

(51) International Patent Classification:
G09B 19/00 (2006.01)

(21) International Application Number:

PCT/US2014/016354

(22) International Filing Date:

14 February 2014 (14.02.2014)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

61/765,105 15 February 2013 (15.02.2013) US

(71) Applicant: **VOXY, INC.** [US/US]; 632 Broadway, New York, NY 10012 (US).(72) Inventors: **NIELSON, Katharine**; 615 Wythe Avenue, Richmond, VA 23226 (US). **TYSON, Na-im**; 106 S 12th Avenue, Mount Vernon, NY 10550 (US). **BREEN, Andrew**; 523 E 14th Street, Apt. 5A, New York, NY 10009 (US). **KIRKHAM, Kasey**; 369 Broome St#1, New York, NY 10013 (US).(74) Agent: **SOPHIR, Eric**; Dentons US LLP, P.O. Box 061080, Wacker Drive Station, Willis Tower, Chicago, IL 60606 (US).

(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, BN, 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, IR, IS, JP, KE, KG, 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, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, 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) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV,

[Continued on next page]

(54) Title: LANGUAGE LEARNING SYSTEMS AND METHODS

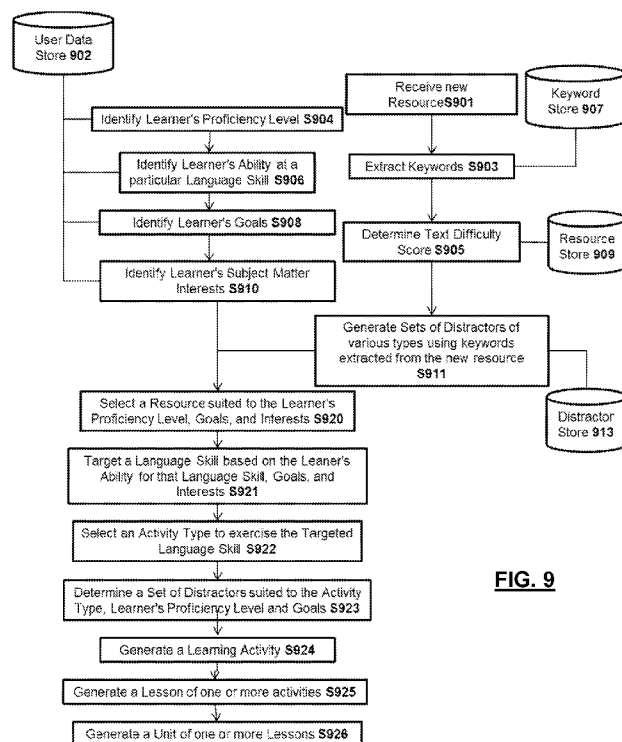


FIG. 9

(57) Abstract: A systems, methods, and products for language learning software that automatically extracts text from corpora using various natural-language processing product features, which can be combined with custom-designed learning activities to offer a needs-based, adaptive learning methodology. The system may receive a text, extract keywords pedagogically valuable to non- native language learning, assign a difficulty score to the text using various linguistic attributes of the text, generate a list of potential distractors for each keyword in the text to implement in learning activities, and topically tag the text against a taxonomy based on the content. This output is then used in conjunction with a series of learning activity-types designed to meet learners' language skill needs and then used to create dynamic, adaptive activities for them.



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, KM, ML, MR, NE, SN, TD, TG).

— *before the expiration of the time limit for amending the
claims and to be republished in the event of receipt of
amendments (Rule 48.2(h))*

Published:

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

(88) Date of publication of the international search report:

16 October 2014

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 14/16354

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - G09B 19/00 (2014.01)

USPC - 434/156

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)

IPC(8): G09B 19/00 (2014.01)

USPC: 434/156

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

USPC: 434/322 or 434/156 or 434/178 or 434/322 or 434/365 or 704/1 or 704/200

IPC(8): G09B 19/00 (2014.01)

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

Patbase; Google (Scholar, Patents); learn, learning, language, skills, goal, objective, distractors, difficulty, proficiency,

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	US 2010/0273138 A1 (EDMONDS et al.) 28 October 2010 (28.10.2010) entire document (especially para [0061]-[0065], [0070], [0073]-[0074], [0085]-[0087], [0093], [0096])	2, 4, 6 ----- 1, 3, 5, 7-27
Y	US 2012/0322035 A1 (Julia et al.) 20 December 2012 (20.12.2012) entire document (especially para [0071],[0115], [0120]-[0122], [0124], [0183])	1, 3, 5, 7-27
Y	US 2007/0206017 A1 (Johnson et al.) 06 September 2007 (06.09.2007) (para [0096]-[0097], [0162], [0165], [0188]-[0193])	22-25

☐ 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

"&" document member of the same patent family

Date of the actual completion of the international search

17 July 2014 (17.07.2014)

Date of mailing of the international search report

18 AUG 2014

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/US 14/16354

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.:
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:
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. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I: Claims 1-27 are drawn to a computer-implemented method comprising: selecting, by a computer, from a resource store a resource having content, wherein the selected resource is related to a content interest of a learner and has a resource difficulty level based upon a proficiency level of the learner; identifying, by the computer, a specific language skill for improvement based upon the one or more abilities of the learner; updating, by the computer, the ability for the specific language skill in the user data store according to a result from the learning activity; and updating, by the computer, the proficiency level of the learner in the user data store.

