FUJITSU

Developing International Standards (Esp. of CJK Ideographs)

Masahiro Sekiguchi（関口正裕）
Fujitsu Limited / Chair, Japanese SC 2

shaping tomorrow with you

Contents

FUJITSU

- Introduction
  - Codes, Fonts, …
  - UCS/Unicode
- International Standards
  - What is ISO?
  - Standards bodies

Some Items may not be covered due to time constraints

- Coded character set Standard for CJK Ideographs
  - Ideographs (Kanji)
  - Why *International* Standard?
- Standardization Process
- Recent Topics
  - Ideographic Variants
  - *Emoji* (ケータイ絵文字)

The IRG meeting

FUJITSU

- IRG (Ideographic Rapporteur Group) had a meeting this week, here at **Nagaoka University of Technology**.
  I appreciate good hospitality.
- IRG met here to discuss standardization issues of CJK ideographs (and related issues.)

… I'm afraid most of you don't understand what I said, though.

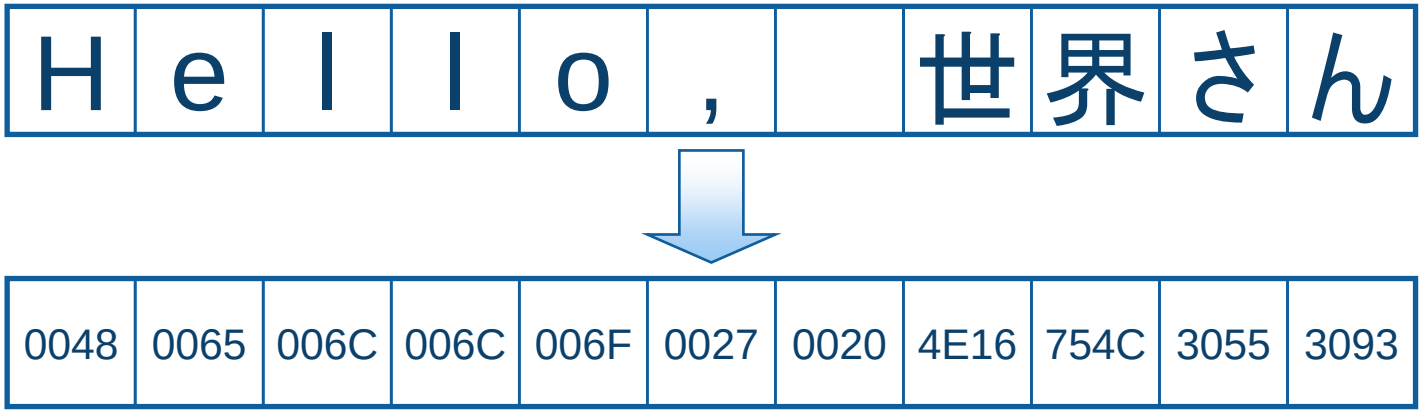
Introduction

- What is
  “Coded character set”?
- What is Fonts?
- UCS/Unicode

Coded character set

FUJITSU

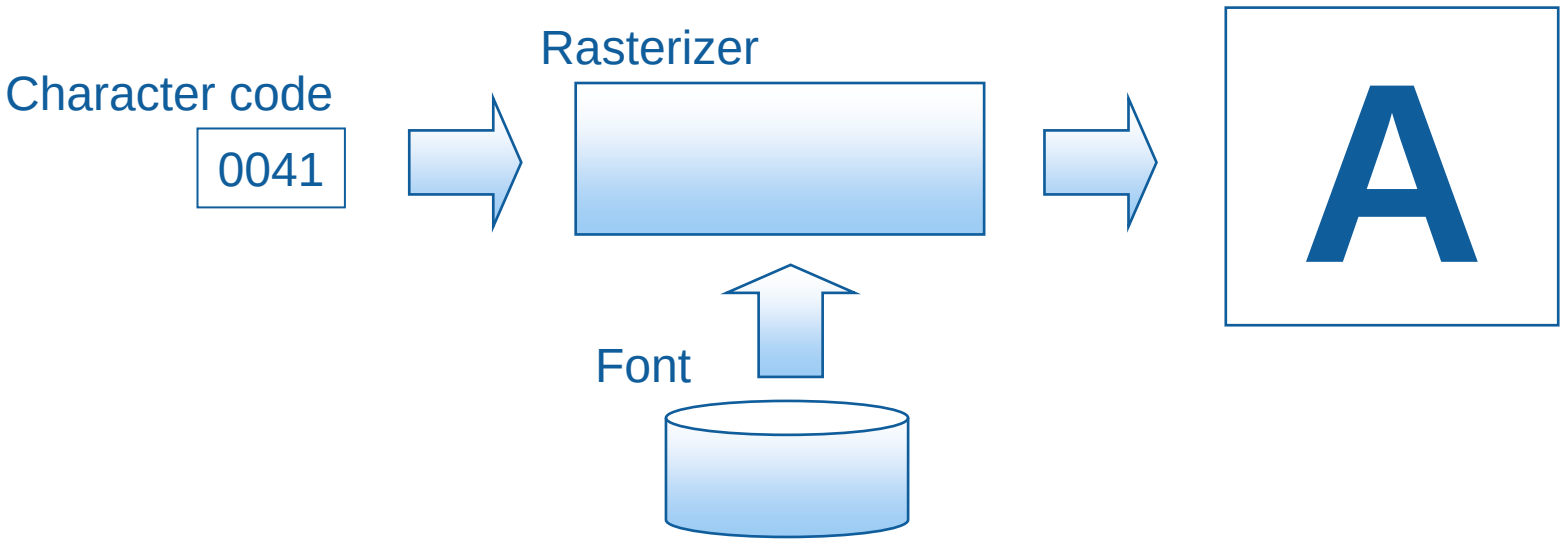
- Computers process numbers. Written texts need to be turned into numbers before we can process them. *Code character set* is a rule of such assignment.
- Also called charset, character code, encoding, … In Japanese we usually call it 文字コード.
- The term coded characgter set is used in the formal standards such as ISO. In Japanese (JIS), the term 符号化文字集合 is used, but it is rarely used in the real world.
- ASCII, EBCDIC, Shift JIS, ISO-2022-JP, ISO-8859-1, Unicode, … are examples.



