

An Advanced Multilingual MT System: Facilitating Patent Examination and Search

Cao Jingcheng

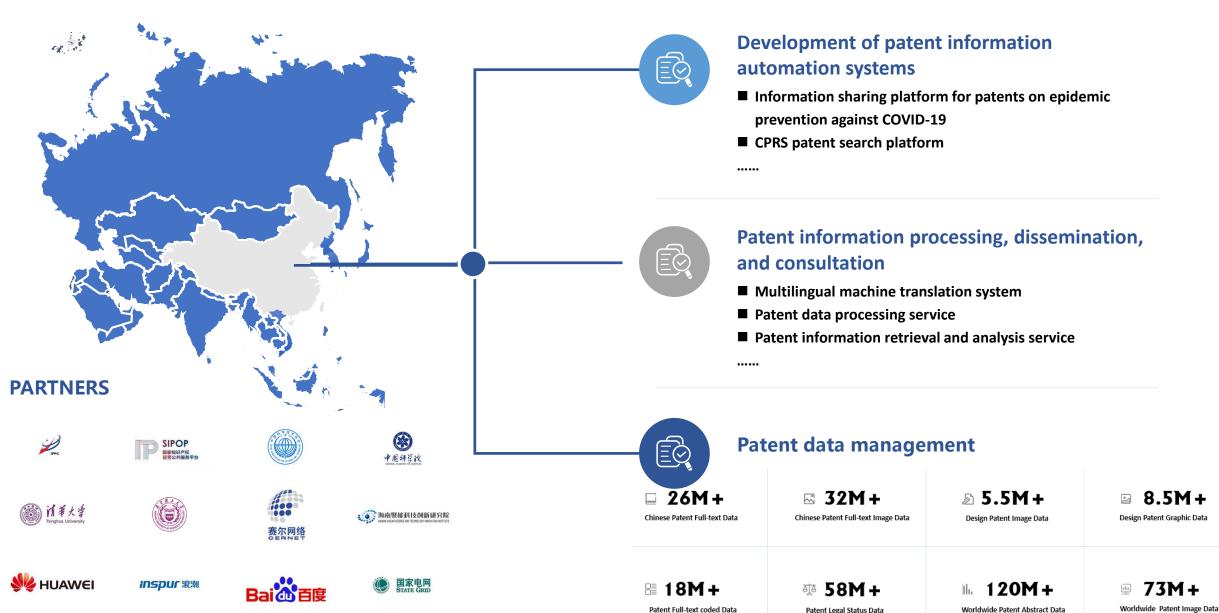
China Patent Information Center, CNIPA

February 2023



CNPAT & MT History

For 34 years, CNPAT has been continuously committed to the development of patent information automation systems and related services including patent information processing, dissemination and consultation.

