

Construction and application research of knowledge graph in aviation risk field

Qian Zhao¹, Qing Li¹ and Jingqian Wen²

¹ School of Mechanical Engineering and Automation, Beihang University, Beijing 100191, China

² School of Mechanical Engineering, Beijing Institute of Technology, Beijing 100081, China

Abstract. Since the causes of aviation accidents and risks are complicated, concealed, unpredictable and difficult to be investigated, in order to achieve the efficient organization and knowledge sharing of the historical cases of aviation risk events, this paper put forward the method of constructing vertical knowledge graph for aviation risk field. Firstly, the data-driven incremental construction technology is used to build aviation risk event ontology model. Secondly, the pattern-based knowledge mapping mechanism, which transform structured data into RDF (Resource Description Framework) data for storage, is proposed. And then the application, update and maintenance of the knowledge graph are described. Finally, knowledge graph construction system in aviation risk field is developed; and the data from American Aviation Safety Reporting System (ASRS) is used as an example to verify the rationality and validity of the knowledge graph construction method. Practice has proved that the construction of knowledge graph has a guiding significance for the case information organization and sharing on the field of aviation risk.

1 Introduction

With the rapid development of the air transport industry, the modern society's demands for the aviation safety have been continually improved [1]. The contingency and huge losses of aviation accidents have made people pay close attention to the analysis and summary of historical cases. In the 1970s, nearly ten countries and regions, led by the United States, established the Aviation Safety Reporting System [2] to record air risks and accident cases in the form of reports. The data of reports are from pilot, air traffic controller, dispatcher, cabin crew, maintenance technicians and others. ASRS encodes case reports from the aspects of aircraft, environment and incident evaluation, and store the valuable data in relational database to realize the structured organization and persistent storage of the aviation risk and accident case information. As the world's largest information sources in aviation safety, ASRS in America has formed more than 1.3 million reports after nearly 40 years of accumulation. Especially in recent years, up to an average of 1,774 per week rapid growth.

However, the flat storage form based on relational database can't intuitively show the complex network relations between aviation cases, and also the keywords-based retrieval method is not easy for us to query and use the multi-dimensional information, so that the historical cases accumulated in ASRS system can't be well utilized, and the knowledge contained is difficult to be effectively explored. That is why there are many historical cases, but failed to be fully used to avoid similar incidents. We take

the ice accident as an example, an air police-200 early warning aircraft from the Chinese Air Force, which is modified by the Yun-8AEW, crashed in the eastern part of Anhui province on June 3, 2006 due to aircraft wing icing [3]. The situation exists a similarity with the US Eagle Air Flight 4184 air crash [4] on October 31, 1994. This shows that ASRS system has laid a good foundation for information collection and sharing to avoid aviation risks and analyse aviation accidents. While how to make more effective use of the large number of historical cases accumulated by ASRS, how to gain experience and lessons from it and how to improve security rules to provide auxiliary analysis support to avoid similar risks have been the key issue which need to be solved urgently.

2 Related work and technology

As an effective tool for describing massive knowledge entities, attributes and relationships between them, knowledge graph has been widely used by academics and industries since its successful application to search engines by Google in 2012 [5]. The knowledge graph can be divided into general field and vertical field according to its knowledge range. General field knowledge graph is characterized by large amount of data, strong versatility and wide range, but vertical field knowledge graph has the characteristics of relatively small amount of data, high quality, high density, high concentration, high authority and high knowledge accuracy [6].

At present, the research of general knowledge graph at home and abroad are developing rapidly. There are

Probase [7], DBpedia [8], Freebase [9], Baidu Zhixin, Sogou knowCube, CN-DBpedia [10] and so on. Foreign vertical field knowledge graph involved the GeoNames in geographical field, DBLife [11] in academic field, UniProtKb [12] in biological field and so on, but the domestic vertical field knowledge graph is limited to academic research. For example, Ruan Tong et al. [13] proposed a data-driven incremental vertical knowledge graph construction method; Li Wenpeng et al. [14] proposed a software knowledge graph construction method for open source software projects; Ge Bin et al. [15] proposed a method and a computational framework for military knowledge graph construction.