What is Fonts?

FUJITSU

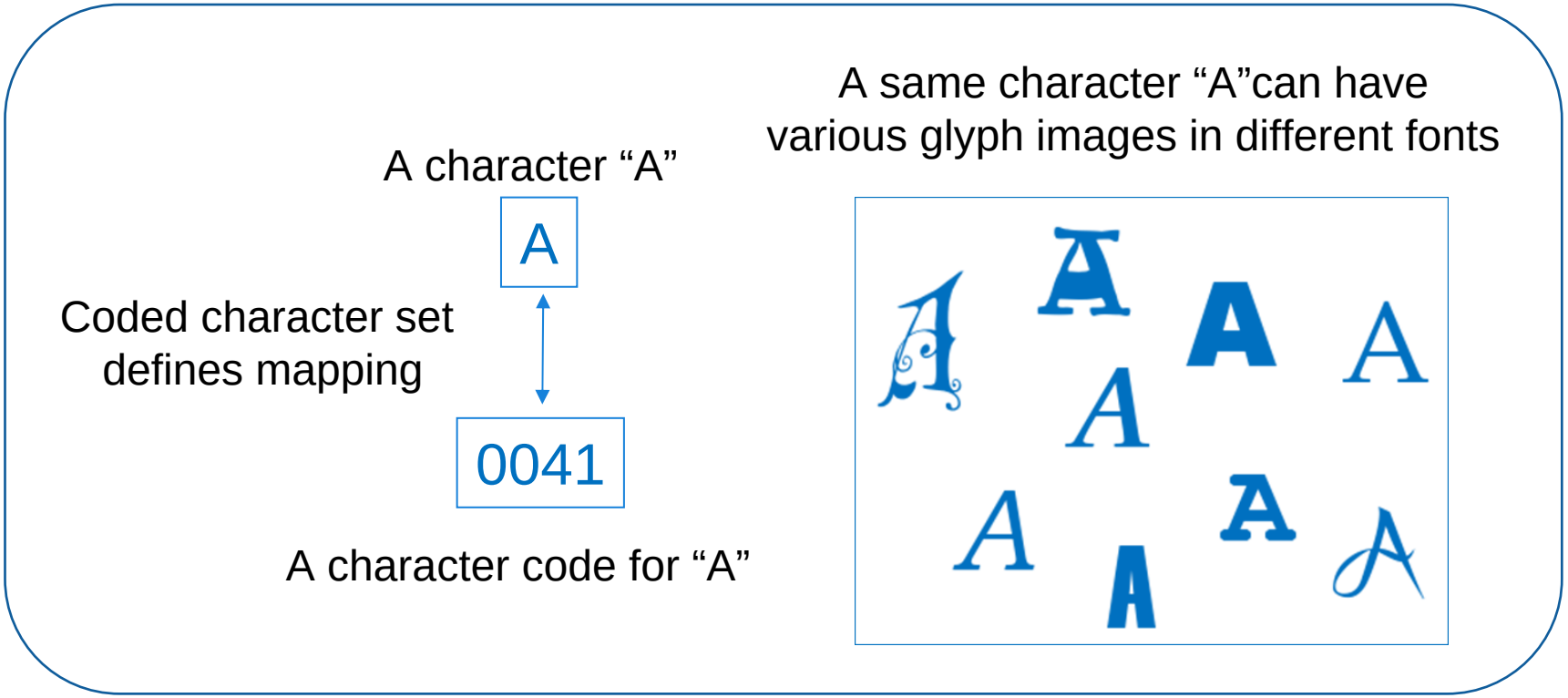
- A set of data to map numbers (character codes) to graphic shapes (and glyph images finally.)
- Bundled in computer/devices,
  as well as sold as separate products.
- A component called *rasterizer* uses the font information to produce the glyph images of characters.



