Internal distribution code:

(A) [-] Publication in OJ
(B) [-] To Chairmen and Members
(C) [-] To Chairmen
(D) [X] No distribution

Datasheet for the decision of 22 October 2021

Case Number: T 1059/19 - 3.5.07

Application Number: 12755516.7

Publication Number: 2672395

IPC: G06F17/22

Language of the proceedings: EN

Title of invention:
TRANSLITERATION DEVICE, PROGRAM,RECORDING MEDIUM, AND METHOD

Applicant:
Rakuten Group, Inc.

Headword:
Transliteration device/RAKUTEN

Relevant legal provisions:
EPC Art. 83

Keyword:
Sufficiency of disclosure - (yes)

Europäisches Patentamt
European Patent Office
Office européen des brevets

Beschwerdekammern

Boards of Appeal

Chambres de recours

Boards of Appeal of the
European Patent Office
Richard-Reitzner-Allee 8
85540 Haar
GERMANY
Tel. +49 (0)89 2399-0
Fax +49 (0)89 2399-4465

Case Number: T 1059/19 - 3.5.07

D E C I S I O N
of Technical Board of Appeal 3.5.07
of 22 October 2021

Appellant: (Applicant)	Rakuten Group, Inc. 1-14-1 Tamagawa, Setagaya-ku, Tokyo 1580094 (JP)
Representative:	Hoffmann Eitle Patent- und Rechtsanwälte PartmbB Arabellastraße 30 81925 München (DE)
Decision under appeal:	**Decision of the Examining Division of the European Patent Office posted on 17 October 2018 refusing European patent application No. 12755516.7 pursuant to Article 97(2) EPC.**

Composition of the Board:

Chairman	J. Geschwind
Members:	C. Barel-Faucheux
	R. de Man

Summary of Facts and Submissions

I. The appeal lies from the examining division's decision to refuse European patent application No. 12755516.7, which was filed as international application PCT/JP2012/054956 and published under Article 153(4) EPC as EP 2 672 395.

II. The examining division refused the application for the following reasons: the application did not disclose the invention as defined by the main and first to fourth auxiliary requests in a manner sufficiently clear and complete for it to be carried out by a person skilled in the art, contrary to the requirements of Article 83 EPC; the amendments to the independent claims of the third auxiliary request added subject-matter which extended beyond the content of the application as filed, contrary to the requirements of Article 123(2) EPC; and claims 1 and 8 to 10 of the third auxiliary request lacked clarity under Article 84 EPC.

III. With the statement of grounds of appeal, the appellant filed an amended third auxiliary request and a new, fifth auxiliary request. It requested that the decision under appeal be set aside and that a patent be granted on the basis of the main request, one of the first, second and fourth auxiliary requests considered in the decision under appeal, or one of the amended third request or the new, fifth auxiliary request.

IV. The appellant was summoned to oral proceedings. In a subsequent communication sent in advance of the oral proceedings, the board expressed its preliminary opinion that. Accordingly, the question to be discussed during the oral proceedings with regard to Article 83 EPC was whether it was sufficiently clearly and completely described, in the description as originally filed, "how the original languages of the original spelling strings" were "implicitly classified into classes" and thereby collectively referred to as "implicit languages". The board also noted that the word "rule(s)" did not appear to be present in the description and claims as originally filed and that this might reveal a problem under Article 123(2) EPC.

V. With a letter of reply, the appellant filed a sixth auxiliary request, in which the term "rules" had been replaced by the term "similarities".

VI. Oral proceedings were held as scheduled, during which the appellant submitted two documents ("Lexical similarity", Wikipedia, and "Hauptfach Computerlinguistik", downloaded from the website of the Ludwig Maximilian University of Munich). It also replaced its requests with a single main request corresponding to the sixth auxiliary request.

VII. The appellant's final requests were that the decision under appeal be set aside and that a patent be granted on the basis of the claims of the main and sole request. At the end of the oral proceedings, the Chair announced the board's decision.