In the field of aviation risk, the researchers mostly analyse the impact of a factor on aviation accidents, or use an algorithmic theory to analyse aspects of aviation risk [16, 17]. However, aviation risk and accidents are caused by results of multiple factors, so we need to consider the mutually influences of aircraft, environment, human factors and other aspects in the process of case analysis. Knowledge graph presents a great advantage in building a knowledge network. It provides a new means of acquiring, storing, organizing, managing, updating and displaying [18] for knowledge which contains complex relationships and it also can supply a more knowledgeable approach to cognitive habits.

3 Construction of aviation risk knowledge graph

As a complex event, aviation accident may be influenced by the aircraft performance, crew contingency, bad weather, geographical conditions and other factors. The various influencing factors and the event itself constitute a complex network relationship for each aviation risk event, and the same event set make up a huge knowledge

network system, so two-dimensional relationship table can't show the complex relationships at all. Knowledge graph, as a direct representation of the relationship, is more conducive to our effective analysis of the specific potential risks existing in those complex relationships.

In order to fully organize and share the aviation risk knowledge, this paper combines the characteristics of aviation risk domain and constructs the aviation risk knowledge graph, the graph means a knowledge semantic network which is used to describe the whole event, it is composed of entities that make up the aviation risk event and the associations between these knowledge entities. Among them, the knowledge entity refers to the fact that is distinguishable, identifiable and owing some definite semantic meanings in the aviation risk event. The associations between knowledge entities refer to a certain type of binary relationship between knowledge entities.

The construction process of aviation risk knowledge graph includes domain ontology modeling, instance-to-ontology mapping, visualization analysis and application maintenance. Firstly, establish the ontology model of aviation risk event by combing the top-down and bottom-up methods, which is used as a data model to define the knowledge graph and describe the concepts and relationships between them. Secondly, transform the relational data of ASRS into knowledge entities, attributes and relationships in the knowledge graph by using the pattern-based knowledge mapping mechanism. Finally, carry out the analysis of data relations, and achieve the visualization and maintenance of knowledge graph. The construction process is shown in Figure 1.

Based on the above mentioned construction process, the main contents of this paper are divided into three parts: representation of knowledge, data mapping, management and application of knowledge graph.

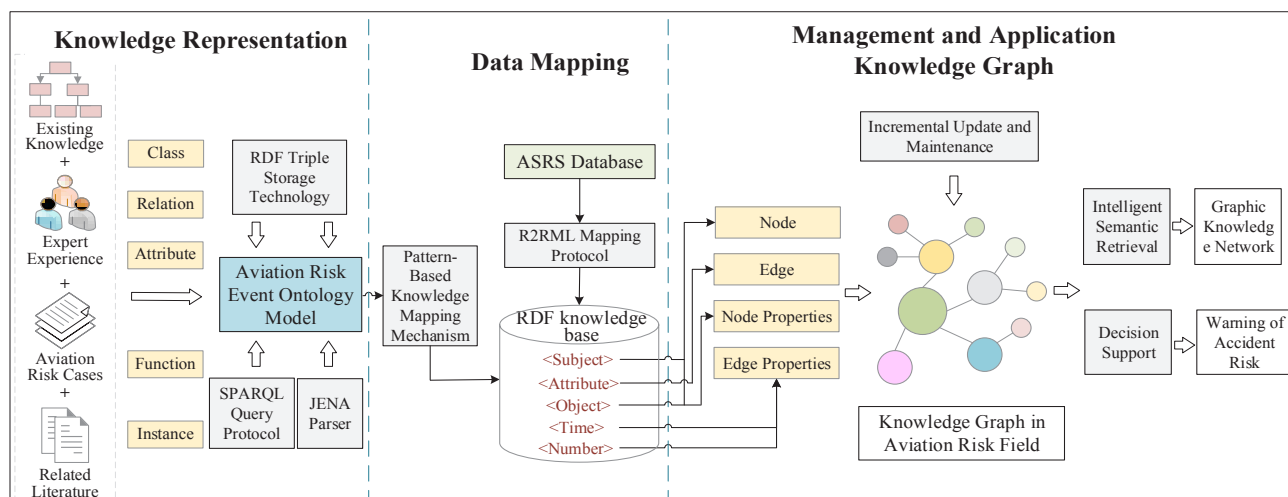


Figure 1. Overall structure diagram of the knowledge graph construction in aviation risk domain.