Codes and fonts

FUJITSU

- A font is a graphical art designed by typographers
  - A lot of different fonts for different purposes
- A coded character set is standardized. Font itself is not a standard.
  - There are standards related to fonts, e.g., font file format, though.



UCS/Unicode

FUJITSU

- A modern coded character set designed in early 1990s. (c.f. ASCII was designed in early 1950s.)
- Published as an International Standard **ISO/IEC 10646**.
  - Also published as identical national standard in many countries.
  - In Japan, it is available as JIS X 0221 in Japanese language.
- Aiming at a single universal coded character set that covers worldwide needs.
- UCS is a short for Universal Coded Character Set. JIS calls it 国際符号化文字集合 in Japanese.
- Unicode is published by the Unicode Consortium and is exactly the same coded character set. You can think UCS and Unicode a synonym.

... That's what we met for!

FUJITSU

- ISO/IEC 10646 is an International Standard
- It is a **huge** standard;
  - consisting of > 2,000 pages + some data files,
  - defining ≈ 110,000 characters, and
  - ≈ 70% are so-called CJK ideographs (漢字・汉字)
- We IRG are developing the CJK ideographs part of ISO/IEC 10646.
  - And some more things related to it.

Summary (1)

FUJITSU

- We need coded character set and suitable font to handle text on computers.
- International Standard ISO/IEC 10646 defines the modern coded character set UCS/Unicode, that contains almost all characters in the world including CJK ideographs.
- IRG is developing the CJK ideographs part of the standard.

International Standards

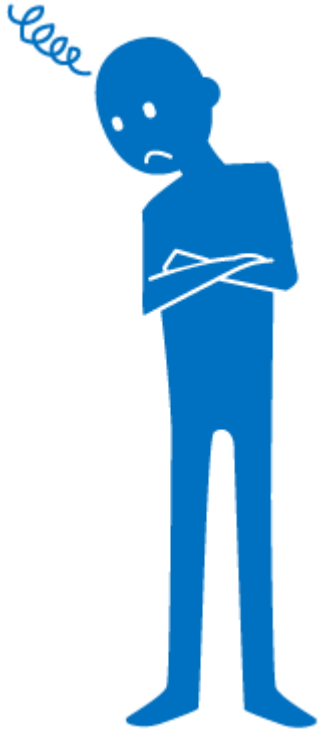
- What is ISO?
- Major SDOs
- Member bodies

I’ve heard of ISO, but...

FUJITSU

... it should be something related to production quality management or environment issues.

Why does ISO matter to “CJK Information Processing” ?



ISO is not just about it

FUJITSU

- There are several famous ISOs
  - ISO 9000: Quality Mgmt Sys.
  - ISO 14000: Environmental Mgmt Sys.
  - ISO 27000: Information Security Mgmt Sys.

- Some people say “ISO” to mean one of the above particular standards, but ISO really means more.
- There are nearly 20,000 ISOs published.

