

(12) 특허협력조약에 의하여 공개된 국제출원

(19) 세계지식재산권기구
국제사무국



(43) 국제공개일
2012년 3월 8일 (08.03.2012)

PCT

(10) 국제공개번호
WO 2012/030053 A3

- (51) 국제특허분류:
G06F 17/26 (2006.01) G06F 17/27 (2006.01)
- (21) 국제출원번호: PCT/KR2011/003832
- (22) 국제출원일: 2011년 5월 25일 (25.05.2011)
- (25) 출원언어: 한국어
- (26) 공개언어: 한국어
- (30) 우선권정보:
10-2010-0085959 2010년 9월 2일 (02.09.2010) KR
- (71) 출원인 (US 을(를) 제외한 모든 지정국에 대하여): 에스케이텔레콤 주식회사 (SK TELECOM. CO., LTD) [KR/KR]; 서울특별시 중구 을지로 2가 11번지, 100-999 Seoul (KR).
- (72) 발명자: 김
- (75) 발명자/출원인 (US 에 한하여): 김상범 (KIM, Sang-Bum) [KR/KR]; 서울특별시 노원구 상계 1동 현대 2차아파트 202동 1207호, 139-893 Seoul (KR). 윤창호 (YIN, Chang Hao) [CN/KR]; 서울특별시 성북구 정릉 2동 푸른동아 아파트 107동 1303호, 136-778 Seoul (KR). 황영숙 (HWANG, Young Sook) [KR/KR]; 서울특별시 성북구 돈암 1동 풍림아파트 105-502,

136-061 Seoul (KR). 임해창 (RIM, Hae Chang) [KR/KR]; 서울특별시 성북구 안암동 5가 고려대학교 자연계캠퍼스 아산이학관 206호, 136-713 Seoul (KR). 이형규 (LEE, Hyoung Gyu) [KR/KR]; 서울특별시 성북구 안암동 5가 고려대학교자연계캠퍼스 아산이학관 237호, 136-713 Seoul (KR).

(74) 대리인: 전철용 (JUN, Chul-Yong) 등; 서울특별시 중구 충무로 3가 60-1번지 극동빌딩 14층, 100-705 Seoul (KR).

(81) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 국내 권리의 보호를 위하여): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

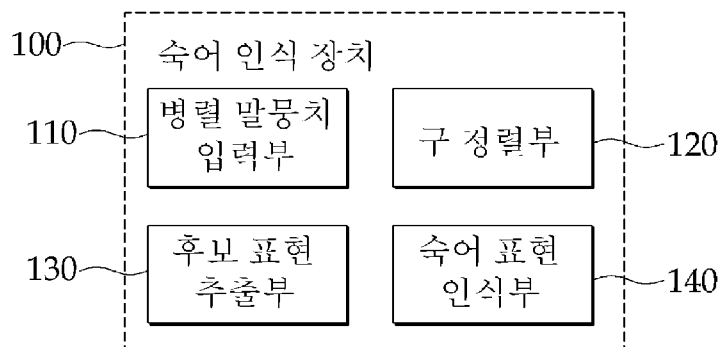
(84) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 역내 권리의 보호를 위하여): ARIPO (BW, GH, GM,

[다음 쪽 계속]

(54) Title: APPARATUS AND METHOD FOR RECOGNIZING AN IDIOMATIC EXPRESSION USING PHRASE ALIGNMENT OF A PARALLEL CORPUS

(54) 발명의 명칭 : 병렬 말뭉치의 구 정렬을 이용한 숙어 표현 인식 장치 및 그 방법

[Fig. 1]



100 ... Idiom recognition apparatus
110 ... Parallel corpus input unit
120 ... Phrase alignment unit
130 ... Candidate expression extraction unit
140 ... Idiomatic expression recognition unit

(57) Abstract: The present invention relates to an apparatus and method for recognizing an idiomatic expression using phrase alignment of a parallel corpus, and more particularly, to an apparatus and method extracting an idiom candidate expression using phrase alignment information of a parallel corpus and measuring an idiomatic expression index for each candidate idiomatic expression in order to recognize an idiomatic expression, thereby correcting errors in the measurement of translation entropy and in the extraction of a representative target word, as well as enhancing the accuracy of recognizing an idiomatic expression.

(57) 요약서: 본 발명은 병렬 말뭉치의 구 정렬을 이용한 숙어 표현 인식 장치 및 그 방법에 관한 것으로서, 상세하게는 병렬 말뭉치의 구 정렬 정보를 이용하여 숙어 후보 표현을 추출하고, 그 추출된 숙어 후보 표현마다 숙어 표현 지수를 측정하여 숙어 표현으로 인식함으로써, 단어의 번역 엔트로피 측정 및 단어의 대표 대역어 추출의 오류를 해결하고 숙어 표현 인식의 정확률을 향상시킬 수 있다.

WO 2012/030053 A3



KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), 유라시아 (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), 유럽 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

공개:

- 국제조사보고서와 함께 (조약 제 21 조(3))
- 청구범위 보정 기한 만료 전의 공개이며, 보정서를 접수하는 경우 그에 관하여 별도 공개함 (규칙 48.2(h))

(88) 국제조사보고서 공개일:

2012 년 4 월 19 일

INTERNATIONAL SEARCH REPORT

International application No.

