร	จำนวนคำเฉลี่ย	ค่าเบี่ยงเบนมาตรฐาน
11	306	172.44	79.37
17	53	286.74	78.24
15	21	225.81	76.05
13	10	321.6	126.50
7	9	180	87.08
4	6	324.5	97.06
20	3	574.67	43.93
2	2	352.5	16.50
6	2	579.5	151.50
12	2	557	156
14	2	406	3
1	1	1392	0
3	1	1342	0
5	1	1099	0
8	1	400	0
9	1	289	0
10	1	1298	0
16	1	318	0

กลุ่ม	จำนวนเอกสาร	จำนวนคำเฉลี่ย	ค่าเบี่ยงเบนมาตรฐาน
18	1	616	0
19	1	965	0

ผลแสดงการกำหนดกลุ่มให้กับข้อสอบถาม แสดงได้ดังตารางที่ ๑.2 ซึ่งได้พิจารณาจากเอกสารที่มีความคล้ายคลึงกับข้อสอบถามมากที่สุด ดังนั้นข้อสอบถามจะถูกกำหนดให้อยู่ในกลุ่มเอกสารนั้น และจากตารางที่ ๑.2 จะพบว่าข้อสอบถามโดยส่วนใหญ่จะถูกจัดให้อยู่ในกลุ่มเอกสารที่ใหญ่ที่สุด (กลุ่มที่ 11) คิดเป็น 98.8% ของข้อสอบถามทั้งหมด

ตารางที่ ๑.2 ตารางแสดงผลการจัดกลุ่มให้กับทุกข้อสอบถาม

กลุ่ม	จำนวนข้อ สอบถาม	เปอร์เซ็นต์ ข้อสอบถาม
11	82	98.8
17	0	0
15	0	0
13	0	0
7	1	1.2
4	0	0
20	0	0
2	0	0
6	0	0
12	0	0

กลุ่ม	จำนวนข้อ สอบถาม	เปอร์เซ็นต์ ข้อสอบถาม
14	0	0
1	0	0
3	0	0
5	0	0
8	0	0
9	0	0
10	0	0
16	0	0
18	0	0
19	0	0

ภาคผนวก ช
การวิเคราะห์ผลการทดลองทางสถิติของเครื่องมือทดสอบ
การค้นคืนเอกสารรูปแบบที่ 1 และรูปแบบที่ 1*

ในการเปรียบเทียบประสิทธิภาพของเครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 1 กับเครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 1* ผู้วิจัยสนใจตัวแปร คือค่าประสิทธิภาพของเทคนิคการค้นคืนเอกสารทั้ง 2 รูปแบบ ที่วัดด้วยค่าความแม่นยำ (Precision), ค่าความระลึก (Recall) และค่าเฉลี่ยฮาร์โมนิก (Harmonic mean) ของการค้นคืนเอกสารที่ทดสอบโดยใช้ข้อสอบถามจำนวน 83 ข้อสอบถาม การตรวจสอบเงื่อนไขพื้นฐานขั้นแรกต้องตรวจสอบการแจกแจงของค่าประสิทธิภาพที่ได้มาว่า มีการแจกแจงปกติหรือไม่ เพื่อเลือกทางเลือกในการทดสอบสมมติฐานว่าจะใช้วิธีการทดสอบสมมติฐานแบบอิงพารามิเตอร์ (Parametric Test) หรือแบบไม่อิงกับพารามิเตอร์ (Non Parametric Test) การทดสอบทางสถิติ แสดงได้ดังต่อไปนี้

1) การวิเคราะห์ข้อมูลค่าความแม่นยำ (Precision)

การวิเคราะห์การแจกแจงข้อมูล

ผู้วิจัยจะตรวจสอบว่าค่าความแม่นยำ (Precision) ของการค้นคืนเอกสารมีการแจกแจงแบบปกติหรือไม่ จากค่าความแม่นยำ (Precision) 83 ค่าที่ได้ในการค้นคืนเอกสารแต่ละรูปแบบ โดยตั้งสมมติฐานของการทดสอบสำหรับทดสอบค่าตัวแปรค่าความแม่นยำ (Precision) ของการค้นคืนเอกสารแต่ละรูปแบบ ดังนี้