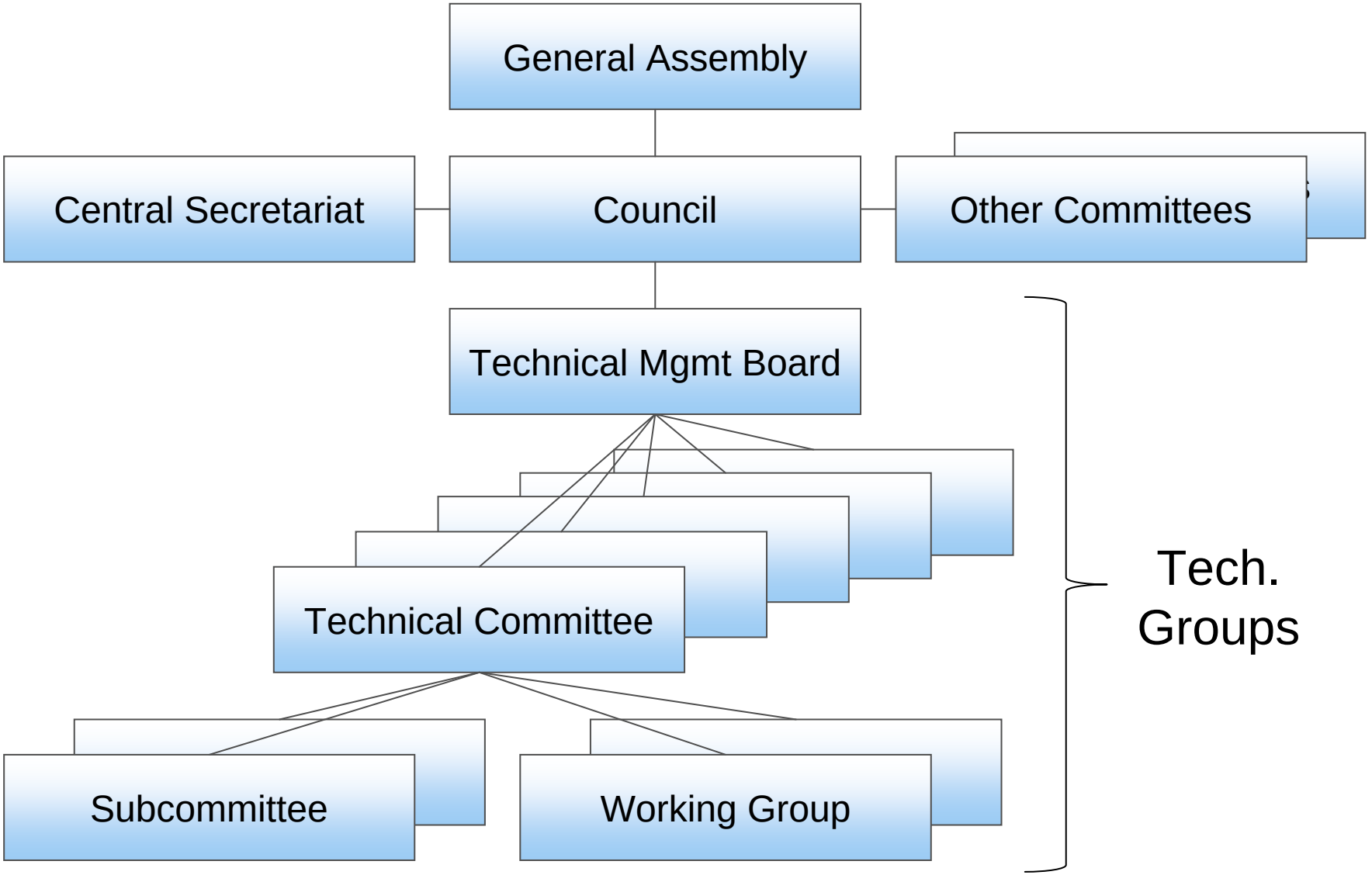
Three Major SDOs

FUJITSU

- ITU (International Telecommunication Union)
  - Communication technologies (incl. radio and broadcasting)
- IEC (International Electrotechnical Commission)
  - Electrotechnologies (Electric/Electronic)
- ISO (International Organization for Standardization)
  - Anything else

ISO is one of the above three, covering **anything else**.

