Feature Name	Feature Description
#ofWords	{0,1,2,3} 0: the number of words larger than 1 but less than 10 1: the number of words larger than 10 but less than 30 2: the number of words larger than 30 3: others
isLi	{0,1} 0: this text unit is not marked by \ in HTML 1: others
isH	{0,1} 0: this text unit is not marked by \<h> in HTML 1: others
FontSize	{0,1,2}, supposed the most frequent font size in the page is *FFS* 0: the font size is less than *FFS* 1: the font size is equal to *FFS* 2: others
FontWeight	{0,1} 0: the text is not marked in bold 1: others
XxxRatio	{0,1,2} 0: the ratio of words whose first letter is capital is less than 0.2 1: the text consists of only one word and the first letter of it is capital 2: others
ContainUrl	{0,1} 0: the text contains hyperlink 1: others
ContainColon	{0,1,2} 0: the text does not contain colon 1: the text contains colon, and colon is not located at the end 2:others
VisionLocation	{0,1,2} 0: the text is located at the top of the page 1: the text is located at the middle of the page 2: others
Depth	{0,1,2} let depth(t) be the depth of t in the vision tree 0: depth(text) /max depth of vision tree <= 1/3 1: 1/3< depth(text) /max depth of vision tree <= 2/3 2: others
#ofSibling	{0,1} 0: the number of sibling nodes of the text in vision tree is equal to 1 1: others
SiblingType	{0,1} 0: the types of sibling nodes are mixed by inner nodes and leaf nodes 1: others

NextNodeType	{0,1}: Let DFS sequence be the sequence of vision tree traversed by Depth First Search (DFS) algorithm 0: the type of the next node in the DFS sequence is the sibling of the text 1: others
LastNodeType	{0,1} 0: the type of the last node in the DFS sequence is the sibling of the text 1: others
NextIsLi	{0,1} 0: the next node in the DFS sequence is marked by in HTML 1: others
LastEndColon	{0,1} 0: the text of last node in the DFS sequence ends with colon 1: others
LastLabel	{0,1,2,3} 0: the first node of DFS sequence 1: the last node in DFS sequence is labeled by "C" 2: the last node in DFS sequence is labeled by "I" 3: the last node in DFS sequence is labeled by "N"
Concept	{0,1} The ratio of words described concept in the text is less than 0.1: 0 others: 1
Country	{0,1} The ratio of words described country in the text is less than 0.1: 0 others: 1
Month	{0,1} The ratio of words described month in the text is less than: 0 others: 1
Total	{0,1} The ratio of all keywords in the text is less than 0.5: 0 others: 1