Group II: Claims 28-33 are drawn to a computer-implemented method for crafting a language learning pedagogy comprising: identifying, by the computer, a set of one or more learner preferences associated with the learner comprising one or more learner content interests; generating, by the computer, a lesson comprising a set of one or more learning activities according to a learner goal in the learner
-- 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.:
1-27

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 III: Lack of Unity

preferences.

Group III: Claims 34-43 are drawn to a computer-implemented method for tailoring language-learning to a learner comprising: determining, by a computer, an ability of a learner in a specific language skill based on a result of a skill assessment performed by the learner; determining, by the computer, a language proficiency level of the learner based on at least one language skill ability of the learner; receiving, by the computer, a learner content interest and a learner goal from the computing device of the learner; and wherein the learner profile comprises one or more language skill abilities of the learner, the language proficiency level, the learner content interest, and the learner goal.

Group IV: Claims 44-52 are drawn to a language learning system for automatically generating activities comprising: a host computer comprising a processor executing a set of software modules for language learning; a learner computer comprising a processor configured to execute a user interface to interact with a set of learning activities generated by a learning activity generator module executed by the host computer, wherein a learning activity is automatically generated using the capabilities and preferences of the learner stored in the learner profile.

Group V: Claims 53-71 are drawn to a computer-implemented method comprising: storing, by the computer, into a keyword store each potential keyword in the set matching a term in a computer file containing a keyword whitelist and each potential keyword in the set matching a collocation in the keyword whitelist; determining, by the computer, for one or more potential keywords in the set of potential keywords, a word difficulty value associated with each of the one or more potential keywords based on scoring rules determining the word difficulty value; and storing, by the computer, into the keyword store each potential keyword having a determined pedagogical value that satisfies a threshold word difficulty value.

Group VI: Claims 72 is drawn to a computer-implemented method for predicting a text difficulty score for a new resource, the method comprising: extracting, by a computer, one or more linguistic features having a weighted value from a plurality of training resources containing text, wherein the text is associated with a metadata label containing a text difficulty score of the text; determining, by the computer, a vector value associated with each training resource based on each of the weighted values of each of the extracted one or more linguistic features; training, by the computer, a statistical model using the vector values associated with each training resource, wherein the statistical model represents a correlation between a set of features selected for extraction, a set of weighted values assigned to the set of features selected for extraction, and a set of text difficulty scores associated with the training resources; extracting, by the computer, one or more linguistic features having a weighted value from a new resource; determining, by the computer, a vector value for the new resource based upon the set of extracted linguistic features; and predicting, by the computer, a text difficulty score for the new resource based upon the vector value for the new resource and the statistical model.

Group VII: Claims 73-89 are drawn to a computer-implemented method comprising: comparing, by a computer, at least a portion of text related to a resource against a lexical feature bank file comprising a listing of semantic features and a listing of syntactic features; identifying, by the computer, a set of lexical features associated with the text based on the comparison, wherein the set of lexical features comprises at least one semantic feature from the text matching a semantic feature in the list and at least one syntactic feature from the text matching a syntactic feature in the list; assigning, by the computer, a first value to each of the lexical features in the set of lexical features; and determining, by the computer, a text difficulty score for the text using each of the first values associated with each of the features in the set of lexical features.

Group VIII: Claims 90-115 are drawn to a computer-implemented method comprising: generating, by a computer, a set of one or more semantic distractors comprising one or more words having a definition that is related to a target word; generating, by the computer, a set of orthographic distractors comprising one or more words having an edit distance satisfying an edit distance amount setting, wherein the edit distance is a number changes to a word required to be identical to the target word, and wherein the edit distance amount setting determines the number of changes to the word; and generating, by the computer, a set of phonetic distractors comprising one or more homophones of the target word.

The inventions listed as Groups I through VIII 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 technical features for the following reasons:

Special technical features

The special technical feature of Group I is a computer-implemented method comprising: selecting, by a computer, from a resource store a resource having content, wherein the selected resource is related to a content interest of a learner and has a resource difficulty level based upon a proficiency level of the learner; identifying, by the computer, a specific language skill for improvement based upon the one or more abilities of the learner; updating, by the computer, the ability for the specific language skill in the user data store according to a result from the learning activity; and updating, by the computer, the proficiency level of the learner in the user data store, not required in any other group.

