

This thesis deals with analysis and system design for automatic document recognition. The system explores the document and converts it into text data with information about the position of the word in original document. These data will then be reviewed and some of them will be assigned their importance. The way the data will be assigned is based on rules which may vary according to user needs. According to the data, their assignment and the importance of their position, the system finds a similar document and identifies the current document.