With improvements in technology and access to the internet, people are increasingly using the Internet to research their health concerns. According to Pew Research Center’s Internet & American Life Project, more than 35% of adults in the United States regularly use the Internet to self-diagnose their ailments, using it both for non-urgent symptoms and for urgent symptoms. There are many systems in English that can support self-diagnosis as well as provide medical information such as WebMD, Mayo Clinic, NHS Choice, etc. However, there are not many such systems in Vietnam that can provide reliable and easy to understand clinical information. **Healthkee** aims to become a smart health assistant for Vietnamese that provide personalized health information, which is accurate, fast and easy to understand anywhere, anytime.

### Medical data
- Collaborated with a group of students from Hanoi Medical University.
- Provided by:
  - Hanoi Medical University Hospital
  - Vietnam – Germany Hospital
  - Bach Mai Hospital
  - National Hospital of Obstetrics and Gynecology
  - Vietnam National Hospital of Pediatrics

<table>
<thead>
<tr>
<th>Concept</th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conditions</td>
<td>170</td>
</tr>
<tr>
<td>Symptoms</td>
<td>724</td>
</tr>
<tr>
<td>Risk factors</td>
<td>384</td>
</tr>
<tr>
<td>Medications</td>
<td>575</td>
</tr>
<tr>
<td>Tests and procedures</td>
<td>81</td>
</tr>
<tr>
<td>Complications</td>
<td>180</td>
</tr>
</tbody>
</table>

### Medical concepts searching
- Working offline on Mobile.
- Using Inverted Index technique.
- Searching with/without accents and variants
- Ranking used element-wise comparison.
- The relevance between user’s query q and medical concepts c is a vector:
  \[
  \text{score}(q,c) = \langle f(q,c) \rangle \cdot g(s, c) \in \mathbb{R}^{|C| + 1}
  \]
  where:
  - \( f(q,c) = \sum_{i=1}^{n} \log(1 + s_{i}) \cdot g(s_{i}, c) \)
  - \( g(s_{i}, c) = \exp(-|\text{length}(q) - \text{length}(c)|) \)
  - \( s_{i} = \log \left( 1 + \frac{\text{occur}(q,c)}{\text{doc}(q) \times \text{doc}(c)} \right) \)

### Symptom suggestion
- Suggesting the most relevant symptoms which users may concern but not selected and thereby help to lead users to appropriate diseases.
- Idea: find all symptom candidates then rank them by calculating the relevance between the set of symptoms chosen by users and those candidates:
  \[
  \text{score}(s_i, \Sigma s_{\text{can}}) = \log(1 + h) \cdot \max \left[ \left( 1 - e^{-\omega} \right) \right]
  \]
  where:
  - \( h = \sum_{i=1}^{n} r(s_i, c) \cdot \sum_{j=1}^{m} r(c_j, c) \in [0,1] \) with \( r(s_i, c) \) demonstrates the relevance between \( s_i \) and \( c \).

### Medical recommendation system
- Personalized distributing over 150k+ medical articles from 20 well-known newspapers based on user interests.
- Using Latent Dirichlet Allocation to analyze hidden topics from articles.
- Topic-sensitive recommendation with latest user reading articles analysis.

### Condition ranking
Find all possible conditions based on chosen symptoms and demographics, then rank them in an appropriate order: the most relevant conditions go first, the less relevant conditions go after or may be eliminated.

\[
\hat{f}(s_i, c) = \sum_{c_j \in \mathcal{C}} r(s_i, c_j) \cdot g(s_i, c_j) + h(s_i, c)
\]

where:
- \( \mathcal{D}(s_i) = \{(s_i, c_j) | s_i \in \mathcal{S}, c_j \in \mathcal{C} \} \)
- \( g(s_i, c_j) = \left( 1 + \log \left( \frac{\text{doc}(s_i)}{\text{doc}(c_j)} \right) \right) \log \left( 1 + \text{doc}(s_i) \times \text{doc}(c_j) \right) \)
- \( h(s_i, c_j) = 1 + \sum_{c_k \in \mathcal{C}} r(s_i, c_k) \cdot r(c_j, c_k) \)

### Conclusion & Future works
- The system can provide reliable healthcare information for daily usage.
- At the early stage in clinical level because of the lack of information about patient’s disease profile or conducted tests and procedures.
- Plan to collect more medical data from trusted sources to broaden the database.
- Improve the algorithms (eg: using Bayes network for Symptom Checker)

### Table 1: The accuracy of Healthkee’s symptom checker with others

<table>
<thead>
<tr>
<th>Type of SPV or diagnosis</th>
<th>Number of SPV</th>
<th>Listed to 5% (%)</th>
<th>Listed to top 10% (%)</th>
<th>Listed to top 30% (%)</th>
<th>Not found (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All signatures</td>
<td>65</td>
<td>34</td>
<td>43</td>
<td>51</td>
<td>61</td>
</tr>
</tbody>
</table>

### Table 2: The accuracy of Healthkee’s symptom checker with others

| Type of SPV | Evaluation
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3.5</td>
</tr>
<tr>
<td></td>
<td>3.5</td>
</tr>
<tr>
<td></td>
<td>3.5</td>
</tr>
<tr>
<td></td>
<td>3.5</td>
</tr>
<tr>
<td></td>
<td>3.5</td>
</tr>
</tbody>
</table>

*Number of cases: SPV evaluation divided by applicable SPV evaluation.