ISBN 9781495167324

IJRI

International Journal of Research and Investigation

Volume 1, Issue 2
July 2015

Issue Index

Content

Full Paper:

Design of the Computer-Assisted Translation Program for Chinese and Korean

Yimeng Wang

Jilin Huaqiao University of Foreign Languages

The widely used technology of computer-aided translation by using computer software benefits our modern life for the language translation. The typical languages we studied in this work are Chinese and Korean, including the spell, grammar, word order, collocation and language habits. In this paper, we used and discussed a translation program based on C++ language. The translation program mainly had the parts including the texts translation, the collocation correction, order adjusting, language habits and grammar-based modification processes. When the Chinese was translated to the Korean, the honorific language system and the different order of sentence had to be considered in the program. When the Korean was translated to the Chinese, the collocation of words had to be considered. The target texts were readable and understandable after translation from source texts.

Key words: computer-aided translation, C++ language, Chinese, Korean

1. Introduction

The technology of computer-aided translation (CAT) has been utilized in our modern life to assist the translation process by using computer software. It connects the human work with the computer software (Sergio 2009). However the available CAT technologies today cannot result good translations when translating one language to another, because the language habits exist everywhere in each different culture. As important as the dictionary-based translations which such programs store a large

* I am very much indebted to Professor Yan Yin, who read and summarized this paper for me. Much The research was conducted during periods of research at Korea University, Seoul, South

number of source and target databases and change the source texts to the targets texts, the correction based on language habits is also utilized to improve the CAT. After the corrections including checking spell and grammar, adjusting word order and optimizing collocation, the output texts become more readable and understandable. Besides, the language habits of users can improve the correction effectiveness, by assisting the dictation program of translation methods and the edition of recognized texts (S. Khadivi 2005; S. Khadivi 2006).

On the other hand, the language habits have a very important role in translation, especially when our society becomes more and more international. The typical languages we studied in this work are Chinese and Korean. The spell, grammar, word order and collocation are quite different for the two kinds of language. The Chinese is a branch of the family of Sino-Tibetan language, while the Korean belongs to the family of putative Altaic language (Sanchez 2008). However, the Chinese is also a part of the language which has been used in Korean history, during when China and Korea have a very close relationship. Even for over a millennium in the past, modified Chinese characters which was called Hanja was used in Korea until the Hangul got widespread use in the 20th century, which first commissioned by the Great Sejong for the national writing system. Based on the similarities and differences of the two kinds of languages habits, we aim to improve of the CAT technology in order to explore a smarter translation system.

2. Matters of method

2.1. Database of languages

The computer language we used to store the database and process the translation was C++ language (X. Zhang 2010; Shi 2011). The database structure included the word structure and grammar structure. The word structure concluded several attributes, including the number (sequence), the spelling (sometimes one word has several kinds of spelling), characteristic (such as verb, preposition, conjunction, preposition, adverb, noun, adjective), collocation rule (in order to find a better meaning when translated to target text), language habits (cross-checking offline and online to decide which meaning

Korea, from September of 2014 to June of 2015. During the time studying in South Korea, I have cooperated and completed this work.

is the optimum) and target meanings (one word generally has more than one meaning when translated to another language). The basic database structure of one source language is shown as:

- Number
- Spelling
- Characteristic
- Collocation rule
- Language habits
- Target meaning 1
- Target meaning 2
- ….

Then we input all the attributes of each word into the database. For example, shown in figure1, 'I' in Chinese has six basic attributes: '100001' (sequence in the database); '我 (wǒ/I)' (the spelling); 'noun' (the characteristic), '我 (wǒ/I)' +verb or verb+'我 (wǒ/I)' (the collocation rule); 'www.baidu.com' (Chinese domesticated website of search engine), 'www.google.com' and the offline database of reference sentences (the links of database used for online and offline cross-checking); '나 (na/I)', '저 (chŏ/I)' and '제가 (chega/I)' (target texts translated to Korean) with their number in Korean database 200001, 200002 and 200003 respectively. 'Eat' in Chinese has the six attributes including '100002', '吃 (chī/eat)', 'verb', the collocation rule ('吃 (chī/eat)' +subject or object+'吃 (chī/eat)'), the online and the offline databases, '먹다 (meokda/eat)' and '드시다 (deusida/eat or drink)' with database number 200011 and 200012 respectively. 'meal' in Chinese has six attributes as well: '100003', '饭 (fàn/meal)', 'noun', verb+'饭 (fàn/meal)' or '饭 (fàn/meal)'+verb, online and offline databases, '식사 (siksa/meal)' and '밥 (bap/meal)' with database number 200021 and 200022 respectively.