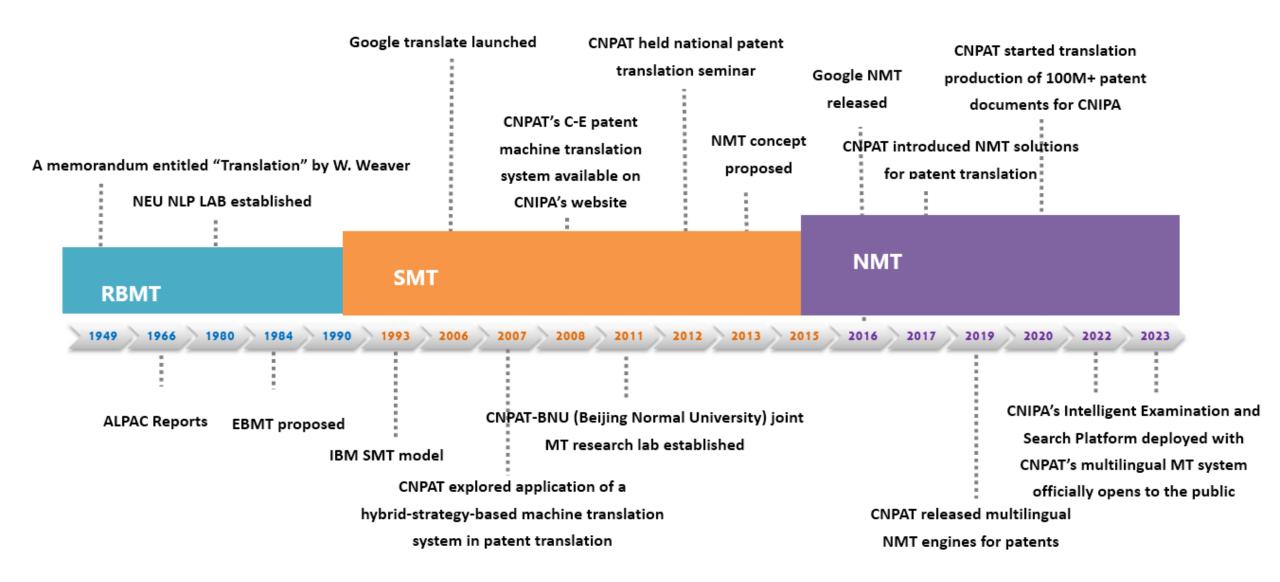


About CNPAT



CNPAT Machine Translation History





Multilingual MT system in application



			oogle Chrome							- 0	
		247/q	uickView?pn	Id=US2022355487A	1						
355487	A1										
亮设置	聚焦设	置			公开日: 202	22.11.10 - 言高完设置	聚焦设置 高密				
	()) 1/23 K) Q Q O O ● 上						多个传感器从航走机器人的环境收集环境数据,以及通过航法通信接口发送要由航送远程操作系统接收的航送环境数据。				
									说明书		
						★ 技术领域	全 (1)(1)(1)(1)(1)(1)(1)(1)(1)(1)(1)(1)(1)(
			US 2022/0355487 A1 Nov. 10, 2022				文本发明的机器人、远程操作系统和方法通常涉及使用所述方法和远程操作系统控制所述机器人的操作,并且特别涉及机器人和远程操 文本 作系统之间的通信、控制范例、反馈和训练。				
		ROBOTS, TELE-OPERATION SYSTEMS, communication interface, control selection input indicating					▶ 背景技术				
	paradigms based on at least one property selected from a TPCHNICAL RELD ensuing consisting of a task to be reformed by the relate a			新す	摘						
			[0001] The present	robots, tele-operation systems, and late to controlling operation of sais hods and tele-operations systems, and	description of a human operator of the tele-operation sys-			贴、VR套件或控制面板, 各,或者可以是远程的,诸		多不同变型是可能的。远程通信可以是短程的, 四49	
			particularly relate to	hods and tele-operations systems, and communication, control paracligns ig between robots and tele-operation	control paradigms including a respective control paradigm	通 明 发明内容	80,70,818,90,919,42,42,42,42,42,42,42,42,42,42,42,42,42,	治,助治19以7定以1911,1泊	(山西)辺像白峡(州山) /2)	M198.	
			sysiens.	ACKGROUND	 specific to each respective candidate human operator; receiving the control selection input may include receiving an identification of a particular human operator from the 		术了机器人和沅程操作系	<u>统,其中可以基于操作者</u> 的	的身份、操作者的角色。	成要由机器人执行的预期任务从多个控制范例中	
perioderly value to comanization, outility perioder including in specific register in an including an approximation of the perioder including an approximation of the perioder including an approximation of the perioder including and the period and the perio											
					-03202233340/AT						
	0	15202	22355487A	1							
	■ 高売设置 聚焦设置						■ 高亮设置	聚焦设置		公开日: 20	
		ENTXTC WPABSC ENTXT WPABS OETXT USTXT ALL					蕃录	< >	9 / 23 < >	ର୍ଷ ଚାତ 🖕 🛓	
		PA - SANCTUARY COGNITIVE SYSTEMS CORP					↓ ↓ </td				
		w		SANCTOARTC	Soon The State Ma Conf		w	plurality of control paradigms r	ay include receiving, by the	at least processing unit, the select control paradigm based	
		全 文 文 本	AC -	CA 翻译						×	
		文本	IN -	ROSE GE							
					英语 マ	▲ 翻译	清空」」	▲文件 未选择任何文件			
		摘要附图	AB -	The pres	The present disclosure describes robots and tele-operation sys		公开描述了机器人和远	程操作系统,其中基于操作	■者的身份、操作者的	角 Nov. 10, 202	
				control pa	where a select control paradigm is selected from control paradigms based on an identity of an ope	and an and a set and		期任务从多个控制范例中) 范例中的任何一个来操作,		用	
		2 12		of an ope	operator, or expected tasks to be performed by the robots can be operated in accordance with any o	ne robot. Generic f the plurality of	王何一个可以通过选择	适当的控制范例来充当操作	著的角色或充当操作	ă I I I I I I I I I I I I I I I I I I I	
		说 明 书		Generic r	control paradigms, such that any of said generic a role or act as an assistant to an operator by	TODOUS CALL SCIVE		反馈机制,机器人或远程挑 控制范例,以随时间逐渐;		pecific to the particular human operator identi trol selection input.	
		书附图		paradign to an ope	appropriate control paradigm. Control paradigms ca specific, or specific to a faction or role which an	an be operator				he tele-operation system may be operable by f candidate human operators; a plurality of subs	
	۱ ۲	PDF		be opera	or specific to a set of tasks to be performed by present disclosure also describes feedback mechan	the robot. The				operators may be defined among the plurality human operators, the plurality of control pa	
				to a set c	robot or tele-operator system receive operator feed	lback and update				include a respective control paradigm specific it of human operators; receiving, by the at le device, control selection input indicating a sel	
		全 文 图 像		feedback	a control paradigm, to gradually improve the contritime.	or paradigm over				adigm from a plurality of control paradigms n reiving, by the at least one input device, cont	
				feedback						nput indicating a description of a human opera -operation system; selecting the select com may comprise selecting the control paradi	
				over time						a subset of human operators corresponding to a of the human operator in the control select	
										ach control paradigm in the plurality of cont	
			ICAI -	<u>B25J9/16</u>						may be specific to a respective set of tasks in if sets of tasks to be performed by the rob	
			IPC8 -	B25J 9/1(by the at least one input device, control selecti ating a select control paradigm from a plurality adigms may include receiving, by the at least c	
				20060101		861				re, control selection input indicating at least of performed by the robot; and selecting the selecting the selecting the selecting the selecting the selection.	
			20060	20060101						udigm may comprise selecting a control paradig a set of tasks which includes the at least one ta n the control selection input.	
				20060101ALIZU	1220/13BHCA			robot via a tele-operation syste	n the tale operation proton	[0021] the method may further comprise receiving, by t	
								coor via a wie-operation syste	ing unit, at least one input	communication interface, an indication of at least one ca didate action to be performed by the at least one robot bas	