1) ทดสอบการแจกแจงค่าความแม่นยำของการค้นคืนเอกสารรูปแบบที่ 1

H_0 : ข้อมูลค่าความแม่นยำของการค้นคืนเอกสารรูปแบบที่ 1 มีการแจกแจงแบบปกติ

H_1 : ข้อมูลค่าความแม่นยำของการค้นคืนเอกสารรูปแบบที่ 1 ไม่ได้มีการแจกแจงแบบปกติ

2) ทดสอบการแจกแจงค่าความแม่นยำของการค้นคืนเอกสารรูปแบบที่ 1*

H_0 : ข้อมูลค่าความแม่นยำของการค้นคืนเอกสารรูปแบบที่ 1* มีการแจกแจงแบบปกติ

H_1 : ข้อมูลค่าความแม่นยำของการค้นคืนเอกสารรูปแบบที่ 1* ไม่ได้มีการแจกแจงแบบปกติ

การตรวจสอบการแจกแจงของข้อมูลเชิงปริมาณมีสถิติทดสอบที่ใช้คือ Kolmogorov-Smirnov เมื่อขนาดตัวอย่างมากกว่า 50 หน่วย และของ Shapiro-Wilk เมื่อขนาดตัวอย่างน้อยกว่า 50 หน่วย (กัลยา วานิชย์บัญชา, 2546) สำหรับในงานวิจัยนี้ตัวอย่างในแต่ละกลุ่มมีขนาดมากกว่า 50 หน่วย จึงใช้วิธีตรวจสอบการแจกแจงข้อมูลเทคนิคของ Kolmogorov-Smirnov โดยจะยอมรับสมมติฐาน H_0 ถ้าค่า Sig. (Significance) มีค่ามากกว่าค่านัยสำคัญ α ซึ่งงานวิจัยนี้กำหนดให้เท่ากับ 0.05 ดังตารางต่อไปนี้

ตารางที่ ข.1 ตารางแสดงค่าสถิติทดสอบการแจกแจงปกติ (Normality Test) ของค่าความแม่นยำ

	เครื่องมือทดสอบการ คั่นคั่นเอกสาร	Kolmogorov-Smirnov		
		Statistic	df	Sig.
ค่าความแม่นยำ	1	0.278	83	0.000
	1*	0.197	83	0.000

ผลการทดสอบในตารางที่ ข.1 ข้างต้นพบว่าค่าสถิติค่า Sig. ของตัวแปรของการคั่นคั่นเอกสารทั้ง 2 รูปแบบเป็นดังนี้

- 1) การคั่นคั่นเอกสารรูปแบบที่ 1 มีค่า Sig. เท่ากับ 0.000 ซึ่งมีค่าน้อยกว่าค่า $\alpha = 0.05$ ดังนั้นจึงปฏิเสธสมมติฐาน H_0
- 2) การคั่นคั่นเอกสารรูปแบบที่ 1* มีค่า Sig. เท่ากับ 0.000 ซึ่งมีค่าน้อยกว่าค่า $\alpha = 0.05$ ดังนั้นจึงปฏิเสธสมมติฐาน H_0

ดังนั้นสรุปได้ว่าค่าความแม่นยำ (Precision) ของการคั่นคั่นเอกสารทั้ง 2 รูปแบบนั้น ไม่ได้มีการแจกแจงแบบปกติ ผู้วิจัยจึงใช้การทดสอบสมมติฐานแบบไม่อิงพารามิเตอร์ (Non Parametric Test) โดยเลือกใช้การทดสอบแบบแมนวิทนีย์ (Mann-Whitney Test)

การวิเคราะห์เปรียบเทียบค่าความแม่นยำ (Precision) ของการคั่นคั่นเอกสารทั้ง 2 รูปแบบ

จากการวิเคราะห์การแจกแจงของข้อมูลเห็นได้ว่า การแจกแจงของตัวแปรค่าความแม่นยำ (Precision) ของการคั่นคั่นเอกสารทั้ง 2 รูปแบบนั้น ไม่ได้มีการแจกแจงแบบปกติ ดังนั้นผู้วิจัยจึงทดสอบสมมติฐานด้วยการทดสอบแบบแมนวิทนีย์ (Mann-Whitney Test) โดยใช้โปรแกรม SPSS กำหนดให้