The Korean database was edited by us in the same way. Besides the above vocabulary, many Korean words have different meanings when translated to Chinese. In instance, '저 (chŏ/I)' in Korean database has six basic attributions: '200002', '저 (chŏ/I)', 'noun', '저 (chŏ/I)'+는 (neun) (auxiliary word), 'www.naver.com' (Korean domesticated website of search engine), '我 (wǒ/I)' and '那 (nà/that)' with database number 100001 and 100004 respectively. '드시다 (deusida/eat or drink)' in Korean database has six basic attributions: '200012', '드시다 (deusida/eat or drink)', 'verb', noun+'를 (reul)' (auxiliary word) +'드시다 (deusida/eat or drink)', reference databases, '吃 (chī/eat)' and '喝 (hē/drink)' with database number 100002 and 100005 respectively. '밥 (bap/meal)' in Korean database has six basic attributions: '200022', '밥 (bap/meal)', 'noun', noun+'는 (eul)' (auxiliary word) +verb, reference databases, '饭 (fàn/meal)' and '屑 (xiè/sawdust)' with database number 100003 and 100006 respectively.

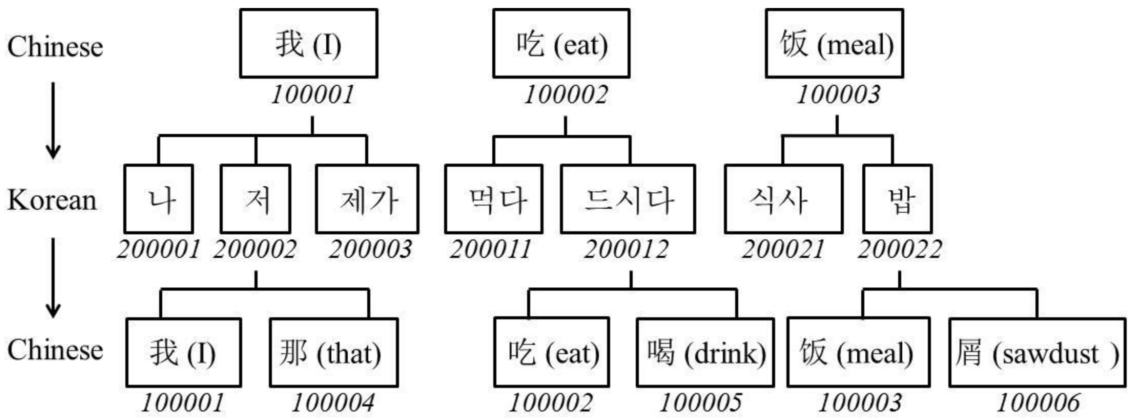

Figure 1. Translation from source texts to target texts for the correction based on collocation rule.

2.2. The structure of transition program

The translation program included the texts translation, the collocation correction, order adjusting, language habits and grammar-based modification processes (Fig. 2). The collocation correction was designed by cross-collocation of the subject, verb, object, auxiliary word and so forth based on the information shown in figure 1. When translate Chinese to Korean, '나 (na/I)'+는 (neun), '저 (chǒ/I)'+는 (neun)and '제가 (chega/I)' were able to collocate with '먹다 (meokda/eat)' and '드시다 (deusida/eat or drink)', and were able to collocate with '식사 (siksa/meal)'+를 (reul) and '밥 (bap/meal)'+는 (eul) as well. However, the language system in Korea has to fit the language habits which would be discuss in the following. The basic collocation when translate Korean to Chinese was very simple. For example, 'I' was able to collocate with 'eat' or 'drink', while 'that' could not. Both 'eat' and 'drink' were able to collocate with 'meal', but not able to collocate with 'sawdust'. Thus after the correction of collocation rule, a better translation for one word would be presented.

As shown in figure 2, after a correction of collocation rule, the order of the sentence would be modified. In Chinese, '我 (wǒ/I)'(subject)+'吃 (chī/eat)'(verb)+'饭 (fàn/meal)'(subject) was a sentence, whereas in Korean '나 (na/I)'(subject)+는 (neun)+'밥 (bap/meal)'(subject) +는 (eul)+'먹다 (meokda/eat)'(verb) was a sentence.