MT application in the new generation intelligent patent examination and search system of CNIPA



₩P 快i A 2

Multilingual MT system in application

0



 7 languages

 Chinese, English, Germany, French, Japanese, Korean, Russian

 Multiple data formats

 XML/Word/PPT/Excel/PDF

Patent and non-patent literature

100 million+ from over 100 countries, regions and organizations



Clarifying Needs

Facing Pain Points

2018 2019 Added Value Data of Patent-intensive Industries (RMB 100 million)

11.6

107090

Share of Added Value of Patent-intensive Industries in GDP (%)

(Source: 2022 China's IP Development Status Evaluation Report, released by CNIPA on December, 2022)

114631



IP for innovation



300000

250000

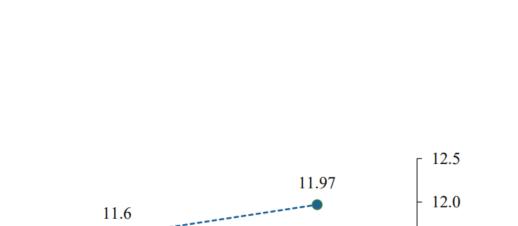
200000

150000

100000

50000

0



121289

2020



- 11.5

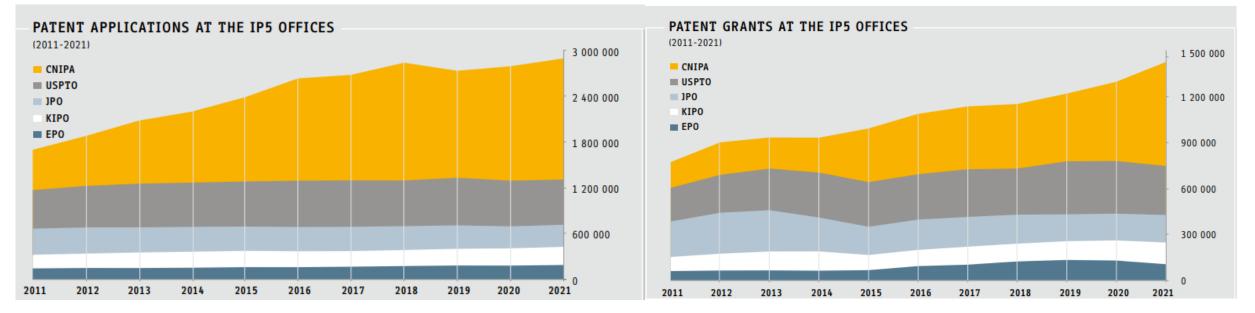
11.0

10.5

10.0

Urgent need for high-quality patent machine translation





(Source: KEY IP5 STATISTICAL INDICATORS 2021)