$\mu_{1,1}$ คือ ค่าเฉลี่ยค่าความแม่นยำของการคั่นคั่นเอกสารรูปแบบที่ 1

μ_{1AA} คือ ค่าเฉลี่ยค่าความแม่นยำของการค้นคืนเอกสารรูปแบบที่ 1*
วิเคราะห์เปรียบเทียบค่าความแม่นยำของการค้นคืนเอกสารทั้ง 2 รูปแบบว่ามีความแตกต่างกันหรือไม่ โดยตั้งสมมติฐานดังนี้

$$H_0 : \mu_{1A} - \mu_{1AA} = 0$$

$$H_1 : \mu_{1A} - \mu_{1AA} \neq 0$$

ผลการทดลองที่ได้ เมื่อใช้สถิติทดสอบแบบแมนวิทนี (Mann-Whitney Test) สามารถแสดงค่าสถิติทดสอบได้ดังตารางที่ ข.2 และจากสมมติฐานข้างต้นจะปฏิเสธสมมติฐาน H_0 เมื่อค่า Asymp. Sig. (2-tailed) น้อยกว่าระดับค่านัยสำคัญ α (กัลยา วานิชย์บัญชา, 2548)

ตารางที่ ข.2 ตารางแสดงค่าสถิติทดสอบค่าความแม่นยำของการค้นคืนเอกสาร

	ค่าความแม่นยำของการค้นคืนเอกสาร
Mann-Whitney U	2435.500
Z	-3.290
Asymp. Sig. (2-tailed)	0.001

จากตารางที่ ข.2 จะเห็นได้ว่าค่า Asymp. Sig. (2-tailed) มีค่าเท่ากับ 0.001 ซึ่งมีค่าน้อยกว่าค่านัยสำคัญ $\alpha = 0.05$ ดังนั้น จึงปฏิเสธสมมติฐาน H_0 กล่าวคือ ค่าเฉลี่ยค่าความแม่นยำของการค้นคืนเอกสารทั้ง 2 รูปแบบแตกต่างกันอย่างมีนัยสำคัญ

สรุปการวิเคราะห์ข้อมูลประสิทธิภาพของค่าความแม่นยำ (Precision)

ค่าความแม่นยำ (Precision) ของการค้นคืนเอกสารรูปแบบที่ 1 แตกต่างกับรูปแบบที่ 1* อย่างมีนัยสำคัญ และเมื่อพิจารณาจากค่าเฉลี่ย (Mean) ของค่าความแม่นยำ จะพบว่าค่าเฉลี่ยค่าความแม่นยำของการค้นคืนเอกสารรูปแบบที่ 1 น้อยกว่าการค้นคืนเอกสารรูปแบบที่ 1* ดังนั้นค่าความแม่นยำของการค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ด้วยวิธีการวัดความคล้ายคลึงเชิงระยะห่างยูคลิเดียน ด้วยการพิจารณาค้นคืนเฉพาะเอกสารที่มีค่าในข้อสอบถามปรากฏภายในกรอบค่าความคล้ายคลึงที่กำหนดด้วยผลลัพธ์ที่ได้จากเทคนิคการจัดกลุ่มข้อมูล (Clustering) มีค่ามากกว่าค่าความแม่นยำของการค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ด้วยวิธีการวัดความคล้ายคลึงเชิงระยะห่างยูคลิเดียนภายในกรอบค่าความคล้ายคลึงที่กำหนดด้วยผลลัพธ์ที่ได้จากเทคนิคการจัดกลุ่มข้อมูล (Clustering)

2) การวิเคราะห์ข้อมูลค่าความระลึก (Recall)

การวิเคราะห์การแจกแจงข้อมูล

ผู้วิจัยจะตรวจสอบว่าค่าความระลึก (Recall) ของการค้นคืนเอกสารมีการแจกแจงแบบปกติหรือไม่ จากค่าความระลึก (Recall) 83 ค่าที่ได้ในการค้นคืนเอกสารแต่ละรูปแบบ โดยตั้งสมมติฐานของการทดสอบสำหรับทดสอบค่าตัวแปรค่าความระลึก (Recall) ของการค้นคืนเอกสารแต่ละรูปแบบ ดังนี้