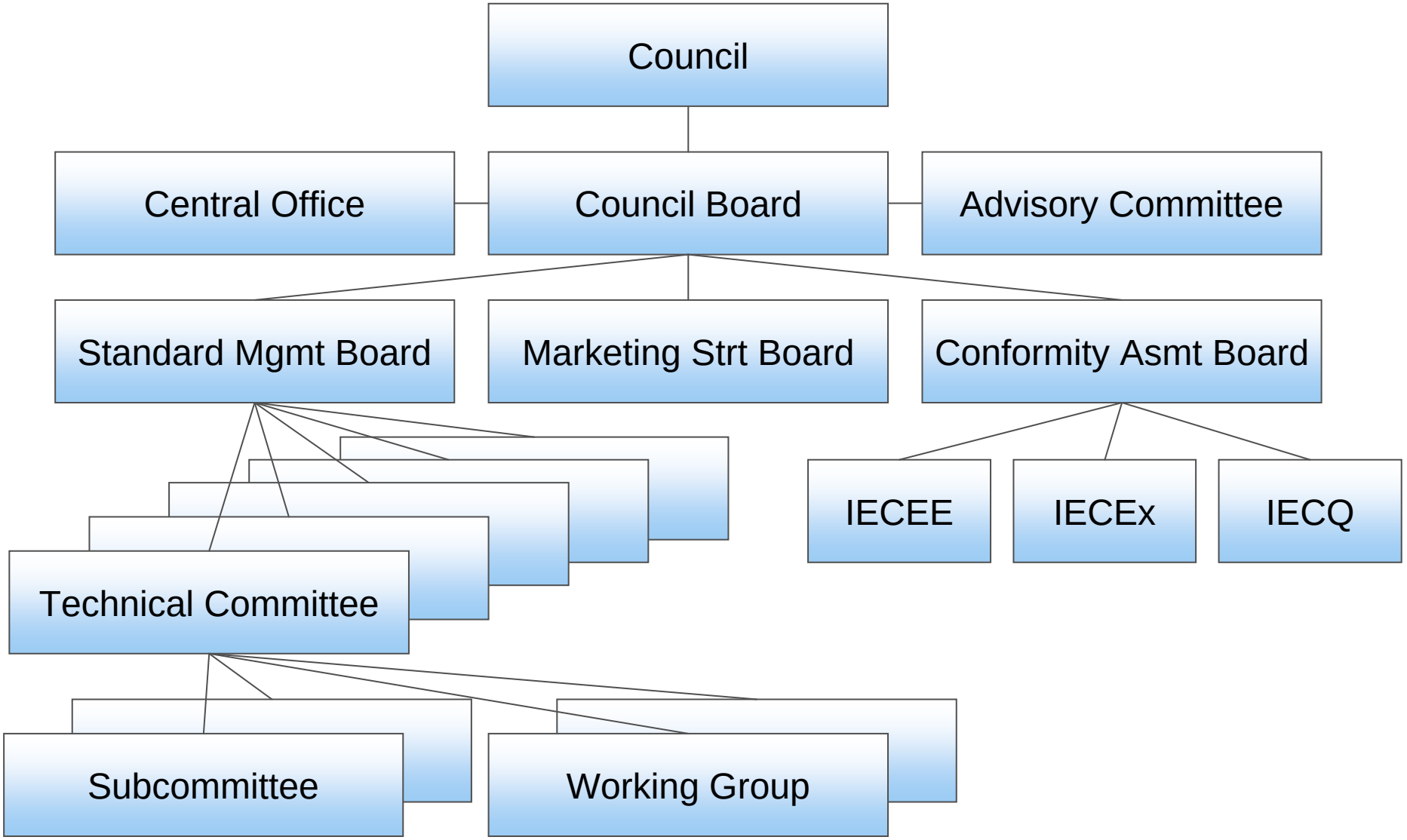
ISO Structure



ISO TCs

TC 1	Screw threads	TC 71	Concrete, reinforced concrete and pre-stressed concrete	TC 138	Plastics pipes, fittings and valves for the transport of fluids	TC 204	Intelligent transport systems
TC 2	Fasteners	TC 72	Textile machinery and accessories	TC 142	Cleaning equipment for air and other gases	TC 205	Building environment design
TC 4	Rolling bearings	TC 74	Cement and lime	TC 145	Graphical symbols	TC 206	Fine ceramics
TC 5	Ferrous metal pipes and metallic fittings	TC 76	Transfusion, infusion and injection equipment for medical and pharmaceutical use	TC 146	Air quality	TC 207	Environmental management
TC 6	Paper, board and pulps	TC 77	Products in fibre reinforced cement	TC 147	Water quality	TC 209	Cleanrooms and associated controlled environments
TC 8	Ships and marine technology	TC 79	Light metals and their alloys	TC 148	Sewing machines	TC 210	Quality management and corresponding general aspects for medical devices
TC 10	Technical product documentation	TC 81	Common names for pesticides and other agrochemicals	TC 149	Cycles	TC 211	Geographic information/Geomatics
TC 11	Boilers and pressure vessels	TC 83	Sports and recreational equipment	TC 150	Implants for surgery	TC 212	Clinical laboratory testing and in vitro diagnostic test systems
TC 12	Quantities and units	TC 84	Devices for administration of medicinal products and intravascular catheters	TC 153	Valves	TC 213	Dimensional and geometrical product specifications and verification
TC 14	Shafts for machinery and accessories	TC 85	Nuclear energy	TC 154	Processes, data elements and documents in commerce, industry and administration	TC 214	Elevating work platforms
TC 17	Steel	TC 86	Refrigeration and air-conditioning	TC 155	Nickel and nickel alloys	TC 215	Health informatics
TC 20	Aircraft and space vehicles	TC 87	Cork	TC 156	Corrosion of metals and alloys	TC 216	Footwear
TC 21	Equipment for fire protection and fire fighting	TC 89	Wood-based panels	TC 157	Mechanical contraceptives	TC 217	Cosmetics
TC 22	Road vehicles	TC 91	Surface active agents	TC 158	Analysis of gases	TC 218	Timber
TC 23	Tractors and machinery for agriculture and forestry	TC 92	Fire safety	TC 159	Ergonomics	TC 219	Floor coverings
TC 24	Particle characterization including sieving	TC 93	Starch (including derivatives and by-products)	TC 160	Glass in building	TC 220	Cryogenic vessels
TC 25	Cast irons and pig irons	TC 94	Personal safety -- Protective clothing and equipment	TC 161	Control and protective devices for gas and/or oil burners and appliances	TC 221	Geosynthetics
TC 26	Copper and copper alloys	TC 96	Cranes	TC 162	Doors and windows	TC 223	Societal Security
TC 27	Solid mineral fuels	TC 98	Bases for design of structures	TC 163	Thermal performance and energy use in the built environment	TC 224	Service activities relating to drinking water supply systems and wastewater systems - Quality criteria of the service and performance indicators
TC 28	Petroleum products and lubricants	TC 100	Chains and chain sprockets for power transmission and conveyors	TC 164	Mechanical testing of metals	TC 225	Market, opinion and social research
TC 29	Small tools	TC 101	Continuous mechanical handling equipment	TC 165	Timber structures	TC 226	Materials for the production of primary aluminium
TC 30	Measurement of fluid flow in closed conduits	TC 102	Iron ore and direct reduced iron	TC 168	Prosthetics and orthotics	TC 227	Springs
TC 31	Tyres, rims and valves	TC 104	Freight containers	TC 170	Surgical instruments	TC 228	Tourism and related services
