



华建集团 HuaJian Group

Research & Development of Multi-lingual Machine Translation and Applications


Dr. Huang Heyan
 Heyan.huang@hjteck.com, www.hjteck.com

Research Center of Computer & Language
 Information Engineering (CCLIE), CAS
 HuaJian Group Co. of CAS, China
 2005年9月


华建集团 HuaJian Group


Contents

- I. Background
- II. Multi-lingual MT Research in CCLIE
- III. Multi-lingual MT System Development
- IV. Series Products and Applications
- V. Future works
- VI. Some considerations


华建集团 HuaJian Group


I. Backgrounds

- 1.1 IMT research team in 2nd Division of ICT
- 1.2 IMT Research Center of ICT, CAS
- 1.3 CCLIE of CAS
- 1.4 HuaJian Group Co. of CAS


华建集团 HuaJian Group


1.1 IMT research team in 2nd Division of ICT

- ❖ 1985-1989 founder of HuaJian Co., Dr. Chen Zhaoxiang began to do MT research as his PhD research topic at the ICT of CAS, and set up intelligent machine translation (IMT) theoretical system.
- ❖ 1988-1990 The research had gotten the support by national Hi-tech program of 863. The IMT/EC prototype system had been completed.
- ❖ 1990-1992 IMT Research obtained 863 program continually support, the practical system IMT/EC-863 had been developed.


华建集团 HuaJian Group

1.2 IMT Research Center of ICT, CAS

- ❖ In 1992, the IMT Research Center of ICT was setup. And a minisize IMT system used in pocket computer "Instant Translator" based on the prototype system of IMT/EC was successfully developed, which brought more than 7.6 million dollars revenue.
- ❖ 1992-1997, Supported by 863 program, we succeeded in realizing a English-Chinese system with high performance in common use and in more than 40 specialties.
- ❖ In 1995, the system IMT/EC-863 was awarded the First Prize of National Advance in Science & Technology, which is the top prize for Science and technology in China.


华建集团 HuaJian Group


1.3 CCLIE of CAS

- ❖ In 1997, the Research Center of Computer & Language Information Engineering(CCLIE) of CAS was founded, which is a research institute directly under the CAS.
- ❖ Undertaking key research projects of the state and international cooperation
- ❖ Aim at developing products and engineering technologies of language processing, such as machine translation, etc.
- ❖ Entitled to educate Master's and Doctor's Degree, more than 20 PhD and master students study in CCLIE every year.

华建集团 HuaJian Group

1.4 Huajian Group Co. of CAS

- ❖ In 1997, The HuaJian Group Co. invested by CAS(55% shares) and the core team of IMT research center(45% shares) was setup, with the registered capital of RMB 1 Million (in 2002 added to RMB 50 Million by profits).
- ❖ Total number of staff: about 200 persons
- ❖ Average age: 28
- ❖ Ph.D degree: 8%
- ❖ Master degree: 17.6%
- ❖ Others: 74.4%

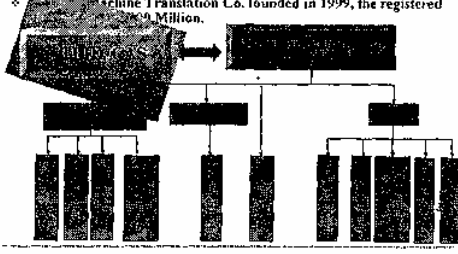


© Copyright © 2002 by HuaJian Group

华建集团 HuaJian Group

Structure of the Huajian Group

- ❖ The Huajian Group consists of 16 companies, in which 6 companies' registered capital is more than RMB 40 million;
- ❖ Machine Translation Co. founded in 1999, the registered capital is 50 Million.