1) ทดสอบการแจกแจงค่าความระลึกของการค้นคืนเอกสารรูปแบบที่ 1

H_0 : ข้อมูลค่าความระลึกของการค้นคืนเอกสารรูปแบบที่ 1 มีการแจกแจงแบบปกติ

H_1 : ข้อมูลค่าความระลึกของการค้นคืนเอกสารรูปแบบที่ 1 ไม่ได้มีการแจกแจงแบบปกติ

2) ทดสอบการแจกแจงค่าความระลึกของการค้นคืนเอกสารรูปแบบที่ 1*

H_0 : ข้อมูลค่าความระลึกของการค้นคืนเอกสารรูปแบบที่ 1* มีการแจกแจงแบบปกติ

H_1 : ข้อมูลค่าความระลึกของการค้นคืนเอกสารรูปแบบที่ 1* ไม่ได้มีการแจกแจงแบบปกติ

การตรวจสอบการแจกแจงของข้อมูลเชิงปริมาณมีสถิติทดสอบที่ใช้คือ Kolmogorov-Smirnov เมื่อขนาดตัวอย่างมากกว่า 50 หน่วย และของ Shapiro-Wilk เมื่อขนาดตัวอย่างน้อยกว่า 50 หน่วย (กัลยา วานิชย์บัญชา, 2546) สำหรับในงานวิจัยนี้ตัวอย่างในแต่ละกลุ่มมีขนาดมากกว่า 50 หน่วย จึงใช้วิธีตรวจสอบการแจกแจงข้อมูลเทคนิคของ Kolmogorov-Smirnov โดยจะยอมรับสมมติฐาน H_0 ถ้าค่า Sig. (Significance) มีค่ามากกว่าค่านัยสำคัญ α ซึ่งงานวิจัยนี้กำหนดให้เท่ากับ 0.05 ดังตารางต่อไปนี้

ตารางที่ ข.3 ตารางแสดงค่าสถิติทดสอบการแจกแจงปกติ (Normality Test) ของค่าความระลึก

	เครื่องมือทดสอบการค้นคืนเอกสาร	Kolmogorov-Smirnov		
		Statistic	df	Sig.
ค่าความระลึก	1	0.158	83	0.000
	1*	0.168	83	0.000

ผลการทดสอบในตารางที่ ข.3 ข้างต้นพบว่าค่าสถิติค่า Sig. ของตัวแปรของการค้นคืนเอกสารทั้ง 2 รูปแบบเป็นดังนี้

1) การค้นคืนเอกสารรูปแบบที่ 1 มีค่า Sig. เท่ากับ 0.000 ซึ่งมีค่าน้อยกว่าค่า $\alpha = 0.05$ ดังนั้นจึงปฏิเสธสมมติฐาน H_0

2) การค้นคืนเอกสารรูปแบบที่ 1* มีค่า Sig. เท่ากับ 0.000 ซึ่งมีค่าน้อยกว่าค่า $\alpha = 0.05$ ดังนั้นจึงปฏิเสธสมมติฐาน H_0

ดังนั้นสรุปได้ว่าค่าความระลึก (Recall) ของการค้นคืนเอกสารทั้ง 2 รูปแบบนั้น ไม่ได้มีการแจกแจงแบบปกติ ผู้วิจัยจึงใช้การทดสอบสมมติฐานแบบไม่อิงพารามิเตอร์ (Non Parametric Test) โดยเลือกใช้การทดสอบแบบแมนวิทนีย์ (Mann-Whitney Test)

การวิเคราะห์เปรียบเทียบค่าความระลึก (Recall) ของการค้นคืนเอกสารทั้ง 2 รูปแบบ

จากการวิเคราะห์การแจกแจงของข้อมูลเห็นได้ว่า การแจกแจงของตัวแปรค่าความระลึก (Recall) ของการค้นคืนเอกสารทั้ง 2 รูปแบบนั้น ไม่ได้มีการแจกแจงแบบปกติ ดังนั้นผู้วิจัยจึงทดสอบสมมติฐานด้วยการทดสอบแบบแมนวิทนีย์ (Mann-Whitney Test)