TC 33	Refractories	TC 105	Steel wire ropes	TC 171	Document management applications	TC 229	Nanotechnologies
TC 34	Food products	TC 106	Dentistry	TC 172	Optics and photonics	TC 230	Psychological assessment (PC)
TC 35	Paints and varnishes	TC 107	Metallic and other inorganic coatings	TC 173	Assistive products for persons with disability	TC 231	Brand valuation (PC)
TC 36	Cinematography	TC 108	Mechanical vibration, shock and condition monitoring	TC 174	Jewellery	TC 232	Learning services for non-formal education and training
TC 37	Terminology and other language and content resources	TC 109	Oil and gas burners	TC 176	Quality management and quality assurance	TC 234	Fisheries and aquaculture
TC 38	Textiles	TC 110	Industrial trucks	TC 177	Caravans	TC 235	Rating services (PC)
TC 39	Machine tools	TC 111	Round steel link chains, chain slings, components and accessories	TC 178	Lifts, escalators and moving walks	TC 236	Project Management (PC)
TC 41	Pulleys and belts (including veebelts)	TC 112	Vacuum technology	TC 180	Solar energy	TC 237	Exhibition terminology (PC)
TC 42	Photography	TC 113	Hydrometry	TC 181	Safety of toys	TC 238	Solid biofuels
TC 43	Acoustics	TC 114	Horology	TC 182	Geotechnics	TC 239	Network services billing (PC)
TC 44	Welding and allied processes	TC 115	Pumps	TC 183	Copper, lead, zinc and nickel ores and concentrates	TC 240	Product recall (PC)
TC 45	Rubber and rubber products	TC 117	Fans	TC 184	Automation systems and integration	TC 241	Road-Traffic Safety Management System (PC)
TC 46	Information and documentation	TC 118	Compressors and pneumatic tools, machines and equipment	TC 185	Safety devices for protection against excessive pressure	TC 242	Energy Management (PC)
TC 47	Chemistry	TC 119	Powder metallurgy	TC 186	Cutlery and table and decorative metal hollow-ware	TC 243	Consumer product safety (PC)
TC 48	Laboratory equipment	TC 120	Leather	TC 188	Small craft	TC 244	Industrial furnaces and associated thermal processing equipment
TC 51	Pallets for unit load method of materials handling	TC 121	Anaesthetic and respiratory equipment	TC 189	Ceramic tile	TC 245	Cross-border trade of second-hand goods (PC)
TC 52	Light gauge metal containers	TC 122	Packaging	TC 190	Soil quality	TC 246	Anti-counterfeiting tools (PC)
TC 54	Essential oils	TC 123	Plain bearings	TC 192	Gas turbines	TC 247	Fraud countermeasures and controls
TC 58	Gas cylinders	TC 126	Tobacco and tobacco products	TC 193	Natural gas	TC 248	Sustainability criteria for bioenergy (PC)
TC 59	Building construction	TC 127	Earth-moving machinery	TC 194	Biological evaluation of medical devices	TC 249	Traditional chinese medicine
TC 60	Gears	TC 130	Graphic technology	TC 195	Building construction machinery and equipment	TC 250	Sustainability in event management (PC)
TC 61	Plastics	TC 131	Fluid power systems	TC 197	Hydrogen technologies	TC 252	Natural gas fuelling stations for vehicles (PC)
TC 63	Glass containers	TC 132	Ferroalloys	TC 198	Sterilization of health care products	JTC 1	Information technology
TC 67	Materials, equipment and offshore structures for petroleum, petrochemical and natural gas industries	TC 134	Fertilizers and soil conditioners	TC 199	Safety of machinery	JTC 2	Energy efficiency and renewable energy sources (JPC)
TC 68	Financial services	TC 135	Non-destructive testing	TC 201	Surface chemical analysis		
TC 69	Applications of statistical methods	TC 136	Furniture	TC 202	Microbeam analysis		
TC 70	Internal combustion engines	TC 137	Footwear sizing designations and marking systems	TC 203	Technical energy systems		