VIII. Claim 1 of the main request reads as follows (itemisation added by the board):

(**Feature 1**) A transliteration device (100) characterized by, comprising:

(**Feature 2**) a generation part (105) for generating, from a training set including multiple transliteration pairs consisting of an original spelling string spelled in any original language and a target spelling string transliterated from the original spelling string and spelled in a given target language and at least including original spelling strings of I original languages,

(**Feature 2.1**) K (where K is a natural number greater than or equal to 2 and less than or equal to I) rewriting tables corresponding to K different implicit languages and including multiple sets of an original segment constituting said original spelling string, a transliterated segment constituting said target spelling string, and a rewriting probability that the original segment is rewritten as the transliterated segment for transliteration,

(**Feature 2.2**) and K transliteration tables corresponding to said K implicit languages and including the multiple transliteration pairs included in said training set,

(**Feature 3**) wherein the number of said implicit languages K is a value presenting into how many linguistically relative classes the original spelling strings forming the multiple transliteration pairs belonging to the training set are implicitly classified according to orthographic similarities and phonetic similarities,

(**Feature 4**) wherein the original languages of the original spelling strings implicitly classified into the same class are collectively termed the implicit language, whereby original spelling strings originating from the linguistically relative classes are orthographically and phonetically more similar to each other than to original spelling strings originating from other linguistically relative classes,

(**Feature 5**) wherein the generation part (105) is arranged to generate the K rewriting tables by generating one of the rewriting tables and making K copies of the generated one rewriting table, wherein the rewriting probability saved in a k-th rewriting table is a probability that the original segment is rewritten as the transliterated segment under a condition that the original language of the original spelling string including the original segment is a k-th implicit language,

(**Feature 6**) wherein the generation part (105) is arranged to generate the K transliteration tables corresponding to the K implicit languages by generating one of said transliteration tables by an αß method using the generated one rewriting table and making K copies of the generated one transliteration table, wherein a k-th transliteration table includes information presenting a transliteration probability, said transliteration probability being a probability that the original spelling string is transliterated into the target spelling string under a condition that the original language of the original spelling string is the k-th implicit language,

(**Feature 7**) an update part (107) for calculating, for each of the multiple transliteration pairs included in

said training set in each of the K transliteration tables, a transliteration probability that the original spelling string of the transliteration pair is transliterated to the target spelling string of the transliteration pair when the original spelling string originates from the implicit language corresponding to the rewriting table using, among rewriting probabilities included in said K rewriting tables, the rewriting probability of a set of the transliterated segment and the original segment obtained from the transliteration pair, saving the transliteration probability in the transliteration table corresponding to the implicit language in association with the transliteration pair, so updating the rewriting probabilities included in said K rewriting tables as to maximize an expected value, which is calculated using the transliteration probability, of a likelihood function calculating a likelihood presenting how likely said K transliteration tables are when said training set is obtained, and repeating said calculation of the transliteration probabilities and said update of the rewriting probabilities; and

(**Feature 8**) an output part (108, 109) for outputting a transliteration of a first string into a second string, and

(**Feature 9**) wherein the generation part (105) is arranged to alter the transliteration probabilities included in the K transliteration tables to be different from each other such that said calculation of the transliteration probabilities and said update of the rewriting probabilities are repeatedly performed.

IX. Claim 8 defines a transliteration program characterized by allowing a computer to function as the

transliteration device of claim 1, claim 9 defines a computer-readable medium on which a transliteration program is recorded, the transliteration program characterized by allowing a computer to function as the transliteration device of claim 1.

X. Claim 10 reads as follows:

A transliteration method executed by a transliteration device (100) provided with a generation part (105), an update part (107) and an output part (108, 109), characterized by comprising:

a generation step in which said generation part (105) generates, from a training set including multiple transliteration pairs consisting of an original spelling string spelled in any original language and a target spelling string transliterated from the original spelling string and spelled in a given target language and at least including original spelling strings of I original languages,

