WHAT IS CLAIMED IS:

1. A method for determining a Human Phenotype Ontology, HPO, comprising:

obtaining symptom description information;

for a segmentation word list in a segmentation word group corresponding to an HPO in a preset HPO collection, querying, from the symptom description information, a segmentation word of the segmentation word list, determining a segmentation word included in both the segmentation word list and the symptom description information as a common segmentation word; determining the number of common segmentation words in the segmentation word list and respective positions of the common segmentation words in the symptom description information; determining a first similarity between the segmentation word list and the symptom description information based on the number of common segmentation words in the segmentation word list and the number of segmentation words in the segmentation word list; determining a measures of dispersion between the HPO and the symptom description information based on the respective positions of the common segmentation words in the symptom description information and respective positions of the common segmentation words in the HPO;

determining a second similarity between the HPO and the symptom description information based on a corresponding first similarity of each segmentation word list;

determining a matching degree between the HPO and the symptom description information based on the second similarity, the measures of dispersion, and the number of segmentation word lists in the segmentation word group; and

determining an HPO, wherein the matching degree corresponding to the HPO satisfies a preset matching success condition.

2. The method according to claim 1, wherein segmentation words are different between each segmentation word list of the segmentation word group, and the number of words of each segmentation word is the same for the same segmentation word list.

3. The method according to claim 1~~ or 2~~, wherein after determining the HPO, in which the matching degree corresponding to the HPO satisfies the preset matching success condition, the method further comprises:

determining, from the symptom description information, a description expression corresponding to the HPO; and

detecting whether an adjacent expression adjacent to the description expression is a preset negative expression, and labeling the adjacent expression in response to determining that the adjacent expression is the preset negative expression.

4. The method according to ~~any one of claims 1 to 3~~claim 1, wherein after determining the HPO, in which the matching degree corresponding to the HPO satisfies the preset matching success condition, the method further comprises:

determining, from the symptom description information, a description expression corresponding to the HPO; and

determining whether the description expression is consistent with the HPO,

determining that the HPO is completely matched with the description expression, in response to the description expression being consistent with the HPO, and

determining that the HPO is roughly matched with the description expression, in response to the description expression being inconsistent with the HPO.

5. The method according to ~~any one of claims 1 to 4~~claim 1, wherein determining the second similarity between the HPO and the symptom description information based on the corresponding first similarity of each segmentation word list comprises:

determining a third similarity between the HPO and the symptom description information based on the corresponding first similarity of each segmentation word list; and

determining the second similarity between the HPO and the symptom description information by multiplying the third similarity by an adjustable weighting factor.

6. The method according to ~~any one of claims 1 to 5~~claim 1, wherein determining the measures of dispersion between the HPO and the symptom description information based on respective positions of the common segmentation words in the symptom description information and the respective positions of the common segmentation words in the HPO, comprises:

determining a first position vector of the segmentation word list in the symptom description information based on the respective positions of the common segmentation words in the symptom description information;

determining a second position vector of the segmentation word list in the HPO based on the respective positions of the common segmentation words in the HPO; and

determining the measures of dispersion between the HPO and the symptom description information based on a corresponding first position vector and a corresponding second position

vector of each segmentation word list.

7. The method according to ~~any one of claims 1 to 6~~claim 1, wherein determining the matching degree between the HPO and the symptom description information based on the second similarity, the measures of dispersion, and the number of segmentation word lists in the segmentation word group, comprises:

$$AEEV=\frac{VAR}{N}\times Score$$

where, $AEEV$ is the matching degree between the HPO and the symptom description information, VAR is the measures of dispersion between the HPO and the symptom description information, and $Score$ is the second similarity between the HPO and the symptom description information, and N is the number of segmentation word lists in the segmentation word group;

$$Score=\gamma\times exp(preScore)$$

where,γ is an adjustable weighting factor, and *preScore* is a third similarity,

$$preScore=\sum_{n=1}^{N}\alpha\beta_n w_n logp_n$$

where, n is a serial number of each segmentation word list, $\alpha\beta_n w_n logp_n$ is the first similarity of the segmentation word list n, where $1\leq n\leq N$, α is an adjustment factor, β_n is a normalization factor corresponding to the segmentation word list n, w_n is a weight corresponding to the segmentation word list n, and p_n is a ratio of the number of common segmentation words in the segmentation word list n to the number of segmentation words in the segmentation word list n;

$$VAR=\sqrt{\frac{\sum_{n=1}^{N}\left(a_n-b_n-\frac{\sum_{n=1}^{N}(a_n-b_n)}{N}\right)^2}{N-1}}$$

where, a_n is a first position vector of the segmentation word list n in the symptom description information, and b_n is a second position vector of the segmentation word list n in the HPO.

