

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
5 February 2009 (05.02.2009)

PCT

(10) International Publication Number  
**WO 2009/016631 A3**

- (51) International Patent Classification:  
*G06F 17/00* (2006.01)
- (21) International Application Number:  
PCT/IL2008/001051
- (22) International Filing Date:  
31 July 2008 (31.07.2008)
- (25) Filing Language:  
English
- (26) Publication Language:  
English
- (30) Priority Data:  
60/953,209 1 August 2007 (01.08.2007) US
- (71) Applicant (for all designated States except US): **GINGER SOFTWARE, INC.** [US/US]; 405 Waltham Street, #371, Lexington, Massachusetts 02421 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): **KAROV ZANGVIL, Yael** [IL/IL]; 15 Kfar Saba Street, Neve Zedek, 65147 Tel-aviv (IL). **ZANGVIL, Avner** [IL/IL]; 15 Kfar Saba Street, Neve Zedek, 65147 Tel-aviv (IL).
- (74) Agents: **SANFORD T. COLB & CO.** et al.; P.O. Box 2273, 76122 Rehovot (IL).
- (81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM,

AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, 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, KR, 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, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

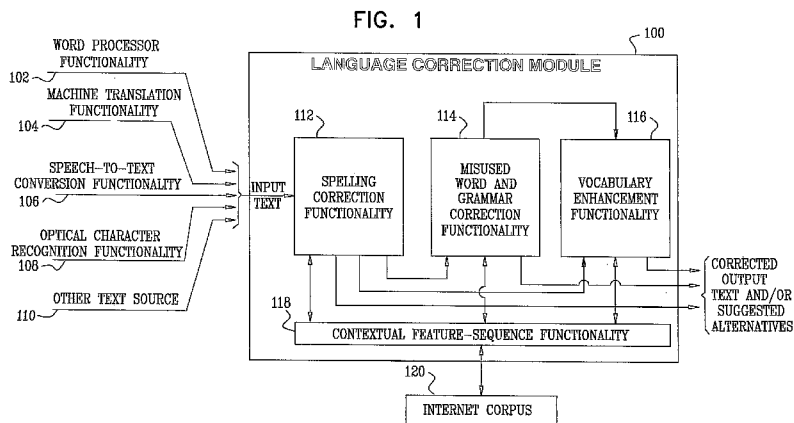
(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— with international search report (Art. 21(3))

(88) Date of publication of the international search report:  
4 March 2010

(54) Title: AUTOMATIC CONTEXT SENSITIVE LANGUAGE CORRECTION AND ENHANCEMENT USING AN INTERNET CORPUS



(57) Abstract: A computer-assisted language correction system including spelling correction functionality, misused word correction functionality, grammar correction functionality and vocabulary enhancement functionality utilizing contextual feature-sequence functionality employing an internet corpus.

WO 2009/016631 A3

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/IL 08/01051

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - G06F 17/00 (2009.01)

USPC - 715/256

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)

USPC: 715/256

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

USPC: 715/200, 255, 256, 257, 262, 265; 707/1, 3, 4, 6, 10; 700/1, 90, 91

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

Electronic Databases Searched: PubWEST (PGPB,USPT,USOC,EPAB,JPAB); GoogleScholar

Search Terms Used: spell, misspell, grammar, syntax, semantic,correct, check, enhance, improve, vocabulary, thesaurus, synonym, adaptive, dislexia,prose, document, text, language, literary, sentence, phrase, paragraph etc.

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2004/0093567 A1 (SCHABES et al.) 13 May 2004 (13.05.2004) Entire document, especially: para [0015]-[0017], [0020], [0022], [0067], [0071], [0110] and Fig. 3	1-5
A	US 2006/0247914 A1 (BRENER et al.) 02 November 2006 (02.11.2006)	1-5
A	US 2005/0257146 A1 (ASHCRAFT et al.) 17 November 2005 (17.11.2005)	1-5
A	BICK, E., "A Constraint Grammar Based Spellchecker for Danish with a Special Focus on Dyslexics," SKY Journal of Linguistics, Vol. 19:2006 (ISSN 1796-279X), pp. 387-396 [retrieved 12 January 2009 (12.01.2009)] Retrieved from the Internet. <URL: http://www.ling.helsinki.fi/sky/julkaisut/SKY2006_1/1.6.1.%20BICK.pdf>	1-5

 Further documents are listed in the continuation of Box C.

\* 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

"&amp;" document member of the same patent family

Date of the actual completion of the international search

13 January 2009 (13.01.2009)

Date of mailing of the international search report

03 FEB 2009

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents

P.O. Box 1450, Alexandria, Virginia 22313-1450

Facsimile No. 571-273-3201

Authorized officer:

Lee W. Young

PCT Helpdesk: 571-272-4300

PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.

PCT/IL 08/01051

**Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)**

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1.  Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
  
2.  Claims Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
  
3.  Claims Nos.: 6-10, 14-17, 22-27, 31-32, 41-45, 50, and 57-58  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

**Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)**

This International Searching Authority found multiple inventions in this international application, as follows:

- Group 1: claims 1-5
- Group 2: claims 11-13, 54- 56, 59
- Group 3: claims 18-21
- Group 4: claims 28-30, 33, 47-49, 51 and 62-66
- Group 5: claims 34-35, 52-53
- Group 6: claims 36
- Group 7: claims 37
- Group 8: claims 38-40, 46
- Group 9: claims 60-61

(See Extra Sheet)

1.  As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
  
2.  As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
  
3.  As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
  
4.  No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:  
Claims 1-5

- Remark on Protest**
- The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
  - The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
  - No protest accompanied the payment of additional search fees.

