Objective
For given query, we try to utilize two query methods to improve the search effectiveness.

Methods
- **Centroid method**
  The centroid method represents q by:
  \[ S_{Cent}(t; q) \overset{\text{def}}{=} \exp(\cos(t, \vec{q}_{Cent})) \cdot \]
  \[ \vec{q}_{Cent} \overset{\text{def}}{=} \sum_{q_i \in q} q_i . \]
  \(^t\): the L2-normalised Word2Vec vector representing term t.
  \( q_i \): query q’s i-th query term.
- **CombMAX method**
  Select 5 most similar terms for each \( q_i \) according to \( \cos(\vec{q}_i, t) \).
  Similarities are softmax-normalised:
  \[ S_{CombMAX}(t; q) \overset{\text{def}}{=} \max_{q_i \in q} p(t|q_i) . \]
  \[ p(t|q_i) \overset{\text{def}}{=} \frac{\exp(\cos(\vec{q}_i, t))}{\sum_{t' \in L_{q_i}} \exp(\cos(\vec{q}_i, t'))} . \]

Query expansion
Select top 3, 5 terms according to the term selection scores, and confuse it with maximum likelihood estimate.

Data
SogouT16

Results
Submitted runs

<table>
<thead>
<tr>
<th>Run name</th>
<th>Term scoring method</th>
<th>#Expansion terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLWWW-C-NH-Base-1</td>
<td>Centroid</td>
<td>3</td>
</tr>
<tr>
<td>SLWWW-C-NH-Base-2</td>
<td>CombMAX</td>
<td>3</td>
</tr>
<tr>
<td>SLWWW-C-NH-Base-3</td>
<td>CombMAX</td>
<td>5</td>
</tr>
<tr>
<td>SLWWW-C-NH-Base-4</td>
<td>Centroid</td>
<td>5</td>
</tr>
</tbody>
</table>

Official results and baseline results. Pity..

Further work
Try other models to utilize the user click information.