华建集团 HuaJian Group

II. Multi-lingual MT Research in CCLIE

- 2.1 National research projects
- 2.2 Rule-based IMT Research
- 2.3 Template-based MT Research
- 2.4 Hybrid Strategy MT Research
- 2.5 Related research
- 2.6 Results

华建集团 HuaJian Group

2.1 National research projects

- ❖ National 863 Hi-tech Program: 4
- ❖ National Natural Science Foundation of China (NSFC) : 8
- ❖ National Defense Scientific and Technological Research : 1
- ❖ National Ninth Five-year Plan: 1
- ❖ National Technological Innovation Projects: 2
- ❖ National Outstanding Youth Scientific Fund
- ❖ Special Fund Sponsored by Ministry of Information Industry: 2
- ❖ Beijing Municipal Sci. & Tech. Commission: 2

华建集团 HuaJian Group

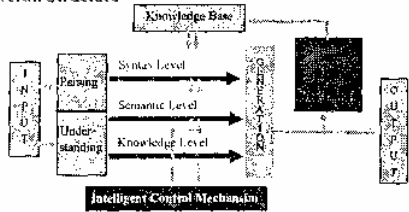
2.2 Rule-based IMT Research

- ❖ Research periods:
 - 1985-1988: Theoretical pursue on rule-based MT approach
 - 1989-1990: Implementation of the prototype system
 - 1990-1992: Implementation of the practical English-Chinese IMT system

华建集团 HuaJian Group

2.2 Rule-based IMT Research

- Overall Structure



华建集团 HuaJian Group

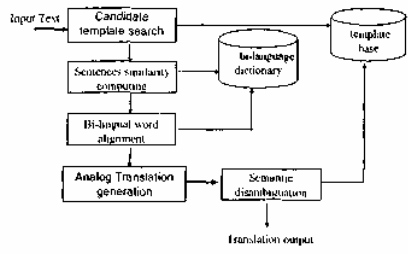
2.2 Rule-based IMT Research

❖ **Features:**

- ❖ Multi-knowledge unified SC (Sub-Category) Grammar
- ❖ Integration of Syntactic, Semantic Knowledge and Common Sense in SC Parsing
- ❖ Hierarchical Adjustable Taxonomy-based Consistent Unification Algorithm
- ❖ Context Sensitive Processing based on forward and feedback test
- ❖ Dynamic selection of parsing level according
- ❖ Generation-oriented source sentence analysis algorithm
- ❖ Extensibility to multiple lingual
- ❖ Objected-oriented Knowledge base Management
- ❖ Effective Post-Editing and Modification Mechanism

华建集团 HuaJian Group

2.3 Template-based MT



```

    graph TD
      Input[Input Text] --> Search[Candidate template search]
      Search --> Similarity[Sentences similarity computation]
      Similarity --> Alignment[Bi-lingual word alignment]
      Alignment --> ATG[Analog Translation generation]
      ATG --> SD[Semantic disambiguation]
      SD --> Output[Translation output]
      Search --- Dict[(bi-language dictionary)]
      Dict --- Base[(template base)]
      Base --- SD
  
```

华建集团 HuaJian Group

❖ **Structured information translation**

- ❖ stated scenario includes:
Weather forecast, food, tourist, Public transport, etc

❖ **High quality translation:**

- ❖ understandability is up to 100%

华建集团 HuaJian Group

2.4 Hybrid Strategy MT Research

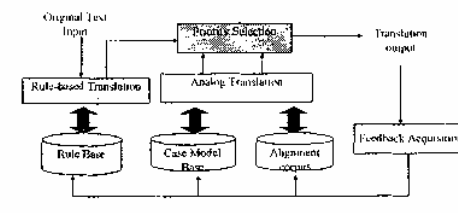
❖ **Research process:**

- 1998-1999:
Theoretical pursue on hybrid strategy machine translation approach
- 2000-2001:
Implementation of the prototype system
- 2001-now:
Implementation of the practical system

华建集团 HuaJian Group

2.4 Hybrid Strategy MT Research

• Overall Structure



```

    graph TD
      Input[Original Text Input] --> Selection[Point Selection]
      Selection --> Output[Translation output]
      Selection --> Rule[Rule-based Translation]
      Selection --> Analog[Analog Translation]
      Rule --- RuleBase[(Rule Base)]
      Analog --- CaseBase[(Case Model Base)]
      CaseBase --- Alignment[Alignment output]
      Alignment --- Feedback[Feedback Acquisition]
      Feedback --> Selection
  
```

华建集团 HuaJian Group

2.4 Hybrid Strategy MT Research

❖ **Main Features**

- ❖ Combination of hybrid translation strategies, including:
Rule-based, corpus-based, case-based heuristic analogy matching, etc.
- ❖ Multi-level abstract knowledge representation of case model
- ❖ Multi-level feature based similarity computation
- ❖ Translation generation based on hybrid strategies
-

华信集团 HuaXin Group

2.5 Related research

- ❖ Large-scale Concept Knowledge System
- ❖ Language Lexical Analysis/pre-processing
- ❖ Hybrid-strategy natural language analysis
- ❖ Information extraction/filtering from internet
- ❖ Text classification based on semantic
- ❖ Word-alignment algorithm
- ❖ Dialog processing for on-line Q&A
- ❖ Topic discovery
- ❖ Name entity recognition

.....

