FindZebra is a search engine for rare diseases intended to act as a diagnosis decision support system (DDSS) capable of assisting the user both during and after a search. Rare diseases are diseases that affect only a small part of the population (less than one in two thousand). Currently around seven thousand rare diseases are known and it is estimated that 6–8% of the population will be affected by a rare disease during their lifetime. Due to their rarity and large number, diagnosis of rare diseases is difficult and often associated with year long delays and diagnostic errors. These difficulties with diagnosis have a profound human and societal cost. This means that even a small increase in success rate when using a tool such as FindZebra could potentially have a great impact on society. In this dissertation we explore four lines of research for improving FindZebra using machine learning methods. The first line of research is on how to improve the retrieval performance of FindZebra. By using a combination of improved models, medical databases and corpus expansion we show that it is possible to obtain a substantial improvement in retrieval performance compared to current state-of-the-art document retrieval systems. Improving retrieval performance is important, but is not the only way of improving the success rate of a DDSS such as FindZebra. Following an unsuccessful search, the search engine should assist the user by indicating what information is likely to be missing. This idea is called Information Completion (IC) and will be explored in the second line of research. In order to represent words (and other discrete tokens) in a neural network it is necessary to transform each word to a vector form. This is typically accomplished by using a word embedding, which is an essential component in any word based neural network. The third line of research is on how to improve this basic component. Users of FindZebra who do not have English as their primary language often have difficulty expressing complex medical queries in English. Optimally, a user should be able to write a query in his or her native language and the search engine should then give a suggestion for a differential diagnosis based on all the information contained in a multilingual corpus, not only in the native corpus. Methods for performing multilingual search will be the fourth line of research explored in this dissertation.