PCT/KR2011/003832**A. CLASSIFICATION OF SUBJECT MATTER****G06F 17/26(2006.01)i, G06F 17/27(2006.01)i**

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F 17/26; G06F 17/28

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Korean Utility models and applications for Utility models: IPC as above

Japanese Utility models and applications for Utility models: IPC as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

eKOMPASS (KIPO internal) & Keywords: alignment, idiom, threshold value

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	KR 10-2001-0027882 A (ELECTRONICS AND TELECOMMUNICATIONS RESEARCH INSTITUTE) 06 April 2001 See abstract and the claims.	1-15
A	KR 10-1999-0047856 A (ELECTRONICS AND TELECOMMUNICATIONS RESEARCH INSTITUTE) 05 July 1999 See abstract and the claims.	1-15
A	KR 10-2003-0094632 A (INTERNATIONAL BUSINESS MACHINES CORPORATION.) 18 December 2003 See abstract and the claims.	1-15



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

06 FEBRUARY 2012 (06.02.2012)

Date of mailing of the international search report

06 FEBRUARY 2012 (06.02.2012)

Name and mailing address of the ISA/KR

Korean Intellectual Property Office
Government Complex-Daejeon, 139 Seonsa-ro, Daejeon 302-701,
Republic of Korea

Facsimile No. 82-42-472-7140

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/KR2011/003832

Patent document cited in search report	Publication date	Patent family member	Publication date
KR 10-2001-0027882 A	06.04.2001	NONE	
KR 10-1999-0047856 A	05.07.1999	NONE	
KR 10-2003-0094632 A	18.12.2003	JP 04-319860 B2	05.06.2009
		JP 2004-070928 A	04.03.2004
		US 2003-0233226 A1	18.12.2003
		US 2008-0077389 A1	27.03.2008
		US 7330810 B2	12.02.2008
		US 7487082 B2	03.02.2009

A. 발명이 속하는 기술분류(국제특허분류(IPC)) <i>G06F 17/26(2006.01)i, G06F 17/27(2006.01)i</i>		
B. 조사된 분야 조사된 최소문헌(국제특허분류를 기재) G06F 17/26; G06F 17/28 조사된 기술분야에 속하는 최소문헌 이외의 문헌 한국등록실용신안공보 및 한국공개실용신안공보: 조사된 최소문헌란에 기재된 IPC 일본등록실용신안공보 및 일본공개실용신안공보: 조사된 최소문헌란에 기재된 IPC 국제조사에 이용된 전산 데이터베이스(데이터베이스의 명칭 및 검색어(해당하는 경우)) eKOMPASS(특허청 내부 검색시스템) & 키워드: 구 정렬,숙어,임제치		
C. 관련 문헌		
카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
A	KR 10-2001-0027882 A (한국전자통신연구원) 2001.04.06 요약 및 청구항 참조.	1-15
A	KR 10-1999-0047856 A (한국전자통신연구원) 1999.07.05 요약 및 청구항 참조.	1-15
A	KR 10-2003-0094632 A (인터넷셔널 비즈니스 머신즈 코퍼레이션) 2003.12.18 요약 및 청구항 참조.	1-15
☐ 추가 문헌이 C(계속)에 기재되어 있습니다. ☒ 대응특허에 관한 별지를 참조하십시오.		
* 인용된 문헌의 특별 카테고리: “A” 특별히 관련이 없는 것으로 보이는 일반적인 기술수준을 정의한 문헌 “E” 국제출원일보다 빠른 출원일 또는 우선일을 가지나 국제출원일 이후에 공개된 선출원 또는 특허 문헌 “L” 우선권 주장에 의문을 제기하는 문헌 또는 다른 인용문헌의 공개일 또는 다른 특별한 이유(이유를 명시)를 밝히기 위하여 인용된 문헌 “O” 구두 개시, 사용, 전시 또는 기타 수단을 언급하고 있는 문헌 “P” 우선일 이후에 공개되었으나 국제출원일 이전에 공개된 문헌 “T” 국제출원일 또는 우선일 후에 공개된 문헌으로, 출원과 상충하지 않으며 발명의 기초가 되는 원리나 이론을 이해하기 위해 인용된 문헌 “X” 특별한 관련이 있는 문헌. 해당 문헌 하나만으로 청구된 발명의 신규성 또는 진보성이 없는 것으로 본다. “Y” 특별한 관련이 있는 문헌. 해당 문헌이 하나 이상의 다른 문헌과 조합하는 경우로 그 조합이 당업자에게 자명한 경우 청구된 발명은 진보성이 없는 것으로 본다. “&” 동일한 대응특허문헌에 속하는 문헌		
국제조사의 실제 완료일 2012년 02월 06일 (06.02.2012)		국제조사보고서 발송일 2012년 02월 06일 (06.02.2012)
ISA/KR의 명칭 및 우편주소 대한민국 특허청 (302-701) 대전광역시 서구 청사로 189, 정부대전청사 팩스 번호 82-42-472-7140		심사관 이종익 전화번호 82-42-481-8373

국제조사보고서에서 인용된 특허문헌	공개일	대응특허문헌	공개일
KR 10-2001-0027882 A	2001.04.06	없음	
KR 10-1999-0047856 A	1999.07.05	없음	
KR 10-2003-0094632 A	2003. 12. 18	JP 04-319860 B2	2009.06.05
		JP 2004-070928 A	2004.03.04
		US 2003-0233226 A1	2003. 12. 18
		US 2008-0077389 A1	2008.03.27
		US 7330810 B2	2008.02.12
		US 7487082 B2	2009.02.03