华信集团 HuaXin Group

2.6 Results

- ❖ Academic Papers: more than 80
- ❖ Software Copyrights: about 50 items
- ❖ Patents: 8 items(approved)
- ❖ Awards: more than 8 items
 - ❖ First Prize of National Advance in Science & Technology in 1995
 - ❖ United Nation Science & Technology Invention and Innovation Star in 1996
 - ❖ The National Excellent Projects in 2000

.....

华信集团 HuaXin Group

Some Papers

- ❖ Chen Zhaoxiang et. al., The English-Chinese Machine Translation System IMTEC, The 12th Inter. Conference on Computational Linguistics (COLING'88), Budapest, 1988.
- ❖ Chen Zhaoxiang et. al., The Principle of Intelligent Machine Translation, International Conference on CPCOL'88, Toronto, 1988.
- ❖ Chen Zhaoxiang, et. al., IMTKB: A Knowledge Base System for Machine Translation, International Conference on CPCOL'88, Toronto, 1988.
- ❖ Chen Zhaoxiang & Huang Heyan, Lexical Knowledge Organization in Machine Translation, Proc. of the Post-COLING'94, Inter. Workshop on Directions of Lexical Research, Aug. 1994, Beijing, China, P106-111.
- ❖ Chen Zhaoxiang & Huang Heyan, Future Direction for Machine Translation System Development, PRICAI'94, Inter. Workshop on Knowledge Engineering and Applications, Aug. 1994, Beijing, China, P67-70.
- ❖ Huang Heyan & Chen Zhaoxiang, Design and Implementation of the Chinese-English Machine Translation System (MT/CE), The First China-Korea Joint Symposium on MT (CKJSMT'94), 1994, P19-24.
- ❖ Huang Heyan & Chen Zhaoxiang, A Generation Algorithm in Machine Translation, International Conference on Chinese Computing'94 (ICCC'94), Singapore, 1994.

华信集团 HuaXin Group

- ❖ Chen Zhaoxiang et. al., Integration of Syntactic, Semantic Knowledge and Common Sense in SC Parsing, Inter. Symp. on Natural Language Generation and the Processing of the Chinese Language, 1996.
- ❖ Huang Heyan & Chen Zhaoxiang, A Hierarchical Adjustable Taxonomy-based Consistent Unification Algorithm in Rule-based Intelligent System, The 2nd China-Korea Joint Symposium on Oriental Language Processing, Pohang, 1996.
- ❖ Huang Heyan and Chen Zhaoxiang, Forward and Feedback Context Sensitive Processing, ICCPOL'97, 1997.
- ❖ Chen Zhaoxiang, et. al., Multi-Input Information Recognition and Processing System MHRPS, Inter. Conf. On Machine Translation & Computer Language Information Processing, 1999, Beijing.
- ❖ Huang Heyan, et. al., The design and Implementation Principle of an Interactive Hybrid Strategies Machine Translation System IHSMTS, Inter. Conf. on MT & Computer Language Information Processing, 1999, Beijing.
- ❖ Huang Heyan, Chen Zhaoxiang, Wang Bin & Liu Haijun, Organization, Indexing and Retrieval of Multiple Specialty Lexicons in Machine Translation System, TSIT'97.
- ❖ ... Huang Heyan & Chen Zhaoxiang, Heuristic Centered Strategies for SC ... Parsing, The 3rd Chinese-Korea Joint Symposium on Oriental Language Processing and Character Recognition (CKOLPCR'99).