3.1 Knowledge Representation

Knowledge in aviation risk cases has rich hierarchical structure and complex logical relationships. The premise of knowledge graph construction in aviation risk domain

is to classify, summarize and standardize the knowledge, and to construct an effective and extensible knowledge model. The knowledge representation method based on ontology [19] which works as a clear and normative field concepts, the level and logical relationships between

concepts, the attributes and constraints of concepts, etc. The method can guarantee uniqueness and no ambiguity of understanding in the transfer and sharing process of knowledge, so as to become the important technical method for product domain knowledge modeling.

Because of the complexity of aircraft structure and the diversity of aviation risk events, the existing aviation accident knowledge base is not really suitable in China.

Thus this paper focuses on aviation risk case and builds the aviation risk knowledge ontology framework. It use data-driven incremental ontology modeling method to further extend the concepts and relationships of the module details. First, the aviation risk event ontology model is defined as $O = \langle C, R, A, I, F \rangle$ quintuple, the meaning of each letter in the fomular is as shown in Table 1.

Table 1. Meaning of ontology model definition.

Letter	Meaning of letter
O	Ontology - Refers to unified and standardized expression of the aviation risk knowledge.
C	Class - Refers to a collection of entities.
R	Relation - Refers to a logical relationship between concepts, including composition, inheritance and other business relationships.
A	Attribute - Refers to the concept (or instance) attribute relationship.
I	Instance - refers to the basic elements or the specific entity of the class.
F	Function - A function or axiom used to represent an association or constraint that exists between relationships or functions.

According to the above definition, the knowledge of aviation risk cases is systematically classified and organized to form a network of knowledge structure. Each concept contains one or more instances. The instance inherits all attributes of the concept. There is cross concepts or association relationships between instances. Figure 2 shows a partial of the aviation risk event ontology model.

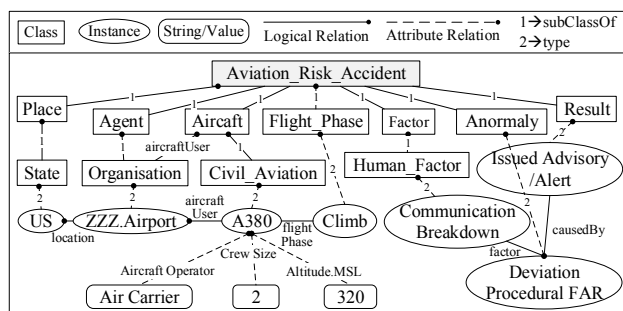


Figure 2. Example of partial aviation risk event ontology model.

After establishing the ontology model, the next step is to store the knowledge in the database. In this paper, we use Jena, a ontology parsing java toolkit, to transform the ontology metadata into the resource description

framework RDF [20], then store and query knowledge in the form of <Subject-Attribute-Object>.

3.2 Data Mapping

The knowledge source of this paper is the structural data of the American Aviation Safety Reporting System (ASRS), whose data have high quality, high reliability and very analytical value. This paper stores the relevant data collected in the relational database and uses the pattern-based data mapping mechanism to complete the transformation structured knowledge, that is, RDB2RDF data conversion process. W3C introduced two mapping language standards in 2012 [21]: Direct and R2RML, where Direct Mapping uses the mapping mechanism that directly output the relational database table structure to RDF graph, but R2RML achieve the transformation through a custom vocabulary table, the latter approach is more customizable and more flexible. Therefore, based on the pre-built aviation risk event ontology model, this paper uses the custom R2RML mapping language to build up pattern-based data mapping.

R2RML mapping [22] design a logical table, it retrieve data from the relational database. We will define a SQL query of table in relational database as a logical table. Each logical table is converted into RDF data by a triples map, that is, each row of instance data in the logical table is mapped to several RDF triples. R2RML mapping mechanism expression is:

$$triplesMap = \{logicalMap, subMap, preobjMAP\} \quad (1)$$

Three elements in the formula are: (a) logicalMap: mainly describe the name of database table. (b) subMap: the common subject of all the RDF triples corresponding to one logical row, is used to generate the subject of RDF triples. (c) preobjMap: each mapping consists of a predicateMap and an objectMap or a valueMap, which is used to make the predicate and object of RDF triples.

