Consider the following documents:

<table>
<thead>
<tr>
<th></th>
<th>doc\textsubscript{1}</th>
<th>doc\textsubscript{2}</th>
<th>doc\textsubscript{3}</th>
<th>doc\textsubscript{4}</th>
<th>doc\textsubscript{5}</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>phone ring person happy person</td>
<td>dog pet happy run jump</td>
<td>cat purr pet person happy</td>
<td>life smile run happy</td>
<td>life laugh walk run run</td>
</tr>
</tbody>
</table>

\textbf{(a) (i)} Construct the inverted index required for ranked retrieval for these five documents. Assume that no stemming or stop-word removal is required. [3 marks]

\textbf{(ii)} What is the complexity of processing a two-term conjunctive query using standard postings lists? Briefly describe one technique that can improve this efficiency. [2 marks]

\textbf{(iii)} Relating to the sample documents above, outline how the processing of the following Boolean query can be optimised:

\text{happy AND run AND pet} [2 marks]

\textbf{(iv)} What is the query-likelihood method in the language modelling approach to information retrieval? How does this differ conceptually from the measure of similarity used in the vector space model? [3 marks]

\textbf{(b) (i)} Smoothing is crucial in the language modelling approach to information retrieval. Why is smoothing important and how is it typically achieved? [2 marks]

\textbf{(ii)} Given the query \{\textit{happy person smile}\}, show how a unigram language modelling approach would rank the documents outlined above. Choose a suitable form of smoothing and include all your workings. State any other assumptions made. [6 marks]

\textbf{(iii)} How might you relax the \textit{term-independence} assumption in the unigram language model and how might it affect subsequent retrieval? [2 marks]