华信集团 HuaXin Group

- ❖ Huang Heyan & Chen Zhaoxiang, The Hybrid Strategies Generation Approach of IHSMTS, 1st Inter. Symp. On Chinese Information Processing, Shanghai, 2001.
- ❖ Chen Zhaoxiang et. al., A Multi-Level Feature-Based Approach to Calculate Similarity of Two Sentences in IHSMTS System, 1st Inter. Symp. On Chinese Information Processing, Shanghai, 2001.
- ❖ Chen Zhaoxiang et. al., The Case-Model Feature Design and Heuristic Restricted Retrieval Approach of IHSMTS, Symp. on IINC, 2001.
- ❖ Huang Heyan & Chen Zhaoxiang, The Overall Design and Researches of Key Approaches for Integrated Chinese Braille Information Processing Platform, CINLP2002.
- ❖ Huang Heyan & Chen Zhaoxiang, Overall Design of the Intelligent Computer-Aided Translation Platform Based on Hybrid Strategy, <Advances on MT Research 2002>, Beijing, 2002.
- ❖ Huang Heyan & Chen Zhaoxiang, The Hybrid Strategy Processing Approach of Complex Long Sentence, Journal of Chinese Information, Vol. 16, No.3, 2002.
- ❖ Huang Heyan & Chen Zhaoxiang, Design and Implementation of the Case Model Acquisition Mechanism in IHSMTS, Journal of Computer Research and development in China, Vol. 39, No.5, 2002.

华信集团 HuaXin Group

III. Multi-lingual MT Systems Development

- 3.1 The development roadmap of MMT
- 3.2 Ready-made MMT systems
- 3.3 Undertaken MMT systems
- 3.4 Scale of Specialty Dictionary Bases
- 3.5 Industry-oriented MMT systems

.....

华建集团 Huajian Group

3.1 The development roadmap

- ❖ 1990-1992 English-Chinese IMT system, >200 PYs
- ❖ 1992-1997 Chinese-English, Russian-Chinese, Germany-Chinese IMT system, >500 PYs
- ❖ 1998-2000 Japanese-Chinese, Chinese-Japanese system
- ❖ 2000-2004 Chinese-French, Chinese-Spanish, Chinese-Russian system, >400PYs
- ❖ 2004-now Chinese-Korea, Korea-Chinese system

华建集团 Huajian Group

3.2 Ready-made translation system

- ❖ 7 kind of languages
 - ❖ Chinese, English, Russian, Japanese, Germany, French, Spanish
- ❖ 9 systems
 - ❖ Chinese↔English, Chinese↔Japanese, Chinese↔Russian, Germany→Chinese, Chinese→French, Chinese↔Spanish

华建集团 Huajian Group

3.2 Ready-made translation system

System	Description	Data	Specialty Bases
English → Chinese	Having the longest development history, stable and mature, having relatively complete specialty bases, mature oral system developed.	Vocabulary: 250K Rules: 5.8K	32
Chinese → English	Having been tested and tuned with Internet corpus, good translation quality and high readability rate of basic sentences and dialogues.	Vocabulary: 250K Rules: 4.2K	23
Russian → Chinese	After many years' development, the system is relatively stable, because it has been trained with a great deal of Internet corpus, the translation result is good.	Vocabulary: 200K Rules: 3.5K	aviation, electron, mechanical
Chinese → Russian	Being developed from the early of 2003, not mature and need tuned and improved with a great deal of Internet corpus.	Vocabulary: 150K Rules: 3.8K	

华建集团 Huajian Group

System	Description	Data	Specialty Bases
Japanese → Chinese	Being developed from 2000. The oral system has better readability for basic sentences and dialogues.	Vocabulary: 133K Rules: 1.3K	
Chinese → Japanese	Being developed from 1999, relatively mature system; oral system developed.	Vocabulary: 190K Rules: 2K	Sports, tourism
Germany → Chinese	Being developed from 1993	Vocabulary: 112K Rules: 2.4K	major
Chinese → French	Being developed from 2000	Vocabulary: 148K Rules: 3.5K	Sports, tourism
Chinese → Spanish	Being developed from 2000	Vocabulary: 160K Rules: 3.7K	

华建集团 Huajian Group

3.3 Undertaken MMT systems

- ❖ 1 kind of languages
 - ❖ Korea
- ❖ 2 systems
 - ❖ Chinese↔Korea

华建集团 Huajian Group

3.4 Scale of Specialty Dictionary Bases

Specialty Bases	E→C(K)	C→E(K)	R→C(K)	C→C(K)
Motor	82	97		70
Machinery	166	193		
Law	63	54		
Audi	79	68		
Textile	73	65		
Aviation	138	127	43	
Medicine	74	217		
Water conservation	65	45		
Trade	30	34	35	
Petroleum	141	49	Finished	
Communication	50	50		
Tourism	6	17		
Computer	40	47		
Chem. & chem. eng.	112	71		