Subsequently, the language habits were checked in the program. Firstly, the honorific language as an important part in Korean, we have to check whether we should use the honorific Korean in a sentence or not when translate language to Korean. Also the sentence '我 (wǒ/I)'(subject)+'吃 (chī/eat)'(verb)+'饭 (fàn/meal)'(subject) has two translation ways to Korean, besides '나 (na/I)' (subject) +는 (neun) +'밥 (bap/meal)'(subject) + 는 (eul)+ '먹다 (meokda/eat)'(verb), the '저 (chǒ/I)'(subject) +는 (neun) +'식사 (siksa/meal)'(subject) +를 (reul)+'먹다 (meokda/eat)'(verb) is the honorific language since '저 (chǒ/I)' and '식사 (siksa/meal)' belong to the honorific vocabulary in Korean. Thus, when the title or position of the speaker and the listener were able to be checked by the system, the target translation texts should be checked whether it fitted the language habits. For example, when the listener got the title 'sir/miss', 'professor' or 'boss', the sentence '나 (na/I)'(subject) +는 (neun) +'밥 (bap/meal)'(subject) +는 (eul)+'먹다 (meokda/eat)'(verb) had to be changed to '저 (chǒ/I)'(subject)+는 (neun)+'식사 (siksa/meal)' (subject) +를 (reul)+'먹다 (meokda/eat)' (verb) which was the honorific language. Secondly, the offline database was used to check whether the translated sentence fitted the language habits as well, such as the reference sentences. If the words collocation fitted the reference sentences, then the translation process was successful. If the collocation was not able to be found in offline database, the online database of search engine was utilized later. However, sometimes some collocations of words were all searchable online. Then the most frequently appeared collocation would be the optimum choice.

Besides, a grammar-based modification was used after the whole program to modify the translated sentences. This program based on grammar knowledge was utilized to add, delete or note some words in the translated sentences when the outputted texts were against the grammar laws.