กำหนดให้

μ_{1B} คือ ค่าเฉลี่ยค่าความแม่นยำของการค้นคืนเอกสารรูปแบบที่ 1

μ_{1BB} คือ ค่าเฉลี่ยค่าความแม่นยำของการค้นคืนเอกสารรูปแบบที่ 1*

วิเคราะห์เปรียบเทียบค่าความระลึกของการค้นคืนเอกสารทั้ง 2 รูปแบบว่ามีความแตกต่างกันหรือไม่ โดยตั้งสมมติฐานดังนี้

$$H_0: \mu_{1B} - \mu_{1BB} = 0$$

$$H_1: \mu_{1B} - \mu_{1BB} \neq 0$$

ผลการทดลองที่ได้ เมื่อใช้สถิติทดสอบแบบแมนวิทนีย์ (Mann-Whitney Test) สามารถแสดงค่าสถิติทดสอบได้ดังตารางที่ ข.4 และจากสมมติฐานข้างต้นจะปฏิเสธสมมติฐาน H_0 เมื่อค่า Asymp. Sig. (2-tailed) น้อยกว่าระดับค่านัยสำคัญ α (กัลยา วานิชย์บัญชา, 2548)

ตารางที่ ข.4 ตารางแสดงค่าสถิติทดสอบค่าความระลึของการคั่นคืนเอกสาร

	ค่าความระลึของการคั่นคืนเอกสาร
Mann-Whitney U	3433.500
Z	-.036
Asymp. Sig. (2-tailed)	0.971

จากตารางที่ ข.4 จะเห็นได้ว่าค่า Asymp. Sig. (2-tailed) มีค่าเท่ากับ 0.971 ซึ่งมีค่ามากกว่าค่านัยสำคัญ $\alpha = 0.05$ ดังนั้น จึงยอมรับสมมติฐาน H_0 กล่าวคือ ค่าเฉลี่ยค่าความระลึของการคั่นคืนเอกสารทั้ง 2 รูปแบบไม่แตกต่างกันอย่างมีนัยสำคัญ

สรุปการวิเคราะห์ข้อมูลประสิทธิภาพของค่าความระลึ (Recall)

... ค่าความระลึ (Recall) ของการคั่นคืนเอกสารรูปแบบที่ 1 ไม่แตกต่างกับรูปแบบที่ 1* อย่างมีนัยสำคัญ ดังนั้น ค่าความระลึของการคั่นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ด้วยวิธีการวัดความคล้ายคลึงเชิงระยะห่างยูคลิเดียน ด้วยการพิจารณาคั่นคืนเฉพาะเอกสารที่มีค่าในข้อสอบถามปรากฏภายในกรอบค่าความคล้ายคลึงที่กำหนดด้วยผลลัพธ์ที่ได้จากเทคนิคการจัดกลุ่มข้อมูล (Clustering) มีค่าไม่แตกต่างกับค่าความระลึของการคั่นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ด้วยวิธีการวัดความคล้ายคลึงเชิงระยะห่างยูคลิเดียนภายในกรอบค่าความคล้ายคลึงที่กำหนดด้วยผลลัพธ์ที่ได้จากเทคนิคการจัดกลุ่มข้อมูล (Clustering)

3) การวิเคราะห์ข้อมูลค่าเฉลี่ยฮาร์โมนิก (Harmonic mean)

การวิเคราะห์การแจกแจงข้อมูล

ผู้วิจัยจะตรวจสอบว่าค่าเฉลี่ยฮาร์โมนิก (Harmonic mean) ของการคั่นคืนเอกสารมีการแจกแจงแบบปกติหรือไม่ จากค่าเฉลี่ยฮาร์โมนิก (Harmonic mean) 83 ค่าที่ได้ในการคั่นคืนเอกสารแต่ละรูปแบบ โดยตั้งสมมติฐานของการทดสอบสำหรับทดสอบค่าตัวแปรค่าเฉลี่ยฮาร์โมนิก (Harmonic mean) ของการคั่นคืนเอกสารแต่ละรูปแบบ ดังนี้

