

CLAIMS

What is claimed is:

1. A text address processing method, comprising:
 - determining, according to social relation circles of users in a service system, at least one address set, each address set in the at least one address set comprising: at least two original text addresses; and
 - performing, for each address set, normalization processing on original text addresses in the address set, to obtain a target text address corresponding to the address set.
2. The method according to claim 1, wherein the determining, according to social relation circles of users in a service system, at least one address set comprises:
 - determining a social relation circle of each user in the service system; and
 - acquiring a text address used by each user and text addresses used by users in the social relation circle of each user, to make up an address set.
3. The method according to claim 1, wherein the performing normalization processing on original text addresses in the address set, to obtain a target text address corresponding to the address set comprises:
 - calculating, according to features of each two original text addresses in the address set, a degree of similarity between said “each two original text addresses”; and
 - determining, according to the degree of similarity between said “each two original text addresses”, whether said “each two original text addresses” can be normalized into one of said “each two original text addresses”, to obtain a target text address corresponding to the address set.
4. The method according to claim 3, wherein the calculating, according to features of each two original text addresses in the address set, a degree of similarity between said “each two original text addresses” comprises:

extracting features of each two original text addresses in the address set, to obtain at least one feature of a standard fragment feature, a longitude and latitude feature, and an alphanumeric feature of said “each two original text addresses”; and
for each feature in the at least one feature, calculating, according to the feature, a degree of similarity between said “each two original text addresses” corresponding to the feature.

5. The method according to claim 4, wherein the calculating, according to the feature, a degree of similarity between said “each two original text addresses” corresponding to the feature comprises:

processing, if the feature is a standard fragment feature, the standard fragment feature of said “each two original text addresses” by using a SimHash algorithm, to obtain a degree of similarity between said “each two original text addresses” on a standard fragment feature dimension;
processing, if the feature is a longitude and latitude feature, the longitude and latitude feature of said “each two original text addresses” by using a longitude and latitude distance algorithm, to obtain a degree of similarity between said “each two original text addresses” on a longitude and latitude feature dimension; and
processing, if the feature is an alphanumeric feature, the alphanumeric feature of said “each two original text addresses” by using a Jaccard coefficient algorithm, to obtain a degree of similarity between said “each two original text addresses” on an alphanumeric feature dimension.

6. The method according to any of claims 1 to 5, after the target text address corresponding to each address set is obtained, further comprising:
determining, according to a corresponding relationship between target text addresses and original text addresses formed during the normalization processing, at least two target text addresses corresponding to a same original text address; and

performing normalization processing on the at least two target text addresses corresponding to the same original text address.

7. The method according to claim 6, wherein the performing normalization processing on the at least two target text addresses corresponding to the same original text address comprises:

acquiring a fragment intersection of standard address fragments contained in original text addresses respectively corresponding to the at least two target text addresses corresponding to the same original text address; and
performing, according to the fragment intersection, normalization processing on the at least two target text addresses corresponding to the same original text address.

8. The method according to claim 7, wherein the performing, according to the fragment intersection, normalization processing on the at least two target text addresses corresponding to the same original text address comprises:

normalizing, if the fragment intersection can represent one of the at least two target text addresses corresponding to the same original text address, the at least two target text addresses corresponding to the same original text address into the target text address that the fragment intersection can represent.

9. The method according to claim 8, further comprising:

storing, if the fragment intersection can represent one of the at least two target text addresses corresponding to the same original text address, the fragment intersection and the target text address that the fragment intersection can represent into a feature knowledge base correspondingly.

10. A text address processing apparatus, comprising:

a determination module configured to determine, according to social relation circles of users in a service system, at least one address set, each

address set in the at least one address set comprising: at least two original text addresses; and

a normalization module configured to perform, for each address set, normalization processing on original text addresses in the address set, to obtain a target text address corresponding to the address set.

11. The apparatus according to claim 10, wherein the determination module is specifically configured to:

determine a social relation circle of each user in the service system; and acquire a text address used by each user and text addresses used by users in the social relation circle of each user, to make up an address set.

12. The apparatus according to claim 10, wherein the normalization module is specifically configured to:

calculate, according to features of each two original text addresses in the address set, a degree of similarity between said “each two original text addresses”; and

determine, according to the degree of similarity between said “each two original text addresses”, whether said “each two original text addresses” can be normalized into one of said “each two original text addresses”, to obtain a target text address corresponding to the address set.

13. The apparatus according to claim 12, wherein the normalization module is further specifically configured to:

extract features of each two original text addresses in the address set, to obtain at least one feature of a standard fragment feature, a longitude and latitude feature, and an alphanumeric feature of said “each two original text addresses”; and

for each feature in the at least one feature, calculate, according to the feature, a degree of similarity between said “each two original text addresses” corresponding to the feature.