K (where K is a natural number greater than or equal to 2 and less than or equal to I) rewriting tables corresponding to K different implicit languages and including multiple sets of an original segment constituting said original spelling string, a transliterated segment constituting said target spelling string, and a rewriting probability that the original segment is rewritten as the transliterated segment for transliteration,

and K transliteration tables corresponding to said K implicit languages and including the multiple transliteration pairs included in said training set,

wherein the number of said implicit languages K is a value presenting into how many linguistically relative classes the original spelling strings forming the multiple transliteration pairs belonging to the

training set are implicitly classified according to orthographic similarities and phonetic similarities,

wherein the original languages of the original spelling strings implicitly classified into the same class are collectively termed the implicit language, whereby original spelling strings originating from the linguistically relative classes are orthographically and phonetically more similar to each other than to original spelling strings originating from other linguistically relative classes,

wherein, in the generation step, the generation part (105) generates the K rewriting tables by generating one of the rewriting tables and making K copies of the generated one rewriting table, wherein the rewriting probability saved in a k-th rewriting table is a probability that the original segment is rewritten as the transliterated segment under a condition that the original language of the original spelling string including the original segment is a k-th implicit language,

wherein, in the generation step, the generation part (105) generates the K transliteration tables corresponding to the K implicit languages by generating one of said transliteration tables by an αß method using the generated one rewriting table and making K copies of the generated one transliteration table, wherein a k-th transliteration table includes information presenting a transliteration probability, said transliteration probability being a probability that the original spelling string is transliterated into the target spelling string under a condition that the original language of the original spelling string is the k-th implicit language,

an update step in which said update part (107) calculates, for each of the multiple transliteration

pairs included in said training set in each of the K transliteration tables, a transliteration probability that the original spelling string of the transliteration pair is transliterated to the target spelling string of the transliteration pair when the original spelling string originates from the implicit language corresponding to the rewriting table using, among rewriting probabilities included in said K rewriting tables, the rewriting probability of a set of the transliterated segment and the original segment obtained from the transliteration pair, saves the transliteration probability in the transliteration table corresponding to the implicit language in association with the transliteration pair, so updates the rewriting probabilities included in said K rewriting tables as to maximize an expected value, which is calculated using the transliteration probability, of a likelihood function calculating a likelihood presenting how likely said K transliteration tables are when said training set is obtained, and repeats said calculation of the transliteration probabilities and said update of the rewriting probabilities; and

an outputting step in which the output part (108, 109) outputs a transliteration of a first string into a second string, and

wherein, in the generation step, the generation part (105) alters the transliteration probabilities included in the K transliteration tables to be different from each other such that said calculation of the transliteration probabilities and said update of the rewriting probabilities are repeatedly performed.

XI. In view of the outcome of the appeal proceedings, the dependent claims of the main request are not relevant for the present decision. The appellant's arguments,

where relevant to this decision, are addressed in detail below.

Reasons for the Decision

The application

1. The application relates to transliteration, i.e. the phonetic translation between languages utilising different orthographic systems, such as a transliteration (between the pronunciation) of the (original American English) name "Barack Obama" and the Japanese writing (of this pronunciation) "バラクオバマ") (paragraphs [0025] and [0026] of the application). A transliteration system 1 is composed of a computer communication network 10, a terminal device 200, a Japanese language database server 301, a foreign language database server 302 and a transliteration device 100 (Figure 1 and paragraph [0021]).

2. The terminal device 200 is composed of a personal computer comprising a display part and an input part (such as a keyboard). The terminal device 200 displays an input screen for inputting an input pair "to verify". The input pair consists of a first string spelled in any original (foreign) language and a second string spelled in a given target language ("Japanese" in the example of Figure 2A) and considered by the user to be the string transliterated from the first string. The input screen also comprises a "VERIFY" button (Figure 2A; paragraphs [0024] and [0025]). The terminal device 200 generates an input pair based on the entered signals, and sends information presenting the generated input pair to the transliteration device 100. Subsequently, the terminal device 200 receives

information returned from the transliteration device 100, and displays a results display screen as shown in Figure 2B expressing the level of probability in percentage that the second string is a correct transliteration of the first string (paragraphs [0027] and [0029]).

3. The application concludes that, by "simply entering a first string and a second string that is a possible transliteration of the first string, the user of the transliteration device 100 can know the transliteration probability between the first and second strings even if he/she does not know what the original language of the first string is. Therefore, the user can determine whether the second string is the correct transliteration of the first string easily and accurately based on the transliteration probability" (paragraph [0101]).