The whole R2RML mappings of a relational database constitute a mapping document, which consists of a series of RDF triples. It can be described as:

$$mapdoc = \{tripleMap1, tripleMap2, K, tripleMapN\} \quad (2)$$

According to logical relationships between concepts in the aviation risk ontology model, the R2RML mapping document is written based on this expression. The mapping principle is illustrated by Figure 3. It contains three parts, the the upper part means the relational database table (wing_icing1 table) and the corresponding attribute columns and tuple row data. The middle part is the definition of logical relations. The lower part is the RDF data generated by the mapping, expressed as a triple.

On the basis of this mapping, this paper adds <Time> (record the time of event occurrence to facilitate the timing analysis) and < Number > (record number of occurrence to make it easy for data analysis) in each triple. Finally the data storage pattern becomes the form of <Subject-Attribute-Object-Time-Number> quintuple.

Input <Relational database tables>	ACN	Organization	Aircraft	Flight Phase	Anomaly	Handling
	873164	ZZZ.Airport	A380	Final Approach	Deviation-Procedural Published Material/Policy	Air Traffic Control Separated Traffic
	955328	SFO.Tower	A350	Taxi	Inflight Event	Air Traffic Control Issued New Clearance
R2RML mapping document	<pre> @prefix map: <#>. @prefix rr: <http://www.w3.org/ns/r2rml#>. @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>. @prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>. @prefix owl: <http://www.w3.org/2002/07/owl#>. @prefix base:<http://www.owl-ontologies.com/2017/ASRS.owl#> map:handle rr:logicalTable [rr:tableName "wing_icing1"]; rr:subjectMap [rr:template 'base:{"Anomaly"}'; rr:class base:Anomaly;]; rr:predicateObjectMap [rr:predicate base:handlingMethod; rr:objectMap [rr:column '{"Handling"}']];]; rr:logicalTable [rr:tableName "wing_icing1"]; rr:subjectMap [rr:template 'base:{"Aircraft"}'; rr:class base:Aircraft;]; rr:predicateObjectMap [rr:predicate base:aircraftUser; rr:objectMap [rr:column '{"Organization"}']]; rr:predicateObjectMap [rr:predicate base:flightPhase; rr:objectMap [rr:column '{"Flight Phase"}']];]; </pre>					
Output <RDF Triple>	base: A380	base: A380	base: Deviation-Procedural Published Material/Policy	base: institution	base: flightPhase	base: ZZZ.Airport base: Final Approach base: Air Traffic Control Separated Traffic

Figure 3. Partial R2RML mapping schematic of entities

3.3 Management and application of knowledge graph

The essence of knowledge graph is a semantic network of entities and relationships between them. The ontology modeling in 3.1 defines the basic data pattern for knowledge graph. 3.2 greatly enriches the relationships between entities by realizing the mapping from structured data to RDF. So the knowledge are finally stored in RDF triples and quintuple, respectively.

The RDF storage pattern includes two types of tuple from the ontology mapping and data mapping. We extract the <Subject> and <Object> as entities in the knowledge graph, and extract <Attribute>, <Time>, and <Number> as the attribute and the association in the knowledge graph, so we can construct the knowledge graph.

In the construction of general field knowledge graph, there is a need for knowledge fusion, that is entity links and knowledge merging to eliminate the concept ambiguity and misconceptions. In contrast with it, the knowledge graph construction in the aviation risk field does not require subsequent processing, because the data source is pre-processed structured data and the manually constructed ontology model also ensures the high quality and uniqueness of the knowledge. With the increasing number of data, this paper proposes a data-driven incremental ontology modeling technology, it expand concepts and instances on the basis of the original ontology. Concept refers to a class with the same entities, the changes of concepts and attributes will not be particularly frequent, so we mainly pay more attention to the automatic update of instances. The richness of relationships between entities depends on the pattern-based data mapping mechanism. Due to the consistency of the database table structure, the pattern rules can be reused without frequent modification after being defined.

The application of knowledge graph in aviation risk field is mainly in two aspects: (1) intelligent semantic retrieval. The knowledge graph makes the large number of hidden knowledge and associated data to be materialized and explicit, it organizes and stores all

concepts, entities, attributes and relationships in the form of graph structure. We analyze the user query statement by using knowledge graph, get the keywords and map it to one concept or a group of concepts of knowledge graph. Then show users the graphic knowledge network about keywords. (2) decision support. It is critical to use the historical experience data on aviation risk events efficiently to obtain valuable information. In contrast to the flattened information, the graphic structure of knowledge will focus on the effective organization and association of the entire event at the semantic level, that is more direct to guide personnel and more conducive to follow-up knowledge reasoning, thus supply deeper level of decision support for the enterprise or government.