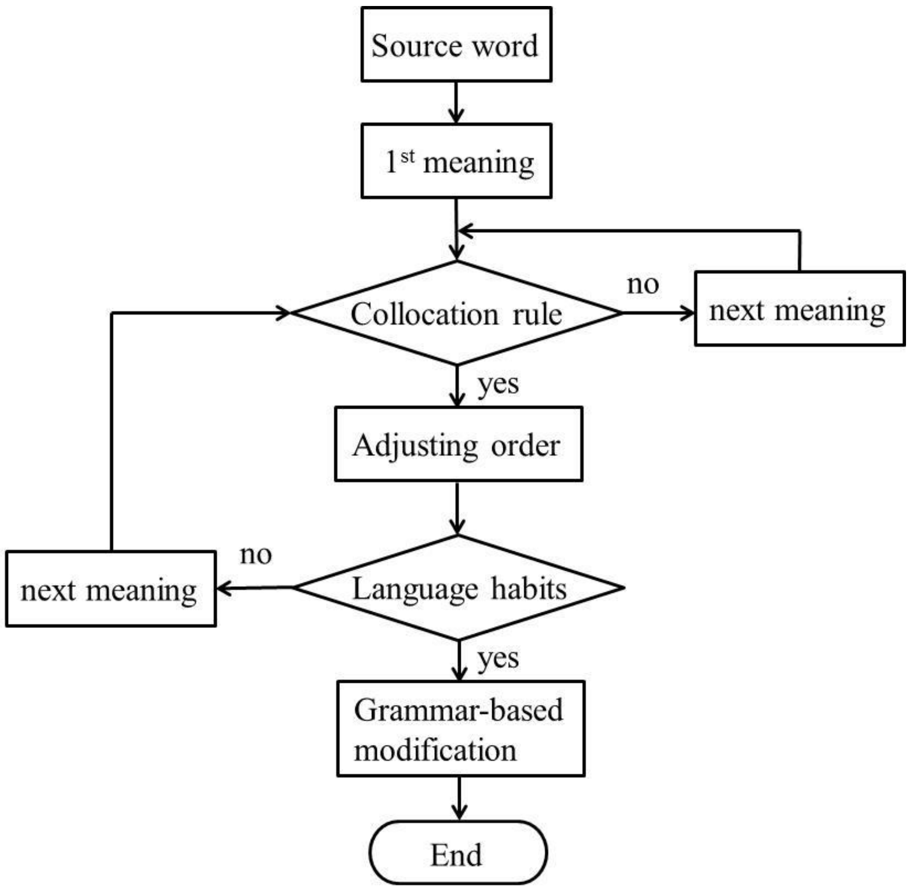

Figure 2. The program flow chart of translation based on C++ language

3. Conclusion

The technology of computer-assisted translation by using computer software is widely used in our international society for the language translation. Typically for the translation between Chinese and Korean, the spell, grammar, word order, collocation and language habits are quite different for the two kinds of language. In this work, we designed and built two dictionary databases and a translation program based on C++ language. The translation program concluded the texts translation, the collocation correction, order adjusting, language habits and grammar-based modification processes. After the translation, the target texts were readable and understandable. Hence via the method we discussed in this work, the CAT technology based on the language habits might be more of interests to linguists and engineers.

References

Sergio Barrachina, Oliver Bender, Francisco Casacuberta, et al. 2009. Statistical Approaches to Computer-Assisted translation. *Computational Linguistics*. 35 (1): 3–28.

S. Khadivi, A. Zolnay and H. Ney. 2005. Automatic text dictation in computer-assisted translation. Paper presented at the European Conference on Speech Communication and Technology (INTERSPEECH 05-EUROSPEECH), September 4-8. Lisbon, Portugal.

S. Khadivi, R. Zens, and H. Ney. 2006. Integration of speech to computer-assisted translation using finite-state automata. Paper presented at the 44th Annual Meeting of the Association for Computational Linguistics and 21th International Conference on Computational Linguistics (COLING/ACL 06), July 22, Sydney, Australia.

Sanchez-Mazas, Blench, Ross, Lin, Pejros, (eds.) 2008. Stratification in the peopling of China: how far does the linguistic evidence match genetics and archaeology? Human migrations in continental East Asia and Taiwan: genetic, linguistic and archaeological evidence. Taylor & Francis Press.

Shi Tong. 2011. C++ Language Programming Technelogy Analysis, *Computer Knowledge and Technology*, 14: 19-22.

X. Zhang & X. Wan. 2010. Analysis and Research on C++ Language Programming Technology, *Computer Knowledge and Technology*, 12: 3282-3284.

An Investigation of the Use of Instrument for Science Researches

S. Tan
Ewa Women's University

Fourier transform infrared (FTIR) spectra were obtained using a Perkin-Elmer Spectrum GX1 fitted with a non-polarized infrared beam and a MIR-TGS detector. Samples were dispersed in KBr pellets, and were analyzed according to 512 individual scans with a resolution of 2 cm-1 at 20 ℃.

For surface morphology and microstructure, field emission scanning electron microscopy (FE-SEM) was carried out using a Hitachi S-4300 operating at 15 kV after deposition of the samples with platinum, and high-resolution transmission electron microscopy (HR-TEM) was collected using a TECNAI G2 20 S-Twin operating at 200 kV after preparation by dispersing the polydiacetylene smallflowers onto a continue Formvar film coated copper mesh grid (PELCO No. 160).

Pore size distribution, surface area, and surface volume were performed from a nitrogen adsorption/desorption using a Micromeritics ASAP 2010 at 77.5 K after the porous starburst polydiacetylene was vacuum-dried at 120 ℃ overnight. These data determined based on Brunauer-Emmett-Teller (BET) theory and (Barrett–Joyner–Halenda) BJH desorption model.

X-ray diffraction (XRD) spectra were obtained using film mode of a Rigaku D/MAX-2500V/PC (Cu anode) at 20 ℃. Samples were evenly dispersed on a 1cm*2.5cm clean glass and vacuum-dried at room temperature for overnight. Typical parameters were set at current 100 mA, voltage 40 kV, step size 0.02o and step time 1.0s.

Bright field and fluorescence images were captured on an Olympus DP50 digital camera connected with an Olympus BX51 microscope. Fluorescence was recorded with an exposure time of 1/500 s in the porous starburst polydiacetylene for confirmation experiment of optical characteristics and of 1/2 s in the polydiacetylene vesicle for detection experiments, and of 1/5 s in the porous starburst polydiacetylene for detection experiments. Fluorescence quantification of the polydiacetylene systems was provided using a photoshop-based image analysis (Adobe Photoshop CS5) in nearly circle pixels (red channel).

