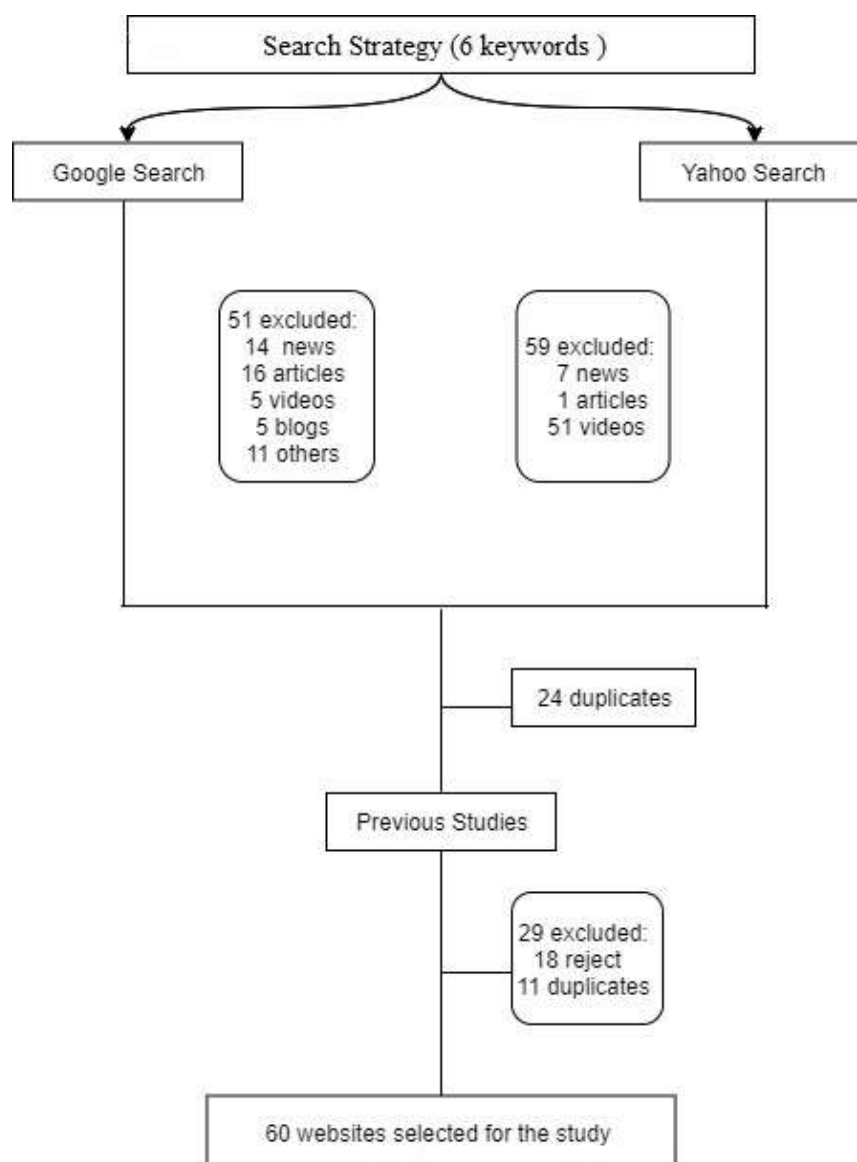


**Figure 1 shows the process of selection of websites.**

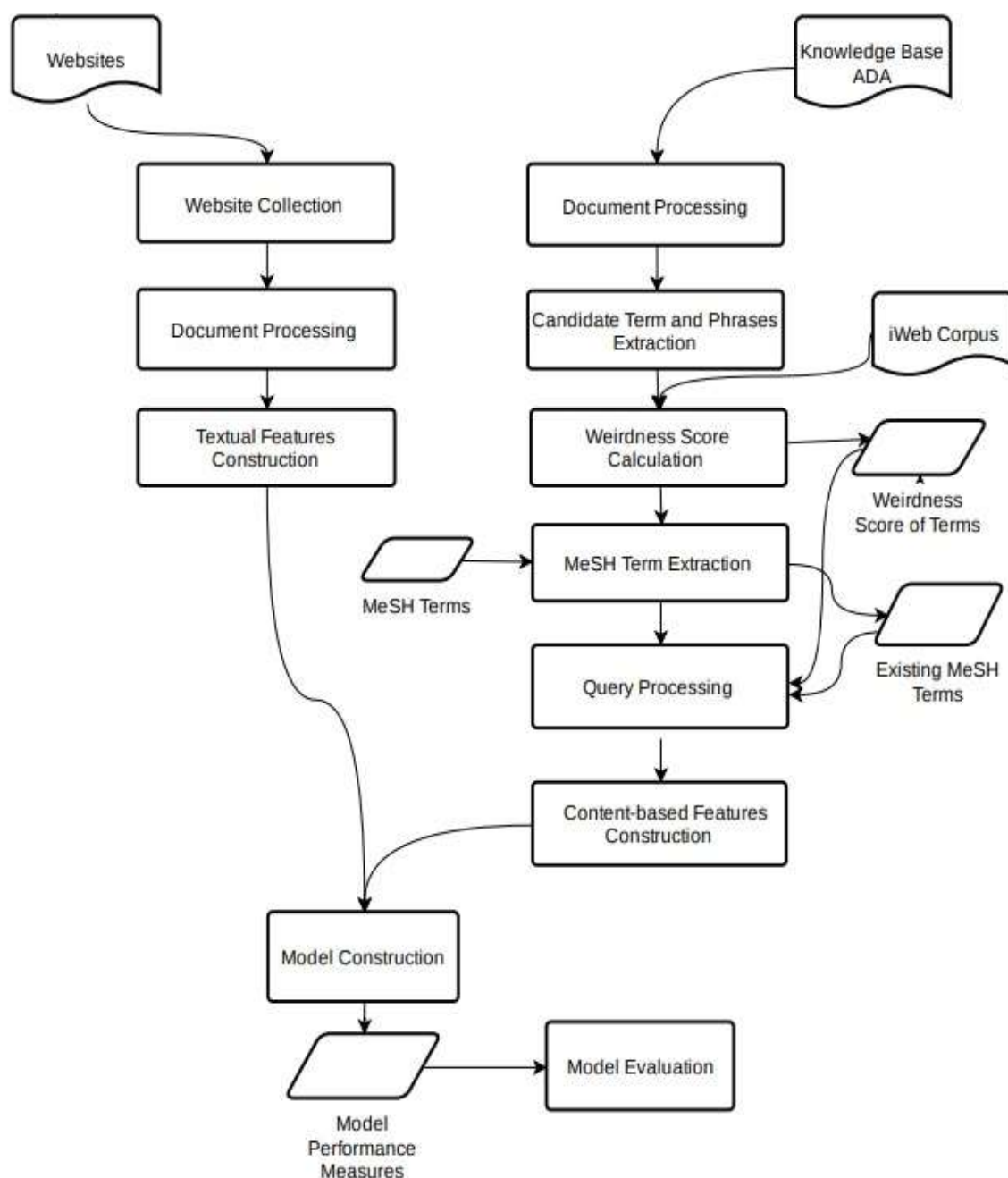
For the collection of websites, keyword searches were performed on Google and Yahoo search engines in June 2016 and August 2016. Six keywords were selected based on type 2 diabetes and DISCERN questions. The following keywords used: “type 2 diabetes”, “type 2 diabetes treatments”, “type 2 diabetes treatments risks”, “type 2 diabetes treatments benefits”, “type 2 diabetes no treatment” and “life with type 2 diabetes”. A wide range of search keywords is used to ensure reaching websites that can be used to assess the coverage of websites related to DISCERN questions.



**Figure 1:** Process of selection of the websites

*The methodology employed in this study is summarized in Figure 2.*

The crawled websites were initially pre-processed before conducting the analyses. If a webpage of the website was not related to type 2 diabetes, it was excluded due to the comprehensiveness criteria and then remaining webpages were gathered together in a document. Some textual features such as paragraph count and image count were obtained using the HTML tags. To extract content-based features, the content was first tokenized, and then all upper-case letters were transformed into lower-case letters. Only one sentence was kept if multiple occurrences of the same sentence were found in the text. In addition, stop words and punctuations were removed from each document, and lemmatization was applied. Finally, the candidate terms were generated from ADA, and their presence was investigated on the websites.



**Figure 2:** Methodology