4 Implementation of aviation risk knowledge map

Based on the construction method of aviation risk field knowledge graph proposed in this paper, the effectiveness of the method is realized and verified by concrete practice.

In the first place, the model of the aviation risk event ontology is established by using the ontology construction tool Protégé. The ontology model regulates the data pattern of the knowledge graph, which contains 230 concepts, 65 logical relations between concepts and 85 attribute relations between concepts.

Secondly, obtain structured data from the [ASRS Database Online] in ASRS. Use "icing" event as an example to illustrate the data acquisition method. We input "icing" in the field of [Narrative] and [Synopsis], and click [Run Search] button, then we can get a total of 3028 cases. Then export the results to excel tables and import them into Oracle database as the data source of the knowledge graph. In the end, based on the rules defined in the manually defined R2RML mapping document, use the open source r2rml-parser tool to implement the mapping from relational data to RDF quintuple.

Finally, develop the aviation risk field knowledge graph construction system by using the java development tool Eclipse IDE and the show middleware Dorado7. The

system consists of two parts: knowledge graph visualization module and knowledge graph update module. The knowledge graph visualization module uses ontology-based RDF query and Gojs foreground visualization javascript plugin to model the knowledge graph, and it can achieve the classification modeling for key concepts. In the meantime, it use ontology-based word segmentation technology and SPARQL-based semantic retrieval technique to realize intelligent search function of knowledge graph for natural query statement. The concrete implementation of this module is shown in Figure 4. The user enters the query statement in the search box, the system will pick out the keywords by using the word segmentation technique based on the ontology dictionary, the keywords refer to key knowledge entities with domain significance in the query statement. Then the system return to the user with graphical

knowledge graph related to keywords. For example, the input word is "Icing event occurred in Washington DC ", the system automatically segments the sentence and gets keywords as "Icing", "event" and " Washington DC ", visualize the entities related to these two key words in the form of knowledge graph and display personnel, hardware facilities, environment and influencing factors of the ice accident in Washington DC. Known from the analysis of figure 4, three top airports in Washington DC which exist icing accidents are: DCA Airport, IAD Airport and ZDC Airport and the main reason leading to the icing accident in Washington DC is aircraft equipment failure. So as to remind relevant maintenance technicians that they must be responsible for equipment maintenance, test and security work to avoid the occurrence of similar incidents before the flight departure.