华建集团 HuaJian Group

Specialty Bases	E→C(K)	C→E(K)	R→C(K)	G→C(K)
Architecture	39	23		
Agriculture	219	1.3		
Military	91(S) 287(L)	77	Finished	
Communication	90	80		
Telecom	101			
Finance	77	78		
News	16			
Compilation	29			
Social Sciences	14			
Fine arts	13			
MBA	3	3.5		
Broadcast	11			
Advertisement	6			
Education	16			

华建集团 HuaJian Group

Specialty Bases	E→C(K)	C→E(K)	R→C(K)	G→C(K)
Food and beverage	16			
Shipping	66		183	
Geology	162			
Energy	17			
Electromechanical	158		110	
Programmer control	35			
Sensing	4			
Optical fiber	8			
Metallurgy	87			
Enterprise management	37	23		
Environment	2.6	2.4		
Electric	2.7			
Sports	24	22		
Food	48			
Engineering	173			

- 华建集团 HuaJian Group
- ### 3.5 Industry-oriented MT Development
- ❖ Patent MT System (State Intellectual Property Office)
 - ❖ Patent Ontology for translation
 - ❖ Large-scale Specialty Dictionaries
 - ❖ Process of Special Name and Formula
 - ❖ Rule of Patent Claims
 - ❖ Aided MT System for Standardization (China Institute of Standardization)
 - ❖ Standard Terms translation
 - ❖ Hybrid Strategy MT
 - ❖ Large-scale bilingual corpus

- 华建集团 HuaJian Group
- ### IV. Series Products and Applications
- 4.1 Large-scale multilingual network information translation platform
 - 4.2 Huajian intelligence aided translation platform
 - 4.3 Embedded MT system
 - 4.4 Multilingual SMS translation platform
 - 4.5 PC-based multilingual MT software
 - 4.6 Main Customers

- 华建集团 HuaJian Group
- ### 4.1 Large-Scale Multilingual Network Information Translation Platform
- ❖ Functions :
 - ❖ Web browsing translation service
 - ❖ On-line translation service of given text
 - ❖ Translation service of uploaded files in specific formats, including *.txt, *.rtf, *.Doc, etc.
 - ❖ Email translation service
 - ❖ Bilingual chatting room, BBS, etc.
 - ❖ International negotiation room
 - ❖ Bilingual information retrieval system
 -

华建集团 HuaJian Group

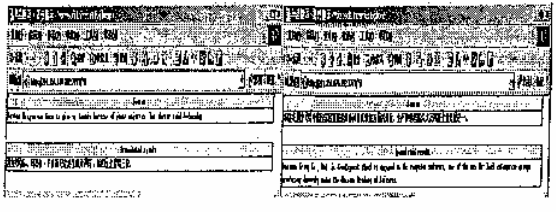
❖ Real time browsing translation:
Provides real time web page browsing translation service

Eng ⇄ Chi Chi ⇄ Eng

华建集团 Huajian Group

❖ **Instant Translation System**
Provides online instant translation service to given text

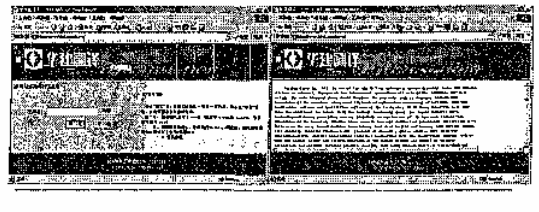
Eng ⇄ Chi Chi ⇄ Eng



华建集团 Huajian Group

❖ **File upload translation System**
❖ Provides instant upload translation service to specific format files, including txt, rtf files.