The special technical feature of Group II is a computer-implemented method for crafting a language learning pedagogy comprising: identifying, by the computer, a set of one or more learner preferences associated with the learner comprising one or more learner content interests; generating, by the computer, a lesson comprising a set of one or more learning activities according to a learner goal in the learner preferences, not required in any other group.

The special technical feature of Group III is a computer-implemented method for tailoring language-learning to a learner comprising: determining, by a computer, an ability of a learner in a specific language skill based on a result of a skill assessment performed by the learner; determining, by the computer, a language proficiency level of the learner based on at least one language skill ability of the learner; receiving, by the computer, a learner content interest and a learner goal from the computing device of the learner; and wherein the learner profile comprises one or more language skill abilities of the learner, the language proficiency level, the learner content interest, and the learner goal, not required in any other group.

-Continuation in Next Supplemental Box-

-Continuation in Preceding Supplemental Box-

The special technical feature of Group IV is a language learning system for automatically generating activities comprising: a host computer comprising a processor executing a set of software modules for language learning; a learner computer comprising a processor configured to execute a user interface to interact with a set of learning activities generated by a learning activity generator module executed by the host computer, wherein a learning activity is automatically generated using the capabilities and preferences of the learner stored in the learner profile, not required in any other group.

The special technical feature of Group V is a computer-implemented method comprising: storing, by the computer, into a keyword store each potential keyword in the set matching a term in a computer file containing a keyword whitelist and each potential keyword in the set matching a collocation in the keyword whitelist; determining, by the computer, for one or more potential keywords in the set of potential keywords, a word difficulty value associated with each of the one or more potential keywords based on scoring rules determining the word difficulty value; and storing, by the computer, into the keyword store each potential keyword having a determined pedagogical value that satisfies a threshold word difficulty value, not required in any other group.

The special technical feature of Group VI is a computer-implemented method for predicting a text difficulty score for a new resource, the method comprising: extracting, by a computer, one or more linguistic features having a weighted value from a plurality of training resources containing text, wherein the text is associated with a metadata label containing a text difficulty score of the text; determining, by the computer, a vector value associated with each training resource based on each of the weighted values of each of the extracted one or more linguistic features; training, by the computer, a statistical model using the vector values associated with each training resource, wherein the statistical model represents a correlation between a set of features selected for extraction, a set of weighted values assigned to the set of features selected for extraction, and a set of text difficulty scores associated with the training resources; extracting, by the computer, one or more linguistic features having a weighted value from a new resource; determining, by the computer, a vector value for the new resource based upon the set of extracted linguistic features; and predicting, by the computer, a text difficulty score for the new resource based upon the vector value for the new resource and the statistical model, not required in any other group.

The special technical feature of Group VII is a computer-implemented method comprising: comparing, by a computer, at least a portion of text related to a resource against a lexical feature bank file comprising a listing of semantic features and a listing of syntactic features; identifying, by the computer, a set of lexical features associated with the text based on the comparison, wherein the set of lexical features comprises at least one semantic feature from the text matching a semantic feature in the list and at least one syntactic feature from the text matching a syntactic feature in the list; assigning, by the computer, a first value to each of the lexical features in the set of lexical features; and determining, by the computer, a text difficulty score for the text using each of the first values associated with each of the features in the set of lexical features, not required in any other group.

The special technical feature of Group VIII is a computer-implemented method comprising: generating, by a computer, a set of one or more semantic distractors comprising one or more words having a definition that is related to a target word; generating, by the computer, a set of orthographic distractors comprising one or more words having an edit distance satisfying an edit distance amount setting, wherein the edit distance is a number changes to a word required to be identical to the target word, and wherein the edit distance amount setting determines the number of changes to the word; and generating, by the computer, a set of phonetic distractors comprising one or more homophones of the target word, not required in any other group.

Common Technical Features

-Groups I and II share the technical feature of identifying in a user data store that stores one or more abilities of the learner for specific language skills; identifying a set of one or more distractors of a type suited to the identified type of learning activity and having a distractor difficulty level based upon the ability of the learner in the specific language skill; generating, by the computer, a learning activity of the identified type of activity utilizing at least one of the set of one or more distractors; identifying, by the computer, a type of learning activity that exercises the identified language skill. However, these shared technical features does not represent a contribution over the prior art:

US 2010/0273138 A1 (EDMONDS et al.) (hereinafter Edmonds) 28 October 2010 (28.10. 2010)