IEC Structure



IEC TCs

FUJITSU

TC 1	Terminology	TC 38	Instrument transformers	TC 85	Measuring equipment for electrical and electromagnetic quantities
TC 2	Rotating machinery	TC 39	Electronic tubes	TC 86	Fibre optics
TC 3	Information structures, documentation and graphical symbols	TC 40	Capacitors and resistors for electronic equipment	TC 87	Ultrasonics
TC 4	Hydraulic turbines	TC 42	High-voltage test techniques	TC 88	Wind turbines
TC 5	Steam turbines	TC 44	Safety of machinery - Electrotechnical aspects	TC 89	Fire hazard testing
TC 7	Overhead electrical conductors	TC 45	Nuclear instrumentation	TC 90	Superconductivity
TC 8	Systems aspects for electrical energy supply	TC 46	Cables, wires, waveguides, R.F. connectors, R.F. and microwave passive components and accessories	TC 91	Electronics assembly technology
TC 9	Electrical equipment and systems for railways	TC 47	Semiconductor devices	TC 93	Design automation
TC 10	Fluids for electrotechnical applications	TC 48	Electromechanical components and mechanical structures for electronic equipment	TC 94	All-or-nothing electrical relays
TC 11	Overhead lines	TC 49	Piezoelectric, Dielectric and Electrostatic Devices and Associated Materials for Frequency Control, Selection and Detection	TC 95	Measuring relays and protection equipment
TC 13	Electrical energy measurement, tariff- and load control	TC 51	Magnetic components and ferrite materials	TC 96	Transformers, reactors, power supply units, and combinations thereof
TC 14	Power transformers	TC 55	Winding wires	TC 97	Electrical installations for lighting and beaconing of aerodromes
TC 15	Solid electrical insulating materials	TC 56	Dependability	TC 99	System engineering and erection of electrical power installations in systems with nominal voltages above 1 kV a.c. and 1,5 kV d.c., particularly concerning safety aspects
TC 16	Basic and safety principles for man-machine interface, marking and identification	TC 57	Power systems management and associated information exchange	TC 100	Audio, video and multimedia systems and equipment
TC 17	Switchgear and controlgear	TC 59	Performance of household and similar electrical appliances	TC 101	Electrostatics
TC 18	Electrical installations of ships and of mobile and fixed offshore units	TC 61	Safety of household and similar electrical appliances	TC 103	Transmitting equipment for radiocommunication
TC 20	Electric cables	TC 62	Electrical equipment in medical practice	TC 104	Environmental conditions, classification and methods of test
TC 21	Secondary cells and batteries	TC 64	Electrical installations and protection against electric shock	TC 105	Fuel cell technologies
TC 22	Power electronic systems and equipment	TC 65	Industrial-process measurement, control and automation	TC 106	Methods for the assessment of electric, magnetic and electromagnetic fields associated with human exposure
TC 23	Electrical accessories	TC 66	Safety of measuring, control and laboratory equipment	TC 107	Process management for avionics
TC 25	Quantities and units	TC 68	Magnetic alloys and steels	TC 108	Safety of electronic equipment within the field of audio/video, information technology and communication technology
TC 26	Electric welding	TC 69	Electric road vehicles and electric industrial trucks	TC 109	Insulation co-ordination for low-voltage equipment
TC 27	Industrial electroheating	TC 70	Degrees of protection provided by enclosures	TC 110	Flat panel display devices
TC 28	Insulation co-ordination	TC 72	Automatic controls for household use	TC 111	Environmental standardization for electrical and electronic products and systems
TC 29	Electroacoustics	TC 73	Short-circuit currents	TC 112	Evaluation and qualification of electrical insulating materials and systems
TC 31	Equipment for explosive atmospheres	TC 76	Optical radiation safety and laser equipment	TC 113	Nanotechnology standardization for electrical and electronic products and systems
TC 32	Fuses	TC 77	Electromagnetic compatibility	TC 114	Marine energy - Wave, tidal and other water current converters
TC 33	Power capacitors and their applications	TC 78	Live working	TC 115	High Voltage Direct Current (HVDC) transmission for DC voltages above 100 kV
TC 34	Lamps and related equipment	TC 79	Alarm and electronic security systems	TC 116	Safety of hand-held motor-operated electric tools
TC 35	Primary cells and batteries	TC 80	Maritime navigation and radiocommunication equipment and systems	JTC 1	Information technology
TC 36	Insulators	TC 81	Lightning protection	JTC 2	Energy efficiency and renewable energy sources
TC 37	Surge arresters	TC 82	Solar photovoltaic energy systems		