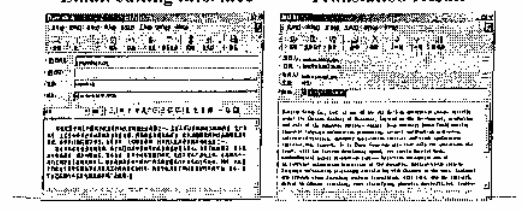
Upload Interface Translation result



华建集团 Huajian Group

❖ **Email automatic translation system**
❖ Provides automatic translation service to Chinese/English emails

Email editing interface Translation Result

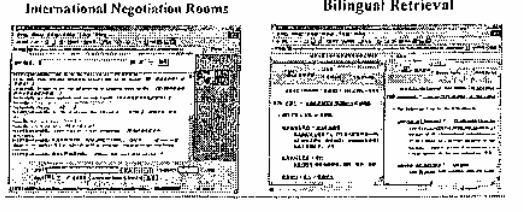


华建集团 Huajian Group

❖ **Other functions**

- ❖ Bilingual Chatting Room ❖ Bilingual BBS
- ❖ International Negotiation Room ❖ Bilingual Retrieval system

International Negotiation Rooms Bilingual Retrieval



华建集团 Huajian Group


Applications

- ❖ **Main Users:**
 - ❖ **Telecom Enterprises:**
 - ❖ China Telecom, China Unicom, China Jitong (Netcom)
 - ❖ **Government Websites:**
 - ❖ State Economic and Trade Committee, Departments of State Development Planning
 - ❖ **Enterprise Information Centers:**
 - ❖ Luneng Elec. Group, Rainbow securities, Australia Tran-sea
 - ❖ **Large-scale International Activities:**
 - ❖ Beijing 2008 Olympic Multilingual Information Service System
 - ❖ 2010 Shanghai Expo Multilingual Information Service System
- ❖ **Contract total: more than RMB 200 Million**

华建集团 Huajian Group

Applications

- ❖ **Related Web Sites:**
 - <http://www.hjtrans.com>
 - www.165net.com
 - www.luneng.com
 - www.wtscyber.com
 - www.netat.net
 - www.cei.gov.cn
 - www.cetin.net.cn
 - www.ctiin.com.cn



华建集团 HuaJian Group

4.2 HuaJian Intelligence Aided Translation Platform

- ❖ Computer Aided Translation platform with Multi-layer architecture, including translation platform, parallel bilingual alignment platform and translation projects and users management platform.
- ❖ Core MT approach
 - ❖ Rule-based IMT approach
 - ❖ Hybrid approach combining with Heuristic analogical matching
- ❖ Functions
 - ❖ Provides an intelligent and high performance aided translation tool platform for professional translator
 - ❖ Provides a modern office tool for multilingual information workers
- ❖ Editions
 - ❖ Having released IAT1.0, IAT2.0, IAT Multilingual Edition
 - ❖ Chi→Eng Rus ↔Chi Jap ↔ Chi Chi→Fre

华建集团 HuaJian Group

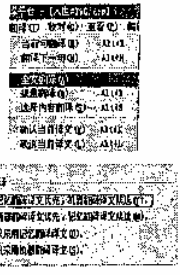
IAT Efficiency and advantages

- ❖ Efficiency
 - ❖ Doc and Txt files translation can save 50% translation time in average
 - ❖ HTML pages translation can save 70% translation time in average
- ❖ Advantages
 - ❖ High automatic translation accuracy rate
 - ❖ Powerful Translation Memory ability
 - ❖ Intelligent man-machine interaction
 - ❖ Synchronous tracking and interacting
 - ❖ Mass vocabulary
 - ❖ English→Chinese: 2.3 million items
 - ❖ Chinese→English: 410 thousand items
 - ❖ Translation transaction processed by flow
 - ❖ Multi formats compatible
 - ❖ supports the translation of Word, RTF, TXT, Html format files

华建集团 HuaJian Group

Translation mode and multi-output recommendation

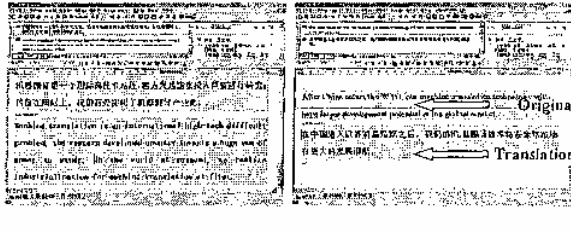
- ❖ Translation mode
 - ❖ Sentence translation
 - ❖ Selected text translation
 - ❖ Batch translation
 - ❖ Full text translation
- ❖ Multi-output recommendation
 - ❖ RBMT output
 - ❖ EBMT output