Fluorescence Spectrophotometer (Hitachi F-7000, Xe lamp) was used to measure the photoluminescence (PL) performance of smallflowers. By 3-D scan model, the specimen were excited by light form 350~600 nm while detection range was set during 450~750 nm. A wavelength scanning model was used for phase-reversing experiment with the exciting wavelength of 488 nm.

Confocal Raman microscope was used for measuring z-sectioning Raman spectra of smallflowers were measured with a confocal Raman microscope (Renishaw, inVia Raman microscope) and an optical microscope (Leica, DM 2500M). A HeNe laser (Renishaw, RL633) was focused on the samples at 633 nm and detected by a CCD camera (renishaw, RenCam CCD detector). The spectra of z-stack collected from the section (1 mm-interval) using 50x microscope objective.

Confocal Laser Scanning Microscope (CLSM, Carl Zeiss, LSM700) was used for measuring z-sectioning fluorescence images of smallflowers. The samples were excited at 555 nm and detected with a filter of 630-800 nm. Also a projection and spectra of z-stack are collected from section interval of 0.625 mm through the x20 objective lens.

A Study of Bacteria on Saline Wastewater Treatment

Prof. I. Acquah
Korea University

E-mail: asdisaac@korea.ac.kr

1 INTRODUCTION

A large volume of saline organic wastewater is discharged into the environment annually. Many industries, including printing, food processing, medicine, textile, papermaking, chemical, oil, and agricultural pesticide-producing industries, release millions of tonnes of saline wastewater. The waste water is high concentration of salt, typically in the range of 10% - 25% w/v (Li et al. 2008). Besides the civil usage, there has also been an upsurge in amount of seawater used to flush toilets. In 2008, the volume of seawater used to flush toilets in Hong Kong alone reached 2.75×10^8 m^3 and it is expected to increase due to the increased demand from the growing population (Cheng 2010). This saline organic saline water end up in the environment and may cause pollution unless adequately treated. Thus, there is an urgent need to treat saline organic wastewater.

As the demand of treating saline wastewater is very large, lots of strategies have been developed and applied on lab-scales and industrial levels. Many are expensive and not economically prudent to implement on a large scale. A low cost strategy using activated sludge is the most widely used for the treatment of wastewater (Keller et al. 2002). By using the salt-domesticated activated sludge, the wastewater with salt concentration under 10g/L had been treated (An & Gu 1993). Wagner's group reported that halophilic microorganisms could be cultured in activated sludgeand utilized for the breakdown of the wastewater (Wagner et al. 2002). Another approach that had been adopted to treat saline wastewater is by doping the sludge with halophilic microorganisms to become activated sludge after a long-term cultivation. The activated sludge was subsequently used for the treatment of saline wastewater (Kargi & Dincer 1996; Panswad & Chadarut 1999). However, during these domesticating processes, the biological diversity of activated sludge showed a sharp diminishing trend, as a result of the fact that the bacteria in the system could adapt to the saline environment (Lee et al. 2005; Yoshie et al. 2006). The developed activated sludge was not quite suitable for treat saline organic wastewater, especially in the long term. Thus, there is still the need to design and implement new schemes for the long term treatmet of saline wastewater.

As the salinity of seawater is about 3.5%, the tolerance ability of marine microorganisms to salinity is much stronger than the land ones (ref needed). Therefore, the strategy of cultivating the marine microbes' niche, sea mud, to become a marine activated sludge (MAS) is presented in this work. This is relatively cheap as there is no need to buy any special bacteria. The effectiveness of MAS in the treatment of saline wastewater and the tolerance of the microorganisms in the MAS to varying salt concentration and temperature as also presented.

2 EXPERIMENT

2.1 Cultivation of Marine Activated Sludge

Sea mud was mixed with seawater (taken from the Yellow Sea of China, 37.54^0 N, 122.07^0 E) to make a 30 L volume in a tank and the resulting mixing cultivated with nutrition at a temperature of 20±1 ^{0}C for 60 days to become activated sludge. This activated sludge is hereafter referred to as marine activated sludge (MAS). Before the cultivation was commenced, the sludge volume index (SVI) of the sea mud-seawater mixture was 19 mL/g. For the entire duration of the cultivation, the concentration of the dissolved oxygen (D.O.) was maintained at 2.5 mg/L by pumping air into the mixture. The sludge was fed once a day at 8:00 a.m. by the addition of nutrients and replenishing the supernatant with fresh seawater, making sure that the volume of the sludge remained at 30 L. The nutrition added to the culture tank were dextrose monohydrate, starch, ammonium chloride and monopotassium phosphate in concentrations of 0.2064 g/L, 0.1689 g/L, 0.0056 g/L and 0.0028 g/L respectively.

