



162162



G00002

TEXT ZONING IN A SEMI-STRUCTURED TEXT

ABSTRACT

5 The present invention relates to a computer-assisted method for extracting domain-specific information from a semi-structured text comprising the steps of preparing a training set by manually annotating resume text or other recruitment related text with boundary information that separates the text into zones and sub zones; inducing decision tree classifiers for zoning and sub zoning by running a decision tree induction algorithm
10 on a pre-classified training text set Then the decision tree classifier is utilised to segment the resume text or other recruitment related text into individual text zones and sub zones containing domain-specific information whereby the domain-specific information is able to be extracted thereafter by any text extraction algorithm.

15 The Most Illustrative Drawing: FIG. 1