1) ทดสอบการแจกแจงค่าเฉลี่ยฮาร์โมนิกการคั่นคืนเอกสารรูปแบบที่ 1

H_0 : ข้อมูลค่าเฉลี่ยฮาร์โมนิกของการคั่นคืนเอกสารรูปแบบที่ 1 มีการแจกแจงแบบปกติ

H_1 : ข้อมูลค่าเฉลี่ยฮาร์โมนิกของการคั่นคั้นเอกสารรูปแบบที่ 1 ไม่ได้มีการแจกแจงแบบปกติ

2) ทดสอบการแจกแจงค่าเฉลี่ยฮาร์โมนิกการคั่นคั้นเอกสารรูปแบบที่ 1*

H_0 : ข้อมูลค่าเฉลี่ยฮาร์โมนิกของการคั่นคั้นเอกสารรูปแบบที่ 1* มีการแจกแจงแบบปกติ

H_1 : ข้อมูลค่าเฉลี่ยฮาร์โมนิกของการคั่นคั้นเอกสารรูปแบบที่ 1* ไม่ได้มีการแจกแจงแบบปกติ

การตรวจสอบการแจกแจงของข้อมูลเชิงปริมาณมีสถิติทดสอบที่ใช้คือ Kolmogorov-Smirnov เมื่อขนาดตัวอย่างมากกว่า 50 หน่วย และของ Shapiro-Wilk เมื่อขนาดตัวอย่างน้อยกว่า 50 หน่วย (กัลยา วานิชย์บัญชา, 2546) สำหรับในงานวิจัยนี้ตัวอย่างในแต่ละกลุ่มมีขนาดมากกว่า 50 หน่วย จึงใช้วิธีตรวจสอบการแจกแจงข้อมูลเทคนิคของ Kolmogorov-Smirnov โดยจะยอมรับสมมติฐาน H_0 ถ้าค่า Sig. (Significance) มีค่ามากกว่าค่านัยสำคัญ α ซึ่งงานวิจัยนี้กำหนดให้เท่ากับ 0.05 ดังตารางต่อไปนี้

ตารางที่ ข.5 ตารางแสดงค่าสถิติทดสอบการแจกแจงปกติ (Normality Test) ของค่าเฉลี่ยฮาร์โมนิก

	เครื่องมือทดสอบการคั่นคั้นเอกสาร	Kolmogorov-Smirnov		
		Statistic	df	Sig.
ค่าเฉลี่ยฮาร์โมนิก	1	0.276	83	0.000
	1*	0.188	83	0.000

ผลการทดสอบในตารางที่ ข.5 ช่างค้นพบว่าค่าสถิติค่า Sig. ของตัวแปรของการคั่นคั้นเอกสารทั้ง 2 รูปแบบเป็นดังนี้

1) การคั่นคั้นเอกสารรูปแบบที่ 1 มีค่า Sig. เท่ากับ 0.000 ซึ่งมีค่าน้อยกว่าค่า $\alpha = 0.05$ ดังนั้นจึงปฏิเสธสมมติฐาน H_0

2) การคั่นคั้นเอกสารรูปแบบที่ 1* มีค่า Sig. เท่ากับ 0.000 ซึ่งมีค่าน้อยกว่าค่า $\alpha = 0.05$ ดังนั้นจึงปฏิเสธสมมติฐาน H_0

ดังนั้นสรุปได้ว่าค่าเฉลี่ยฮาร์โมนิก (Harmonic mean) ของการค้นคืนเอกสารทั้ง 2 รูปแบบ นั้น ไม่ได้มีการแจกแจงแบบปกติ ผู้วิจัยจึงใช้การทดสอบสมมติฐานแบบไม่อิงพารามิเตอร์ (Non Parametric Test) โดยเลือกใช้การทดสอบแบบแมนวิทนี (Mann-Whitney Test)

การวิเคราะห์เปรียบเทียบค่าเฉลี่ยฮาร์โมนิก (Harmonic mean) ของการค้นคืนเอกสารทั้ง 2 รูปแบบ

จากการวิเคราะห์การแจกแจงของข้อมูลเห็นได้ว่า การแจกแจงของตัวแปรค่าเฉลี่ยฮาร์โมนิก (Harmonic mean) ของการค้นคืนเอกสารทั้ง 2 รูปแบบนั้น ไม่ได้มีการแจกแจงแบบปกติ ดังนั้นผู้วิจัยจึงทดสอบสมมติฐานด้วยการทดสอบแบบแมนวิทนี (Mann-Whitney Test)