In another preparation, activated sludge, taking from the Third Wastewater Treatment Plant in Weihai, P.R. China, was acclimated for 60 days. The acclimation process is described as follows. The activated sludge was mixed with sea water; forming a volume of 30 L in a culture tank and the SVI and the Mixed Liquor Volatile Suspended Solids (MLVSS) were respectively 112 mL/g and 2.24 g/L. The DO of the sludge culture during the domestication was maintained at 2.5 mg/L. The supernatant was discarded and replaced with seawater and nutrition (the same as was in the case of MAS) added to the suspension at 8:00 a.m. daily for 60 days. The resulting sludge is referred to us as domesticated activated sludge (DAS).

As a control, another sample of activated sludge was prepared similar to the case of DAS but using saline water (NaCl concentration of 3.04% w/v) instead of seawater. The resulting sludge is termed as conventional activated sludge (CAS).

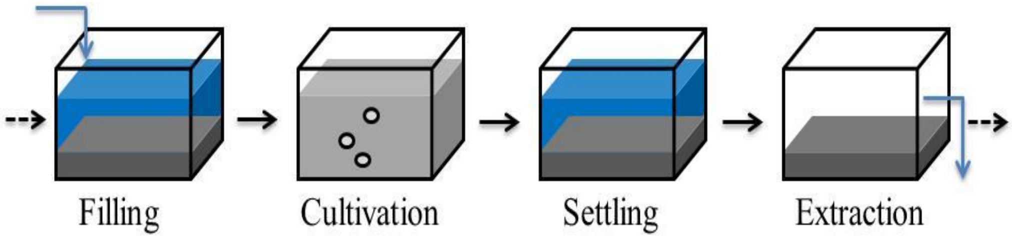

Figure.1 Schematic of cultivation process

2.2 Water treatment experiments

In order to investigate the effectiveness of the sludge in the treatment of saline organic wastewater, parallel experiments were run in three separate sequencing batch reactors (SBR reactors) using the three different types of activated sludge. For each of the experimentation in a SBR reactor, 5 L of the activated sludge with a 36 g/L density was placed in the reactor and 25 L of nutrient-incorporated saline water. The NaCl concentration of the nutrient-incorporated saline water was varied from 0% to 10 % w/v. The final mixture had a chemical oxygen demand (COD) 400mg/L (containing dextrose monohydrate 0.2064 g/L, starch 0.1689 g/L, ammonium chloride 0.0056 g/L and monopotassium phosphate 0.0028 g/L) and the density was 6 g/L. After 24 hours of treatment, we measured the COD degradation rate to evaluate the activity of the sludge, as the changing concentration of NaCl was changed from 0% to 10% w/v. Biological activities were evaluated by comparing the degradation rate of COD_{Mn} for the marine activated sludge (MAS), the domesticated activated sludge (DAS) and the conventional activated sludge (CAS).

