A Test Collection for Research on Depression and Language Use

Abstract: Several studies in the literature have shown that the words people use are indicative of their psychological states. In particular, depression was found to be associated with distinctive linguistic patterns. In this talk, I will describe our first steps to try to identify as early as possible if the writer is developing a depressive state. I will detail the methodology we have adopted to build and make publicly available a test collection on depression and language use. The resulting corpus includes a series of textual interactions written by different subjects. The new collection not only encourages research on differences in language between depressed and non-depressed individuals, but also on the evolution of the language use of depressed individuals. I will also present the new CLEF lab that we will run next year on this topic (eRisk 2017), that includes a novel depression detection task and the proposal of effectiveness measure to systematically compare early detection algorithms and baseline results.