**Multi-document Summarization by Graph Search and Matching**

According to this paper, it describes a method to sum up similarities and difference in a pair of related documents using graph representation for text.The Word,Phrases, Proper names are represented positionally as nodes and the edges represent the semantic relation between them.The algorithm ﬁrst uses a **spreading activation technique** to discover, in each document, nodes semantically related to the topic. The activated graphs of each document are then matched to yield a graph corresponding to similarities and diﬀerences between the pair, which is rendered in natural language.

The main aspect here is the representation of texts. The World Wide Web comprises of millions of informations which are being uploaded day by day. There is a need to summarise and analyse these contents in a meaning full manner.In this method this is done by representing the texts in a unique manner with the help of nodes and the relation via vertices.A Specific reference document is selected with which we ill analyse and search for the summary from other documents.

This method employed here is Spreding Activation Technique.It differs from Neural Nets. In this technique there is mainly two aspects

. First, the structure of our graph reﬂects both semantic relations derived from text as well as linear order in the text (the latter via the positional encoding); the linear order is especially important for natural language.

Second, the set of nodes which become highly activated is a function of link type and distance from entry nodes, unlike other approaches which use a ﬁxed bound on the number of nodes or convergence to a stable state.

**Tools for Building Document Graphs**

The experiments make use of a sentence and paragraph tagger which contains a very extensive regular expression-based sentence boundary disambiguator . The boundary disambiguation module is part of a comprehensive preprocess pipeline which utilizes a list of 75 abbreviations and a series of hand-crafted rules to identify sentence boundaries. Then, the Alembic part-of-speech tagger is invoked on the text. This tagger uses the rule sequence learning approach of (Brill 1994)3. Names and relationships between names are then extracted from the document using SRA’s NetOwl (Krupka 1995), a MUC6-ﬁelded system. Then, salient words and phrases are extracted from the text using the tf.idf metric, which makes use of a reference corpus derived from the TREC (Harman 1994) corpus.

Some mathematical equations and concepts are used here to calculate the ‘weight’.

**Graph search**

Document nodes whose strings are equivalent to topic terms (using a stemming procedure =stem) are treated as entry points into the graph. The weight of neighboring nodes is dependent on the type of node link travelled. For adjacent links, node weight is an exponentially decaying function of activating node weight and the distance between nodes. Distances are scaled so that travelling across sentence boundaries is more expensive than travelling within a sentence, but less than travelling across paragraph boundaries

The Summarisation between two documents are done via FDS (Find-Similarities-and-Diﬀerences) Graph algorithm.

**Effectiveness**

Two methods are followed here

The ﬁrst is an extrinsic evaluation in which the quality of the summary is judged based on how it eﬀects the completion of some other task. The second approach, an intrinsic evaluation, judges the quality of the summarization directly based on user judgements of informativeness, coverage etc.

The summarization exploits the results of recent progress in information extraction to represent salient units of text and their relationships. By exploiting relations between units and the perspective from which the comparison is desired, the summarizer can pinpoint similarities and diﬀerences.