8. An apparatus for determining a Human Phenotype Ontology (HPO), comprising: a symptom description information obtaining unit, a matching degree determining unit, and a HPO determining unit, wherein the matching degree determining unit comprises: a common segmentation word determining sub-unit, a common segmentation word number determining sub-unit, a common segmentation word position determining sub-unit, a first similarity determining

sub-unit, a measures of dispersion determining sub-unit, a second similarity determining sub-unit, and a matching degree determining sub-unit;

the symptom description information obtaining unit is configured to obtain symptom description information;

the common segmentation word determining sub-unit is configured to, for a segmentation word list in a segmentation word group corresponding to an HPO in a preset HPO collection, query, from the symptom description information, a segmentation word of the segmentation word list, and determine a segmentation word included in both the segmentation word list and the symptom description information as a common segmentation word;

the common segmentation word number determining sub-unit is configured to determine the number of common segmentation words in the segmentation word list;

the common segmentation word position determining sub-unit is configured to determine respective positions of the common segmentation words in the symptom description information;

the first similarity determining sub-unit is configured to determine a first similarity between the segmentation word list and the symptom description information based on the number of common segmentation words in the segmentation word list and the number of segmentation words in the segmentation word list;

the measures of dispersion determining sub-unit is configured to determine a measures of dispersion between the HPO and the symptom description information based on the respective positions of the common segmentation words in the symptom description information and respective positions of the common segmentation words in the HPO;

the second similarity determining sub-unit is configured to determine a second similarity between the HPO and the symptom description information based on a corresponding first similarity of each segmentation word list;

the matching degree determining sub-unit is configured to determine a matching degree between the HPO and the symptom description information based on the second similarity, the measures of dispersion, and the number of segmentation word lists in the segmentation word group; and

the HPO determining unit is configured to determine an HPO, wherein the matching degree corresponding to the HPO satisfies a preset matching success condition.

9. The apparatus according to claim 8, comprising: a description expression determining unit,

a negative expression detecting unit, and a negative expression labeling unit, wherein,

the description expression determining unit is configured to determine, from the symptom description information, a description expression corresponding to the HPO, after determining the HPO, in which the matching degree corresponding to the HPO satisfies the preset matching success condition;

the negative expression detecting unit is configured to detect whether an adjacent expression adjacent to the description expression is a preset negative expression, and trigger the negative expression labeling unit in response to detecting that the adjacent expression is the preset negative expression; and

the negative expression labeling unit is configured to label the adjacent expression.

10. The apparatus according to claim 8~~ or 9~~, comprising: a description expression determining unit, an expression determining unit and an expression matching determining unit, wherein,

the description expression determining unit is configured to determine, from the symptom description information, a description expression corresponding to the HPO; and

the expression determining unit is configured to determine whether the description expression is consistent with the HPO, trigger the expression matching determining unit in response to the description expression being consistent with the HPO, wherein the expression matching determining unit is configured to determine that the HPO is completely matched with the description expression, and trigger the expression matching determining unit in response to the description expression being inconsistent with the HPO, wherein the expression matching determining unit is configured to determine that the HPO is roughly matched with the description expression.

11. A storage medium, having a program stored thereon, wherein when the program is executed by a processor, a method for determining a HPO according to any one of claims 1 to 7 is executed.

12. A processor, for running a program, wherein when the program is running, a method for determining a HPO according to any one of claims 1 to 7 is executed.

13. A device, comprising at least one processor, at least one memory and a bus, wherein the at least one memory and the bus are connected to the processor, the processor communicates with the memory through the bus; the processor is configured to call program instructions in the memory to execute a method for determining a HPO according to any one of claims 1 to 7.

14. A computer program product, suitable to a program for executing a method for determining a HPO according to any one of claims 1 to 7 when being executed on a data processing device.