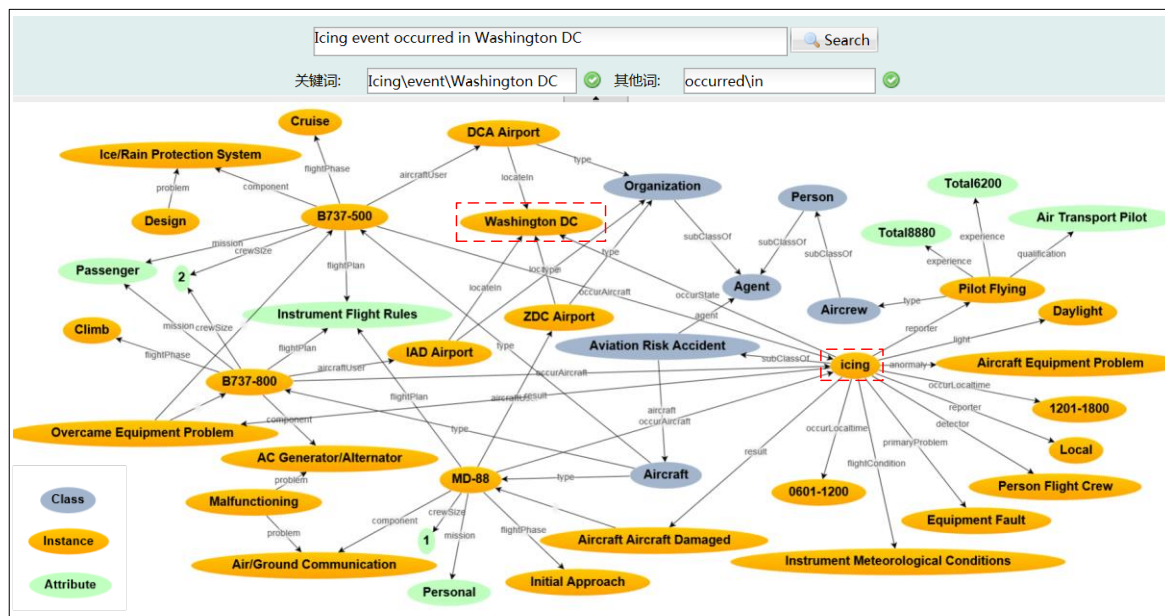


Figure 4. Example of intelligent retrieval based on knowledge graph.

The knowledge graph update module is to realize the update and maintenance of ontology concepts and conceptual attributes. This module includes the ontology hierarchy tree management, ontology file import and export management, ontology data statistics, ontology knowledge editing and knowledge attribute management.

5 Conclusion

Aim to organize and share aviation risk history cases effectively, this paper put forward the knowledge graph construction method in the field of aviation risk, constructed the aviation risk ontology model, proposed the knowledge acquisition method based on structured data, realized the mapping mechanism from the relation data to ontology, developed an aviation risk knowledge map construction system and verified the effectiveness and rationality of the method by certain cases. The application of aviation risk field knowledge graph can effectively improve the efficiency of knowledge retrieval,

improve the way of knowledge organization, and help users to make decision through knowledge relationships.

It also has high application value in the knowledge mining and knowledge sharing.

References

1. J. Liu. Civil Aviation University of China. 7(2007)
2. J. Tian. Civil Aviation University of China. 10(2010)
3. J. Zhang. China Air Force 6·3 air crash: airplane ice caused the plane crash dead[DB/OL]
4. Wikipedia, the US Airways flight 4184 air crash[DB/OL].https://en.wikipedia.org/wiki/American_Eagle_Flight_4184. 2017.06.20
5. F. Hu. East China University Of Science And Technology. 2(2015)
6. T. Li, C. Wang, H. Li. Journal of Nanjing University of Science and Technology. **41**, 30 (2017)
7. W. Wu, H. Li, H. Wang, KQ. Zhu. ACM. 481-492(2012)
8. J. Lehmann. Semantic Web. **6**(2),188(2015)

9. K. Bollacker, C. Evans, P. Paritosh, T. Sturge, J. Taylor. *SIGMOD 2008*. 1247(2008)
10. B. Xu, Y. Xu, J. Liang, C. Xie, B. Liang, W. Cui, Y. Xiao. CN-DBpedia. *Advances in Artificial Intelligence: From Theory to Practice*. 430(2017)
11. P. Derose, W. Shen, F. Chen, Y. Lee, D. Burdick, A. Doan, R. Ramakrishnan. *CIDR 2007*. 169(2007)
12. U.P. Consortium. *Nucleic Acids Research*. 37(2009), (Database issue):D190
13. T. Ruan, M. Wang, H. Wang, H. Hu. *KM Forum*. **3**(2016)
14. W. Li, J. Wang, Z. Lin, J. Zhao, Y. Zou, B. Xie. *Journal of Frontiers of Computer Science & Technology*. **11**(6), 851(2017)
15. B. Ge, Z. Tan, C. Zhang, W. Xiao. *Journal of Command and Control*. **2**(4), 302(2016)
16. K.P. Das, A.K. Dey. *JSM*. (2016)
17. W.R. Knecht. *Accident Analysis and Prevention*. **60**(60C), 50(2013)
18. Y. Wu, A. Yin, K. Lin, X. Yu, G. Lai. *Journal of Fuzhou University(Natural Science Edition)* . **45**(3) 329(2017)
19. M. Zhao, Y. Du, H. She, J. Zhang, H. Wang, Y. Chen. *Transactions of the Chinese Society for Agricultural Machinery*, **47**(9), 278(2016)
20. L. Xiang. *Wuhan University of Science and Technology*. 9(2013)
21. Q. Liu, Y. Li, H. Duan, Y. Liu, Z. Qin. *Journal of Computer Research and Development*. **53**(3), 583(2016)
22. S. Zhou, Z. Xu. "<http://www.paper.edu.cn>". (2013)