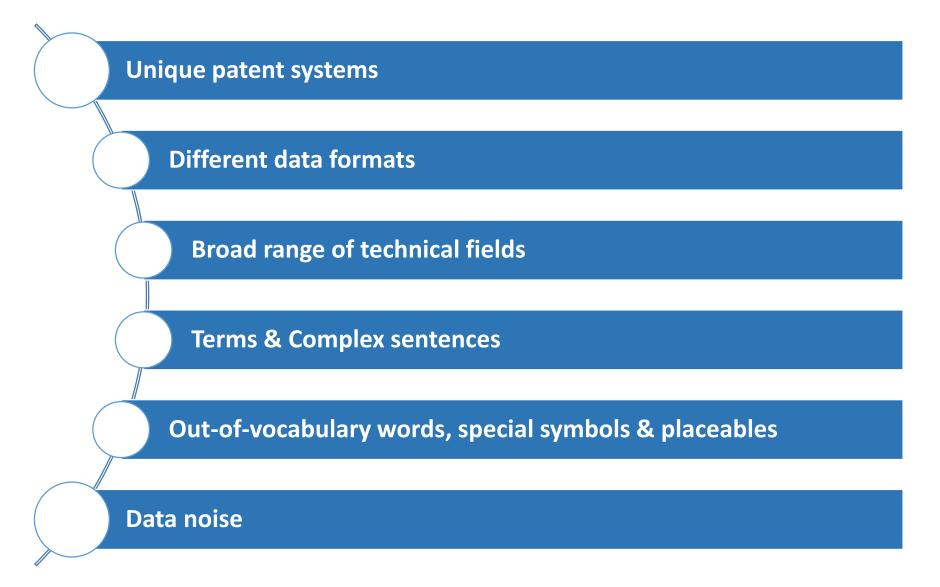
Patents	2020	2021	Growth rate (%)	Share of world total (%)	
Applications worldwide	3,281,900	3,401,100	3.6	100.0	
China	1,497,159	1,585,663	5.9	46.6	
US	597,172	591,473	-1.0	17.4	
Japan	288,472	289,200	0.3	8.5	

(Source: World Intellectual Property Indicators 2022)

9

Challenges in patent machine translation







Customizing MT System

Focusing on Translation Quality

CNPAT's multilingual MT system construction basis



01

03

AI technology

- NMT technology
- Advanced word segmentation technology
- Machine learning error correction model

Corpus resources

• Global patents

02

04

• Advanced corpus processing techniques

30+ years patent translation experience

- Linguistic analysis team
- Experienced patent translation and quality assessment team

System customization service

- Quick response
- Good extensibility

Our work in multilingual MT system construction





About the multilingual MT system



Benefits and features

- Translation quality competitiveness for all language pairs
- Language scalability
- Real-time translation and batch translation
- Automatic language recognition
- Translation memory & glossaries
- Supporting application in multiple scenarios
- Supporting local and cloud deployment
- Confidentiality
- Work standalone and can also be fully integrated



Embracing Innovation

Always on The Road

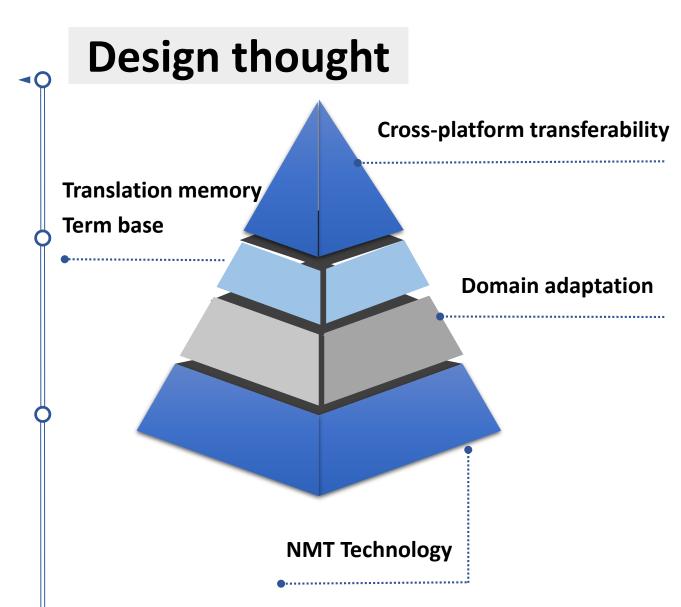
Multilingual MT system



Goals

- To provide fast and accurate translation of multilingual patent documents on the premise of ensuring data security
- To expand the scope of patent retrieval and improving patent browsing efficiency for patent professionals and the public
- To offer translators high-quality machine translation output for quick post-editing











Advance in in-depth study on machine translation needs and machine translation technology

> Consolidate and enhance industrial and technical advantages of machine translation in the field of intellectual property

Þ

Boost translation quality improvement, functional enrichment, and technological upgrade of the multilingual MT system

Deliver high-quality patent information services in the new development stage

THANK YOU!

Contact info: caojingcheng@cnpat.com.cn

CNPAT 中国专利信息中心

守正 创新 专业 乐业