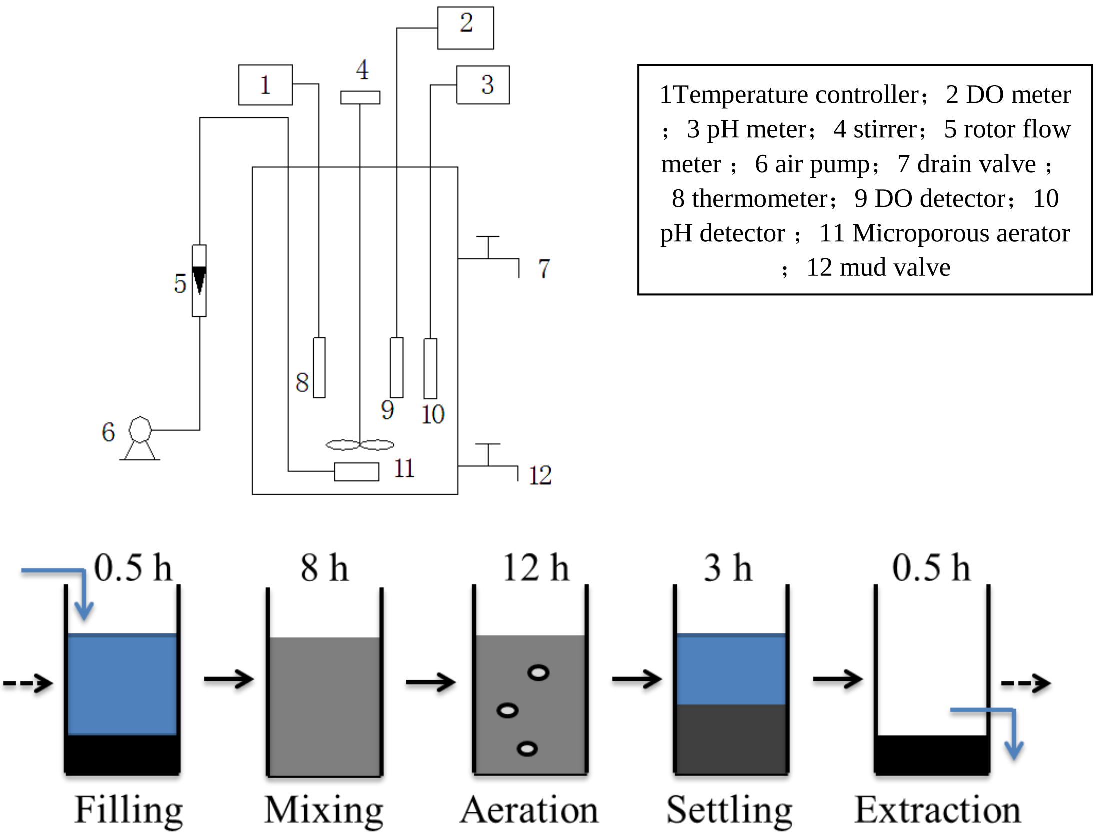

Figure.2 Schematic diagram of SBR reactor

References

Balebona, M. C., et al.

1995 Influence of Salinity and Ph on the Adhesion of Pathogenic Vibrio Strains to Sparus-Aurata Skin Mucus. Aquaculture 132(1-2):113-120.

Chen, M. X., et al.

2014 Impact resistance of different factors on ammonia removal by heterotrophic nitrification-aerobic denitrification bacterium Aeromonas sp HN-02. Bioresource Technology 167:456-461.

Guo, Y., et al.

2013 Heterotrophic nitrification and aerobic denitrification by a novel Halomonas campisalis. Biotechnology Letters 35(12):2045-2049.

Kargi, F., and A. R. Dincer

1996 Effect of salt concentration on biological treatment of saline wastewater by fed-batch operation. Enzyme and Microbial Technology 19(7):529-537.

Lee, J. W., et al.

2005a Operational strategies for an activated sludge process in conjunction with ozone oxidation for zero excess sludge production during winter season. Water Res 39(7):1199-204.

Lee, T. H., et al.

2005b Molecular analysis of bacterial community based on 16S rDNA and functional genes in activated sludge enriched with 2,4-dichlorophenoxyacetic acid (2,4-d) under different cultural conditions. Microbial Ecology 49(1):151-162.

Lu, J., et al.

2014 Impact of salinity on treatment of saline wastewater by sequencing batch biofilm reactor process. Journal of Central South University 21(5):1989-1994.

Panswad, T., and C. Anan

1999 Impact of high chloride wastewater on an anaerobic/anoxic/aerobic process with and without inoculation of chloride acclimated seeds. Water Research 33(5):1165-1172.

Qian, P. Y., et al.

2011 Vertical stratification of microbial communities in the Red Sea revealed by 16S rDNA pyrosequencing. Isme Journal 5(3):507-518.

Uygur, A., and F. Kargi

2004 Salt inhibition on biological nutrient removal from saline wastewater in a sequencing batch reactor. Enzyme and Microbial Technology 34(3-4):313-318.

Wagner, M., et al.

2002 Microbial community composition and function in wastewater treatment plants. Antonie Van Leeuwenhoek 81(1-4):665-80.

Yoshie, S., et al.

2006 Molecular analysis of halophilic bacterial community for high-rate denitrification of saline industrial wastewater. Applied Microbiology and Biotechnology 72(1):182-189.

Zheng, H. Y., et al.
2012 Characterization of a marine origin aerobic nitrifying-denitrifying bacterium. Journal of Bioscience and Bioengineering 114(1):33-37.