Edmond teaches identifying in a user data store that stores one or more abilities of the learner for specific language skills (e.g. learner model 130 stores an estimate of a learner's degree of mastery of learning items, para [0062]-[0065]); identifying a set of one or more distractors of a type suited to the identified type of learning activity and having a distractor difficulty level based upon the ability of the learner in the specific language skill (e.g. consults the learner model 130 in order to generate a stem and distractors at the right level for the learner, para [0074]); generating, by the computer, a learning activity of the identified type of activity utilizing at least one of the set of one or more distractors (e.g. Step 350 can control the difficulty of the generated exercise by selecting the distractors in dependence on the selected difficulty level (step 320) and on the learner model 130. For example, in the case of exercises that use a sentence (e.g., a dictionary definition) in generating the distractors, the difficulty level of the sentence can be taken into account as in step 340. In the case of exercises that use words as distractors, the words can be selected to be at the right difficulty level, para [0086]; para [0074], para [0085],[0087], [0093]); identifying, by the computer, a type of learning activity that exercises the identified language skill (e.g. selects an exercise difficulty level by consulting the learner model 130 to find out the current level of mastery of the selected learning item, para [0073]).

-Groups I and II also share the technical feature of language proficiency level. However, these shared technical features does not represent a contribution over the prior art of Edmonds in view of US 2012/0322035 A1 (Julia et al.) (hereinafter Julia) 20 December 2012 (20.12.2012). Edmonds teaches language learning (para [0016]) and a knowledge level of a learner (abstract). A skilled artisan would immediately recognize this is a language proficiency level in the context of language learning. Nevertheless, Julia teaches language proficiency level (e.g. student's proficiency and language development....performance level in the language being learned, para [0120]-[0122], [0183]) and it would have been obvious to utilize this as such in the system and method of Edmonds.

-Groups I, II, and III share the technical feature of learner content interest and identifying abilities of the learner for specific language skills. However, these shared technical features does not represent a contribution over Julia.

-Continuations in Next Supplemental Box-

-Continuations in Preceding Supplemental Box-

Julia teaches a learner content interest (e.g. the content used by individual users may be different, and to each user's selection or preference, para [0165]) and identifying abilities of the learner for specific language skills (e.g. the user may utilize the language learning system 100 for more than one session, and maintain a history of his or her performance and evaluation results. For example, the user's performance level may be tracked over several sessions to monitor the user's improvement, para [0052]; user's performance is recorded, para [0120]);

-Groups III and IV share the technical feature of storing a learner profile associated with the learner in a user data store, and wherein the learner profile comprises a set of learner language abilities and a set of learner preferences. However, these shared technical features does not represent a contribution over Julia.

Julia teaches storing a learner profile associated with the learner in a user data store, and wherein the learner profile comprises a set of learner language abilities and a set of learner preferences (e.g. profile of that speaker, para [0057]; the data stored in the data structure 230 may be used to study a user's past performance, and to provide comparative basis for the user's progress, para [0067]; student's proficiency and language development....performance level in the language being learned, para [0120]-[0122], [0183]).

-Groups IV and V share the technical feature of parsing/extracting a set of one or more potential keywords from text of a resource. However, these shared technical features does not represent a contribution over the prior art:
US 2011/0111377 A1 (Dekkers) 12 May 2011 (12.05.2011)

Dekkers teaches parsing/extracting a set of one or more potential keywords from text of a resource (e.g. extracting words from individual source texts, para [0087]; abstract, para [0071]).

-Groups V, VI, and VII share the technical feature of difficulty score/value. However, these shared technical features does not represent a contribution over Edmonds.

Edmonds teaches difficulty score/value (e.g. Difficulty level can be a number in a range, for example, the range 0.0-1.0, where 0.0 represents easy and 1.0 represents difficult., para [0073]).

-Groups VI and VII share the technical feature of a statistical model and linguistic/semantic/syntactic features. However, these shared technical features does not represent a contribution over the prior art:
US 2005/0080613 A1 (Colledge et al.) (hereinafter Colledge) 14 April 2005 (14.04.2005)

Colledge teaches a statistical model (statistical models , para [0068], [0076], [0082]) and linguistic/semantic/syntactic features (e.g. analyze the syntax (e.g. semantic relationships that indicate specific semantic and syntactic relationships between word senses may be used, para [0057]; e.g. part of speech, grammatical relations) and semantics (e.g. logical relations) in context, para [0065]).

-Groups I, II and VIII share the technical feature of distractors. However, these shared technical features does not represent a contribution over Edmonds.

Edmonds teaches distractors (para [0074], para [0085],[0087], [0093]).

-Groups I and VI share the technical feature of resources. However, these shared technical features does not represent a contribution over Edmonds.

Edmonds teaches resources (e.g. teaches textual material is for example a textbook or an encyclopaedia, para [0096]).

As the above common features were known in the art at the time of the invention, these cannot be considered a special technical feature that would otherwise unify the groups.

Therefore, Groups I - VIII lack unity under PCT Rule 13 because they do not share a same or corresponding special technical feature.