Continuation of:

Box No. III --- Observations where unity of invention is lacking

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1.

Group 1, claims 1-5, drawn to a system for providing an alternative generator generating on the basis of an input sentence a text based representation providing multiple alternatives for each of the plurality of words in the sentence, a selector for selecting among the multiple alternatives based on an internet corpus and a correction generator which provides correction based on the selection made by the selector.

Group 2, claims 11-13, 54- 56, 59, drawn to a system for providing contextual feature sequence functionality and employing an internet corpus.

Group 3, claims 18-21, drawn to a system for providing an alternative generator generating on the basis of an input a text based representation providing multiple alternatives for each of the plurality of words in the sentence, a selector for selecting among at least said multiple alternatives for each of said plurality of words in the language input, based at least partly on a relationship between selected ones of said multiple alternative for at least some of said plurality of words in said language input; and a correction generator operative to provide a correction output based on selections made by said selector.

Group 4, claims 28-30, 33, 47-49, 51 and 62-66 drawn to a misused suspector evaluating most of the words in an input based on their fit within a context

Group 5, claims 34-35, 52-53 drawn to a selector which uses grading of words as well as ones of said multiple alternatives therefore generated by an alternatives generator according to multiple selection criteria and applying a bias in favor of said suspect word vis-a-vis ones of said multiple alternatives therefore generated by the alternatives generator.

Group 6, claims 36 drawn to a computer-assisted language correction system comprising an incorrect word suspector evaluating at least most of the words in a language input, said suspector being at least partially responsive to an input uncertainty metric indicating uncertainty of a person providing said input, said suspector providing a suspected incorrect word output; and an alternatives generator, generating a plurality of alternatives for suspected incorrect words identified by said suspected incorrect word output; a selector for selecting among each suspected incorrect word.

Group 7, claims 37 drawn to a computer assisted language correction system which has a phonetic similarity functionality operative to propose alternative words based on phonetic similarity to a word in said input and to indicate a metric of phonetic similarity; and character string similarity functionality operative to propose alternative words based on character string similarity to a word in said input and to indicate a metric of character string similarity for each alternative word.

Group 8, claims 38-40, 46 drawn to a computer-assisted language correction system comprising: suspect word identification functionality, receiving a multi-word language input and providing a suspect word output which indicates suspect words; feature identification functionality operative to identify features including said suspect words; an alternative selector identifying alternatives to said suspect words; feature occurrence functionality employing a corpus and providing an occurrence output, ranking various features including said alternatives as to their frequency of use in the corpus.

Group 9, claims 60-61 drawn to a computer-assisted language correction system which has a confidence level assigner operative to assign a confidence level to a selected alternative from said multiple alternatives.

The inventions listed as Groups 1-9 do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: The special technical feature of the Group 1 invention, not included in Groups 2-9, is providing language correction using a selector for selecting among the multiple alternatives based on an internet corpus. The special technical feature of the Group 2 invention, not included in Groups 1 and 3-9, is providing language correction output based on providing contextual feature sequence functionality and employing an internet corpus. Therefore, unity of invention is lacking. The special technical feature of the Group 3 invention, not included in Groups 1-2 and 4-9, is providing language correction output using a selector for selecting among at least said multiple alternatives for each of said plurality of words in the language input, based at least partly on a relationship between selected ones of said multiple alternative for at least some of said plurality of words in said language input. The special technical feature of the Group 4 invention, not included in Groups 1-3 and 5-9, is providing language correction output using a misused suspector evaluating most of the words in an input based on their fit within a context. The special technical feature of the Group 5 invention, not included in Groups 1-4 and 6-9, is providing language correction output using a selector which uses grading of words as well as ones of said multiple alternatives therefore generated by an alternatives generator according to multiple selection criteria and applying a bias in favor of said suspect word vis-a-vis ones of said multiple alternatives therefore generated by the alternatives generator. The special technical feature of the Group 6 invention, not included in Groups 1-5 and 7-9, is providing language correction output using a suspector being at least partially responsive to an input uncertainty metric indicating uncertainty of a person providing said input, said suspector providing a suspected incorrect word output. The special technical feature of the Group 7 invention, not included in Groups 1-6 and 8-9, is providing language correction output using a phonetic similarity functionality operative to propose alternative words based on phonetic similarity to a word in said input and to indicate a metric of phonetic similarity; and character string similarity functionality operative to propose alternative words based on character string similarity to a word in said input and to indicate a metric of character string similarity for each alternative word. The special technical feature of the Group 8 invention, not included in Groups 1-7 and 9, is providing language correction output using suspect word identification functionality, receiving a multi-word language input and providing a suspect word output which indicates suspect words; feature identification functionality operative to identify features including said suspect words; an alternative selector identifying alternatives to said suspect words; feature occurrence functionality employing a corpus and providing an occurrence output, ranking various features including said alternatives as to their frequency of use in the corpus. The special technical feature of the Group 9 invention, not included in Groups 1-8, is providing language correction output correction system which has a confidence level assigner operative to assign a confidence level to a selected alternative from said multiple alternatives. Groups 1, 3, 4, 5, 6, 7, 8 and 9 share the special technical features of an alternative generator, selector and correction generator. 2004/0093567 teaches that these technical features are well known. See figure 4. Schabes et al teach alternative generation (52, 56-57), selection (60), and correction (61). Groups 4-8 share the special technical feature of a misuse suspector. Schabes et al teach that such a suspector is well known (see paragraph 11, element 55). Groups 1 and 2 share the special technical feature of selecting based on an internet corpus. 2007/0127688 teaches analyzing a message against an internet corpus (see claim 56).