Article 83 EPC

4. The board considers that the following parts of apparatus claim 1 of the main request are disclosed in the original application, by the paragraphs and figures cited below, in a manner sufficiently clear and complete for it to be carried out by a person skilled in the art (according to the claim's itemisation in section VIII. above):

- **Feature 1**, **Feature 2** and **Feature 8**: by paragraphs [0039] to [0048] and Figures 4 and 6A, together with the paragraphs cited under sections 1. and 2. above;

- **Feature 2.1**, **Feature 2.2 and Feature 5**: by paragraphs [0054], [0059], [0060] and [0068] and Figure 8A;

- Feature 6 and Feature 9: by paragraphs [0072] to [0074];

- Feature 7: by paragraphs [0074], [0075], [0077], [0083] to [0085], [0088] and [0089].

5. In its decision, the examining division stated that there was no sufficient disclosure in the original application as to "how the original languages of the original spelling strings" were "implicitly classified into classes" and collectively termed "implicit languages", which was a prerequisite of the invention.

The only passage of the description relating to such a classification could be found in paragraphs [0070] and [0071].

Grouping a certain number of original languages, which first had to be identified from the set of all possible languages according to the available strings, into K classes based on some level of phonetic and orthographic similarity was a non-trivial task which had not been further elaborated in the original disclosure. As such a classification was crucial to the functioning of the method of independent claim 10 of the main request and as it was not disclosed in a manner sufficiently clear and complete for it to be carried out by a person skilled in the art, the application failed to meet the requirements of Article 83 EPC.

As the K implicit languages could not be established, none of the subsequent steps of the method could be carried out. Indeed, the subsequent steps were reliant on orthographic and phonetic rules of the implicit languages.

The examining division stressed that, according to the independent claims, the original spelling strings could be in "any language" and not just in "well known languages used throughout the world". It also stressed that the claims required the classification to be carried out by the transliteration device. It was not obvious how to provide a device that was able to group any I languages into K language classes on the basis of their phonetic and orthographic similarity.

6. The above objection relates (merely) to
Feature 3: "*wherein the number of said implicit languages K is a value presenting into how many linguistically relative classes the original spelling strings forming the multiple transliteration pairs belonging to the training set are implicitly classified according to orthographic rules and phonetic rules*",
and
Feature 4: "*wherein the original languages of the original spelling strings implicitly classified into the same class are collectively termed the implicit language, whereby original spelling strings originating from the linguistically relative classes are orthographically and phonetically more similar to each other than to original spelling strings originating from other linguistically relative classes*".

7. The board notes that, in the application, "any original language" in Feature 2 is any language using alphabets in writing including English, French, Spanish, German, Polish, and Russian. However, the application states that "this is not restrictive" (paragraph [0025]). The original language of a first string can be a language specified by the user of the first string or a language

determined by the content presented by the first string (paragraph [0026]).

8. The number of (implicit) languages "K" is described as being a value representing into how many classes the original spelling strings s_n are implicitly classified according to the original language. A case in which the original languages of multiple original strings s_n in the training set are English, French, Spanish, German, Polish, or Russian is discussed. In such a case, if the received number of languages K is "5", the multiple original spelling strings s_n are implicitly classified into 5 classes, i.e. English, French, German, Polish, and Russian, according to their original language. On the other hand, if the received number of languages K is "3", the multiple original spelling strings s_n are implicitly classified into linguistically "relative" (Note from the board: or rather "related") classes such as English/German, French/Spanish, and Polish/Russian classes - or "implicit languages" - because original spelling strings s_n originating from the linguistically relative classes are orthographically and phonetically more similar to each other than to strings s_n originating from other languages (paragraphs [0069], [0070] and [0071]).

9. The board thus understands that the linguistically related classes correspond to groups of "implicit" languages that are orthographically and phonetically more similar to each other.

10. The appellant argued that "[i]f, for example, the training set includes the six original languages English, German, French, Spanish, Russian, and Polish, the parameter K can be any integer number between 2 and 6. If K is set to equal 6, each original language

represents its own class. However, one can set the value of K also below 6 (or generally < I), so that at least two original languages are grouped into one class. For example, K could be set to 3, which would allow to group the 6 original languages into 3 classes and to correspondingly take into account their respective association to one of the language families of Germanic, Romanic and Slavic languages" and "[t]he original languages of the original spelling strings which are implicitly classified into the same class are collectively termed the implicit language, whereby original spelling strings originating from the linguistically relative classes are assumed to be orthographically and phonetically more similar to each other than to original spelling strings originating from other linguistically relative classes. [...] this could reflect the existence of language families and account for the fact that English and German are, for example, more similar to each other than the original spelling strings from the languages French and Spanish".