华建集团 HuaJian Group

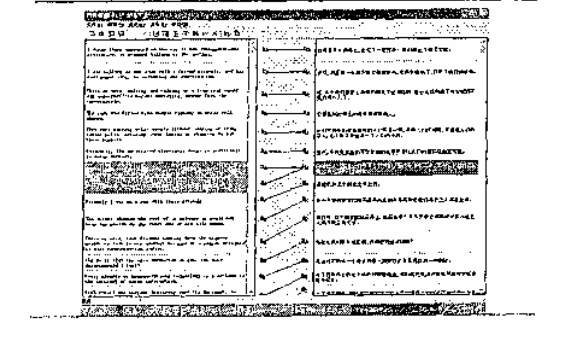
Automatic Translation

Chi ↔ Eng Eng ↔ Chi



华建集团 HuaJian Group

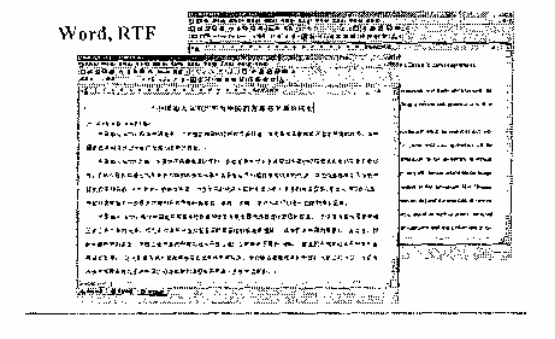
Corpus alignment platform



华建集团 HuaJian Group

Multi-document format processing


Word, RTF



华建集团 HuaJian Group

Applications

- ❖ Beijing Organizing Committee for the Games of the 29th Olympiad
- ❖ CASS-Chinese Academy of Social Sciences
- ❖ First Institute of the Ministry of Aviation Beijing Security Bureau
- ❖ Navy Translation Team
- ❖ Institute of Air Force
- ❖ Navy
- ❖ Deloitte
- ❖ HCC
- ❖ ...



华建集团 HuaJian Group

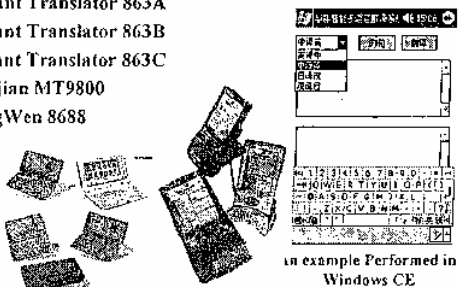
4.3 Embedded MT System

- ❖ Inter-platform virtual machine technology applicable to different hardware platform and OS;
- ❖ The knowledge encoding algorithm to be suitable for limited memory;
- ❖ Optimized kernel machine translation algorithm for low CPU speed ;

华建集团 HuaJian Group

Products Family

- Instant Translator 863A
- Instant Translator 863B
- Instant Translator 863C
- Huajian MT9800
- LangWen 8688



an example Performed in Windows CE

华建集团 HuaJian Group

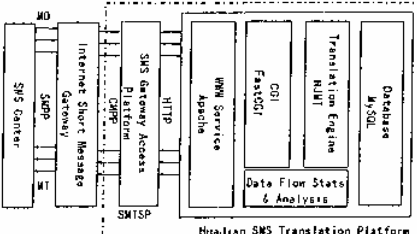
Customers

- ❖ GSL Group of HK
- ❖ HTC International of TW ("HTC", "LEO", "WinTech")
- ❖ BESTA Technology (HK)
- ❖ Sales volume
 - ❖ 530 K
- ❖ Revenue
 - ❖ HK\$1.27 Billion
- ❖ Foreign Exchange Income (up to now)
 - ❖ US\$ 8.6 million + HK \$ 10 million

华建集团 HuaJian Group

4.4 Translation Service on SMS and IM

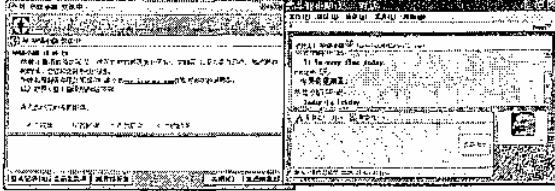
- ❖ Translation Service on SMS (Short Message Service)
 - ❖ Telecom valued-added service oriented SMS MT service platform



