

**CHENNAI – PONDICHERRY**

**FINDING RELATED FORUM POSTS THROUGH CONTENT SIMILARITY OVER INTENTION-BASED SEGMENTATION**

**Abstract:**

We study the problem of finding related forum posts to a post at hand. In contrast to traditional approaches for finding related documents that perform content comparisons across the content of the posts as a whole, we consider each post as a set of segments, each written with a different goal in mind. We advocate that the relatedness between two posts should be based on the similarity of their respective segments that are intended for the same goal, i.e., are conveying the same intention. This means that it is possible for the same terms to weigh differently in the relatedness score depending on the intention of the segment in which they are found. We have developed a segmentation method that by monitoring a number of text features can identify the parts of a post where significant jumps occur indicating a point where a segmentation should take place. The generated segments of all the posts are clustered to form intention clusters and then similarities across the posts are calculated through similarities across segments with the same intention. We experimentally illustrate the effectiveness and efficiency of our segmentation method and our overall approach of finding related forum posts.