กำหนดให้

μ_{1c} คือ ค่าเฉลี่ยค่าความแม่นยำของการค้นคืนเอกสารรูปแบบที่ 1

μ_{1cc} คือ ค่าเฉลี่ยค่าความแม่นยำของการค้นคืนเอกสารรูปแบบที่ 1*

วิเคราะห์เปรียบเทียบค่าเฉลี่ยฮาร์โมนิกของการค้นคืนเอกสารทั้ง 2 รูปแบบว่ามีความแตกต่างกันหรือไม่ โดยตั้งสมมติฐานดังนี้

$$H_0 : \mu_{1c} - \mu_{1cc} = 0$$

$$H_1 : \mu_{1c} - \mu_{1cc} \neq 0$$

ผลการทดลองที่ได้ เมื่อใช้สถิติทดสอบแบบแมนวิทนี (Mann-Whitney Test) สามารถแสดงค่าสถิติทดสอบได้ดังตารางที่ ๖.6 และจากสมมติฐานข้างต้นจะปฏิเสธสมมติฐาน H_0 เมื่อค่า Asymp. Sig. (2-tailed) น้อยกว่าระดับความนัยสำคัญ α (กัลยา วานิชย์บัญชา, 2548)

ตารางที่ ๖.6 ตารางแสดงค่าสถิติทดสอบค่าเฉลี่ยฮาร์โมนิกของการค้นคืนเอกสาร

	ค่าเฉลี่ยฮาร์โมนิกของการค้นคืนเอกสาร
Mann-Whitney U	2432.500
Z	-3.291
Asymp. Sig. (2-tailed)	0.001

จากตารางที่ ๖.6 จะเห็นได้ว่าค่า Asymp. Sig. (2-tailed) มีค่าเท่ากับ 0.001 ซึ่งมีค่าน้อยกว่าความนัยสำคัญ $\alpha = 0.05$ ดังนั้น จึงปฏิเสธสมมติฐาน H_0 กล่าวคือ ค่าเฉลี่ยของค่าเฉลี่ยฮาร์โมนิกของการค้นคืนเอกสารทั้ง 2 รูปแบบแตกต่างกันอย่างมีนัยสำคัญ

สรุปการวิเคราะห์ข้อมูลประสิทธิภาพของค่าเฉลี่ยฮาร์โมนิก (Harmonic mean)

ค่าเฉลี่ยฮาร์โมนิก (Harmonic mean) ของการค้นคืนเอกสารรูปแบบที่ 1 แตกต่างกับรูปแบบที่ 1* อย่างมีนัยสำคัญ และเมื่อพิจารณาจากค่าเฉลี่ย (Mean) ของค่าเฉลี่ยฮาร์โมนิก จะพบว่าค่าเฉลี่ยของค่าเฉลี่ยฮาร์โมนิกของการค้นคืนเอกสารรูปแบบที่ 1 น้อยกว่าการค้นคืนเอกสารรูปแบบที่ 1* ดังนั้น ค่าเฉลี่ยฮาร์โมนิกของการค้นคืนเอกสารด้วยวิธีการวัดความคล้ายคลึงเชิงระยะห่างยูคลิดีเยน ด้วยการพิจารณาค้นคืนเฉพาะเอกสารที่มีค่าในข้อสอบถามปรากฏ มีค่ามากกว่าค่าเฉลี่ยฮาร์โมนิก ของการค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ด้วยวิธีการวัดความคล้ายคลึงเชิงระยะห่างยูคลิดีเยน

ประวัติผู้เขียนวิทยานิพนธ์

นางสาวสุนันทา เปี่ยมพริ้ง เกิดวันที่ 27 ธันวาคม พ.ศ. 2525 ที่จังหวัดตรัง สำเร็จ
การศึกษาปริญญาตรีวิศวกรรมศาสตรบัณฑิต ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะ
วิศวกรรมศาสตร์ มหาวิทยาลัยสงขลานครินทร์ ในปีการศึกษา 2547 และเข้าศึกษาต่อในหลักสูตร
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาการพัฒนาระบบคอมพิวเตอร์ด้านธุรกิจ ภาควิชาสถิติ คณะ
พาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2548