HuaJian SMS Translation Platform

华建集团 HuaJian Group


- ❖ Translation Service on IM (Instant Messenger)
 - ❖ Used on Tencent QQ system, the most widely used IM service in China now (350 million users), ranks 3 in all its SPs.
 - ❖ Being tested on MSN internally



华建集团 HuaJian Group

4.5 PC-based Multilingual MT Software

- ❖ PC-based MT Software tools:
 - ❖ Huajian Net Wizard
 - ❖ Huajian Bilingual Browser
 - ❖ Huajian Easytrans
 - ❖ Huajian Dictionary
 - ❖ Speed Translation-Family EICQ (U.S. Version)



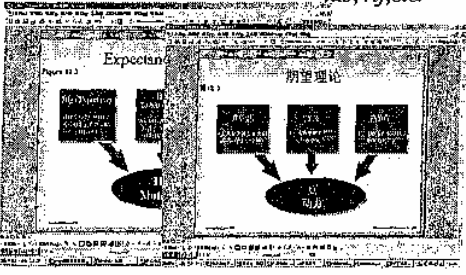
华建集团 HuaJian Group

Main features

- ❖ Be usable on various OS, including: Win98/2000, WinNT/XP, Linux, etc;
- ❖ Seamless link to various document processing software, such as: Microsoft Words, IE, MS Excel, MS PowerPoint, etc;
- ❖ Be able to translate the text contents of various formats, such as HTML, DOC, PPT, PDF, HLP, email, ICQ, etc.
- ❖ Provide several translation modes, including Full-text translation, selected text translation and mouse pointed word translation;

华建集团 HuaJian Group

Multiple format processing *doc, txt, htm, ppt, xls, rtf, etc.*



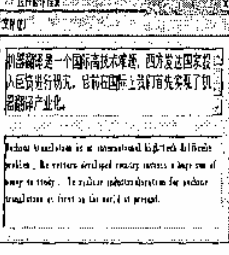
华建集团 HuaJian Group

Full-text Translation

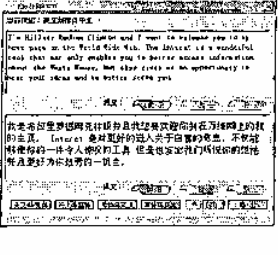


华建集团 HuaJian Group

Selection Translation



Instant Translation and Reading



华建集团 HuaJian Group

Applications

- ❖ Huajian Net Wizard:
 - ❖ 450 thousand sets sold in overseas market (Hong kong PCCW as the general agent)
- ❖ Huajian Bilingual Browser
 - ❖ Bounded with Legend (Lenovo), Hisense, Tongfeng, TCL, annual sales is more than 3 million sets (Hong kong PCCW as the general agent)
- ❖ Huajian Easytrans:
 - ❖ Japan Logo Vista as the agent; Taiwan Inventec-SDK license
- ❖ Huajian Dictionary
- ❖ Speed translation-family EICQ
 - ❖ U.S. Version, USA Loto as the agent

华建集团 HuaJian Group

V. Future Works

- ❖ **Research**
 - ❖ The proper combination of hybrid approaches
 - ❖ To setup an ontology mechanism for different domain application
 - ❖ Bi-lingual alignment (phrase-level, sentence-level)
 - ❖ The decomposition of Long complex source sentence and the composition of target sentence
 - ❖ Extraction of language feature for sentence in example-based MT
 -

华建集团 HuaJian Group

Development

- ❖ **Development**
 - ❖ More languages, such as Arabic, Korea, etc.
 - ❖ More application domains for different languages
 - ❖ Parallel processing ability for large-scale users
Network-oriented service applications
 - ❖ Software implementation technologies for different hardware/software environment, such as term management, format document processing, etc
 -

华建集团 HuaJian Group

VI. Some Considerations

- ❖ Machine Translation (MT) is a notoriously hard work, needs to dedicate to it.
- ❖ Design and develop MT systems on the demand of various fields, applications, and scenarios.
- ❖ Reasonable expectation, then reasonable results
Don't force machine to do what is hard for it.
- ❖ Be careful to process the relationship of 80% to 20%.
- ❖ Pay more attention to the human-machine interface to enhance the usability.
- ❖ Only better, no best – be confident

华建集团 HuaJian Group

Thanks!

Questions?

