System	ROUGE-1	ROUGE-2	ROUGE-SU4
OurSystem	**0.42145**	**0.11016**	**0.15632**
MultiMR	0.41967	0.10302	0.15385
RankBSU	0.39123	0.08742	0.14381
TTG	0.39268	0.09645	0.14553
AveDUC	0.39684	0.09495	0.14671
NIST Baseline	0.33126	0.06425	0.11114

Table 2. Comparison results (F-measure) on DUC 2007 under ROUGE evaluation.

System	OurSystem	MultiMR	RankBSU	TTG
Pyr (Th:0.6)	**0.858**	0.845	0.832	0.834
Pyr (Th:0.65)	**0.743**	0.731	0.718	0.721

Table 3. Comparison results on DUC 2007 under the automated pyramid evaluation with two threshold value 0.6 and 0.65.

the order of the two nodes is sequentially close to each other, which can be formulated as follow:

$$\begin{aligned} & u_i - u_j + n x_{ij} \le n-1 \quad 1 \le i \ne j \le n \\ & 1 \le u_i \le n \qquad\qquad i = 1,...n \\ & u_i \in \mathbb{Z} \qquad\qquad i = 1,...n \end{aligned} \tag{4}$$

At last, we can formulate the objective function as follow:

$$\max \ 1/n \sum_{i=1}^{n} \sum_{j=1, j \ne i}^{n} R_{ij} x_{ij} + \lambda \sum_{i=1}^{n} w_i u_i \tag{5}$$

where parameter λ tunes the effect of the two parts and n is the quantity of *BSU*s in the final *BSU* semantic link network (after reduction).

Sentence Generation. After the summary structure has been planned, sentences are generated for each node in the *BSU* semantic link network. As the *BSU* contains enough semantic and syntactic information, sentence can be generated efficiently according to the following rules:

- Generate a Noun Phrase (NP) based on the actor argument to represent the subject, a NP based on the receiver argument to represent the object if present.
- Generate a Verb Phrase (VP) based on the action verb to link the components above. The tense of the verb is set to the same as in the original sentence, and most modifiers like auxiliaries and negation are conserved.
- Generate complements for the VP when the *BSU* has no receiver. The verb modifiers following the action verb such as prepositional phrases and infinitive phrases can be used as the complement, in case that the verb would have no interesting meaning without a complement.

The process of sentence generation for each node is based on the syntactic structure of the source sentence where the *BSU* is extracted from. The time and location preposition phrases which are important information of new events are kept. The generated sentences are organized according to the summary structure. If some adjacent sentences in the summary have the same subject, the subject of the latter can be substituted by a pronoun (such as it or they) to avoid repetition of noun phrases. One sample summary generated by our system for "Malaysia MH370 Disappear" news is shown in Figure 1.

4 Evaluation Results

4.1 Dataset and Experimental Settings

In order to evaluate the performance of our system, we use two datasets that have been used in recent multi-document summarization shared tasks: DUC2005 and DUC2007. Each task has a gold standard dataset consisting of document clusters and reference summaries. In our experiments, DUC2005 was used for training and parameter tuning, and DUC2007 was used for testing. Based on the tuning set, the parameter λ is set as 10 and δ is set as 0.7 after tuning.

Our system is compared with one state-of-the-art graph-based extractive approach *MultiMR* (Wan and Xiao, 2009) and one abstractive approach *TTG* (Genest and Lapalme, 2011). In addition, we have implemented another baseline *RankBSU* which uses the graph-based ranking methods on the *BSU*s network to rank *BSU*s and select the top ranked *BSU*s to generate sentences.

4.2 Results

ROUGE-1.5.5 toolkit was used to evaluate the quality of summary on DUC 2007 dataset (Lin and Hovy, 2003). The ROUGE scores of the NIST Baseline system (i.e. *NIST Baseline*) and average ROUGE scores of all the participating systems (i.e. *AveDUC*) for DUC 2007 main task were also listed. According to the results in Table 2, our system much outperforms the *NIST Baseline* and *AveDUC*, and achieves higher ROUGE scores than the abstractive approach *TTG*. So the abstract representation of texts and the information extraction process in our system are effective for multi-document summarization. Our system also achieves better performance than the baseline *RankBSU*, which demonstrates that the network reduction method is more efficient than the popular graph-based ranking methods. As compared with the state-of-art graph-based extractive method *MultiMR*, our system also achieves better performance. Furthermore, our system is abstractive with abstract representation and sentence generation. Incorrect