11. The claim does not specify whether the parameter "K" is given as an input to the transliteration device or whether this parameter is determined by determining into how many linguistically relative classes the original spelling strings forming the multiple transliteration pairs belonging to the training set are implicitly classified according to orthographic similarities and phonetic similarities. The only constraint is that "K" can vary between 2 and I (I being the number of original languages).

12. The appellant argued that the applicable skilled person had ordinary knowledge in the field of computational linguistics. It provided the document "Hauptfach

Computerlinguistik 2009/2010", downloaded from the website of the Ludwig Maximilian University of Munich, to prove that this skilled person existed at the priority date (4 March 2011) of the present application. This skilled person would be aware of several possibilities concerning how orthographic and phonetic similarities between different languages can be defined both in terms of quality (qualitatively, including statements such as "French is close to Spanish" or "Russian is closer to Polish than to English") as well as quantity. Similarities between languages can also be formulated using quantitative measures, one usual quantitative measure being so-called "lexical similarity". The latter is a measure of the degree to which the word sets of two given languages are similar. It is indicated by a number from "0" (no common words) to "1" (total overlap between the languages). Once such a measure is provided, the implicit classification of I original languages into K implicit languages according to the measure was a standard exercise in combinatorics and optimisation that could readily be solved by the skilled person. Moreover, starting from the set of the six languages English, French, Spanish, German, Polish and Russian, the skilled person would not have any difficulty in extending the set of languages to languages such as Italian, Portuguese (both being close to French and Spanish), other Slavic languages such as Ukrainian, Czech or Slovak. The set also includes different types of writing systems, i.e. the Latin alphabet as well as the Cyrillic alphabet. Therefore, the skilled person would also understand that this is not an obstacle either and thus that nothing stands in the way of an extension to other alphabets such as the Greek or the Hebrew alphabet.

The appellant argued, during the oral proceedings, that technical tools already existed at the priority date that ran through vocabularies of any (existing) language to determine similarities between languages. This was either just a matter of comparing lists of words (for the orthographic similarities) or analysing sounds transcribed in computer-readable form (for the phonetic similarities).

The board was convinced by these arguments.

13. In particular, the degrees of similarity between the I languages in terms of orthographic and phonetic similarities might be pre-defined by linguists (see, for example, the Wikipedia page dated 4 June 2010 explaining lexical similarity) and given as an input to the transliteration device. The "implicit classification" of the I original languages is then performed by the transliteration device according to this parameter "K" and these degrees of similarity. Some language "families" (classes) might be pre-defined by a linguist and provided to and/or stored somewhere in the transliteration device (in the form of a table or in another form, for example).

14. The examining division considered that a transliteration device falling under the claims had to be able to deal with "any language" and not just with "well known languages used throughout the world". In this connection, the board notes that the claims do not refer to "any language" but to "any original language". There is no suggestion in the application that a transliteration device has to be able to deal with all possible (written) languages. The board thus considers it reasonable to understand the term "original language" as referring to the input languages for which

a transliteration device was designed. At the oral proceedings, the appellant confirmed this interpretation. The board has no doubt that the skilled person, with the help of either a linguist or the documents cited by the appellant, could prepare a table allowing a device to "classify" a predetermined set of original languages into K classes of linguistically related languages by performing a table lookup.

15. The board is thus satisfied that the requirements of Article 83 EPC are met.

Article 123(2) EPC

16. The board notes the replacement, in the independent claims, of the text "according to orthographic rules and phonetic rules" by "according to orthographic similarities and phonetic similarities", which is based on paragraphs [0070] and [0071] of the description. This replacement has fully addressed the board's concerns expressed in its communication in respect of compliance with Article 123(2) EPC.

Conclusion

17. Since the novelty and inventive step of the claims are not addressed in the decision, the board considers it appropriate to give the appellant the opportunity to have these aspects discussed in two instances and therefore to remit the case to the examining division for further prosecution according to Article 11 RPBA 2020.

Order

For these reasons it is decided that:

1. The decision under appeal is set aside.

2. The case is remitted to the department of first instance for further prosecution on the basis of the main request.

The Registrar: The Chairman:

B. Brückner J. Geschwind

Decision electronically authenticated