14. The apparatus according to claim 13, wherein the normalization module is further specifically configured to:

process, if the feature is a standard fragment feature, the standard fragment feature of said “each two original text addresses” by using a SimHash algorithm, to obtain a degree of similarity between said “each two original text addresses” on a standard fragment feature dimension;

process, if the feature is a longitude and latitude feature, the longitude and latitude feature of said “each two original text addresses” by using a longitude and latitude distance algorithm, to obtain a degree of similarity between said “each two original text addresses” on a longitude and latitude feature dimension; and

process, if the feature is an alphanumeric feature, the alphanumeric feature of said “each two original text addresses” by using a Jaccard coefficient algorithm, to obtain a degree of similarity between said “each two original text addresses” on an alphanumeric feature dimension.

15. The apparatus according to any of claims 10 to 14, wherein

the determination module is further configured to: after the normalization module obtains the target text address corresponding to each address set, determine, according to a corresponding relationship between target text addresses and original text addresses formed during the normalization processing, at least two target text addresses corresponding to a same original text address; and

the normalization module is further configured to: perform normalization processing on the at least two target text addresses corresponding to the same original text address.

16. The apparatus according to claim 15, wherein the normalization module is specifically configured to:

acquire a fragment intersection of standard address fragments contained in original text addresses respectively corresponding to the at least two target text addresses corresponding to the same original text address;

and

perform, according to the fragment intersection, normalization processing on the at least two target text addresses corresponding to the same original text address.

17. The apparatus according to claim 16, further comprising:

a feature knowledge base configured to store, when the fragment intersection can represent one of the at least two target text addresses corresponding to the same original text address, the fragment intersection and the target text address that the fragment intersection can represent correspondingly.

18. A non-transitory computer readable medium that stores a set of instructions that is executable by at least one processor of a terminal device to cause the terminal device to perform a text address processing method, the method comprising:

determining, according to social relation circles of users in a service system, at least one address set, each address set in the at least one address set comprising: at least two original text addresses; and performing, for each address set, normalization processing on original text addresses in the address set, to obtain a target text address corresponding to the address set.

19. The non-transitory computer readable medium according to claim 18, wherein the determining, according to social relation circles of users in a service system, at least one address set comprises:

determining a social relation circle of each user in the service system; and acquiring a text address used by each user and text addresses used by users in the social relation circle of each user, to make up an address set.

20. The non-transitory computer readable medium according to claim 18, wherein the performing normalization processing on original text addresses in the address set, to obtain a target text address corresponding to the address set comprises:

calculating, according to features of each two original text addresses in the address set, a degree of similarity between said “each two original text addresses”; and
determining, according to the degree of similarity between said “each two original text addresses”, whether said “each two original text addresses” can be normalized into one of said “each two original text addresses”, to obtain a target text address corresponding to the address set.

21. The non-transitory computer readable medium according to claim 20, wherein the calculating, according to features of each two original text addresses in the address set, a degree of similarity between said “each two original text addresses” comprises:

extracting features of each two original text addresses in the address set, to obtain at least one feature of a standard fragment feature, a longitude and latitude feature, and an alphanumeric feature of said “each two original text addresses”; and
for each feature in the at least one feature, calculating, according to the feature, a degree of similarity between said “each two original text addresses” corresponding to the feature.

22. The non-transitory computer readable medium according to claim 21, wherein the calculating, according to the feature, a degree of similarity between said “each two original text addresses” corresponding to the feature comprises:

processing, if the feature is a standard fragment feature, the standard fragment feature of said “each two original text addresses” by using a SimHash algorithm, to obtain a degree of similarity between said “each two original text addresses” on a standard fragment feature dimension;
processing, if the feature is a longitude and latitude feature, the longitude and latitude feature of said “each two original text addresses” by using a longitude and latitude distance algorithm, to obtain a degree

of similarity between said “each two original text addresses” on a longitude and latitude feature dimension; and
processing, if the feature is an alphanumeric feature, the alphanumeric feature of said “each two original text addresses” by using a Jaccard coefficient algorithm, to obtain a degree of similarity between said “each two original text addresses” on an alphanumeric feature dimension.

23. The non-transitory computer readable medium according to any of claims 18 to 22, after the target text address corresponding to each address set is obtained, further comprising:

determining, according to a corresponding relationship between target text addresses and original text addresses formed during the normalization processing, at least two target text addresses corresponding to a same original text address; and
performing normalization processing on the at least two target text addresses corresponding to the same original text address.

24. The non-transitory computer readable medium according to claim 23, wherein the performing normalization processing on the at least two target text addresses corresponding to the same original text address comprises:

acquiring a fragment intersection of standard address fragments contained in original text addresses respectively corresponding to the at least two target text addresses corresponding to the same original text address; and
performing, according to the fragment intersection, normalization processing on the at least two target text addresses corresponding to the same original text address.

25. The non-transitory computer readable medium according to claim 24, wherein the performing, according to the fragment intersection, normalization processing on the at least two target text addresses corresponding to the same original text address comprises:

normalizing, if the fragment intersection can represent one of the at least two target text addresses corresponding to the same original text address, the at least two target text addresses corresponding to the same original text address into the target text address that the fragment intersection can represent.

26. The non-transitory computer readable medium according to claim 25, further comprising:

storing, if the fragment intersection can represent one of the at least two target text addresses corresponding to the same original text address, the fragment intersection and the target text address that the fragment intersection can represent into a feature knowledge base correspondingly.