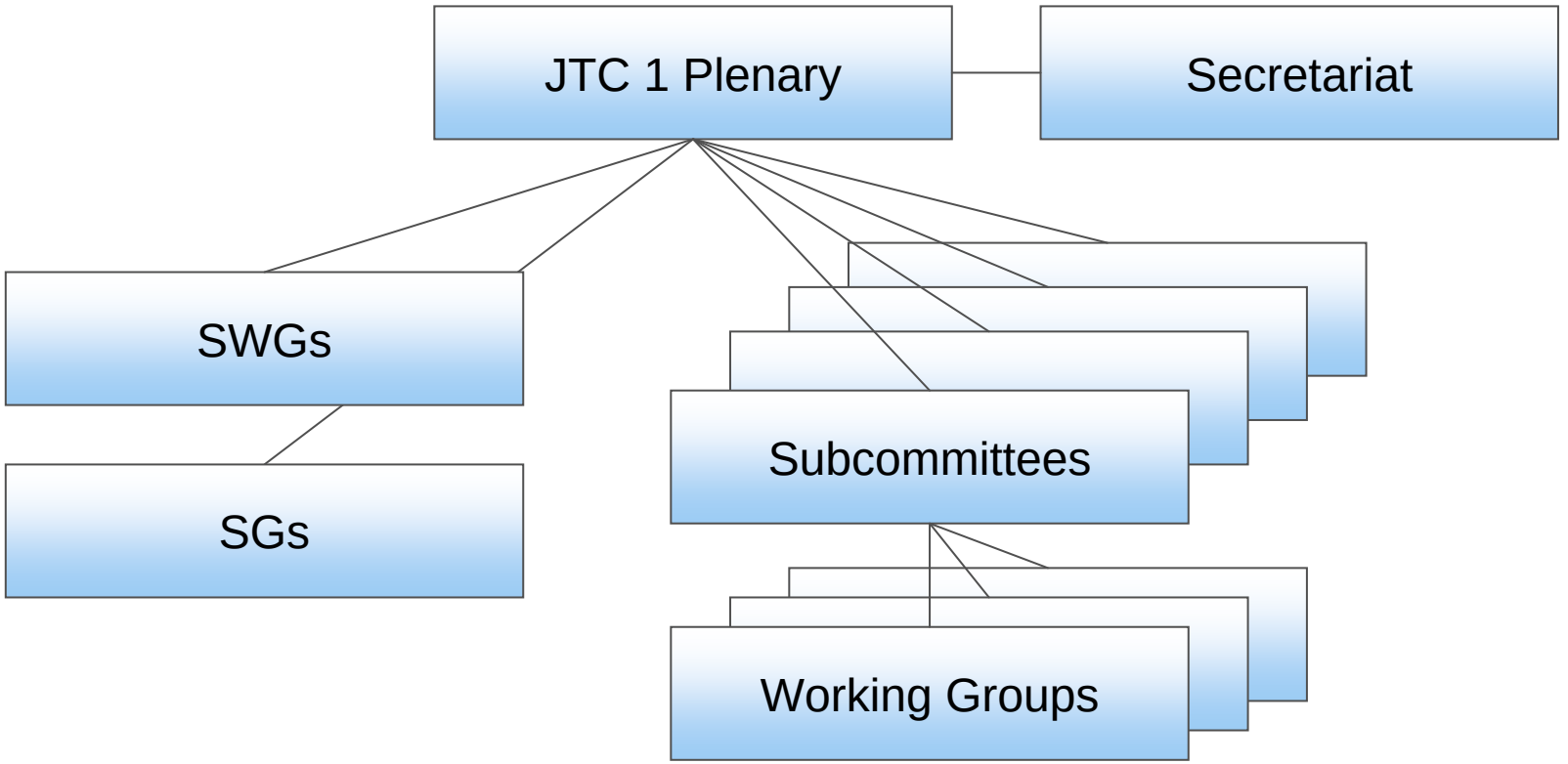
JTC 1 — IT Standardization

FUJITSU

- Joint Technical Committee 1
  - Covers Information Technology
  - Jointly hosted by ISO and IEC
    - JTC 1 reports both to ISO/TMB and to IEC/SMB
  - Also working closely with ITU

- JTC 1/SC 2
  - Covers coded character set (and sorting) standards
  - A subcommittee under the JTC 1

JTC 1 Structure



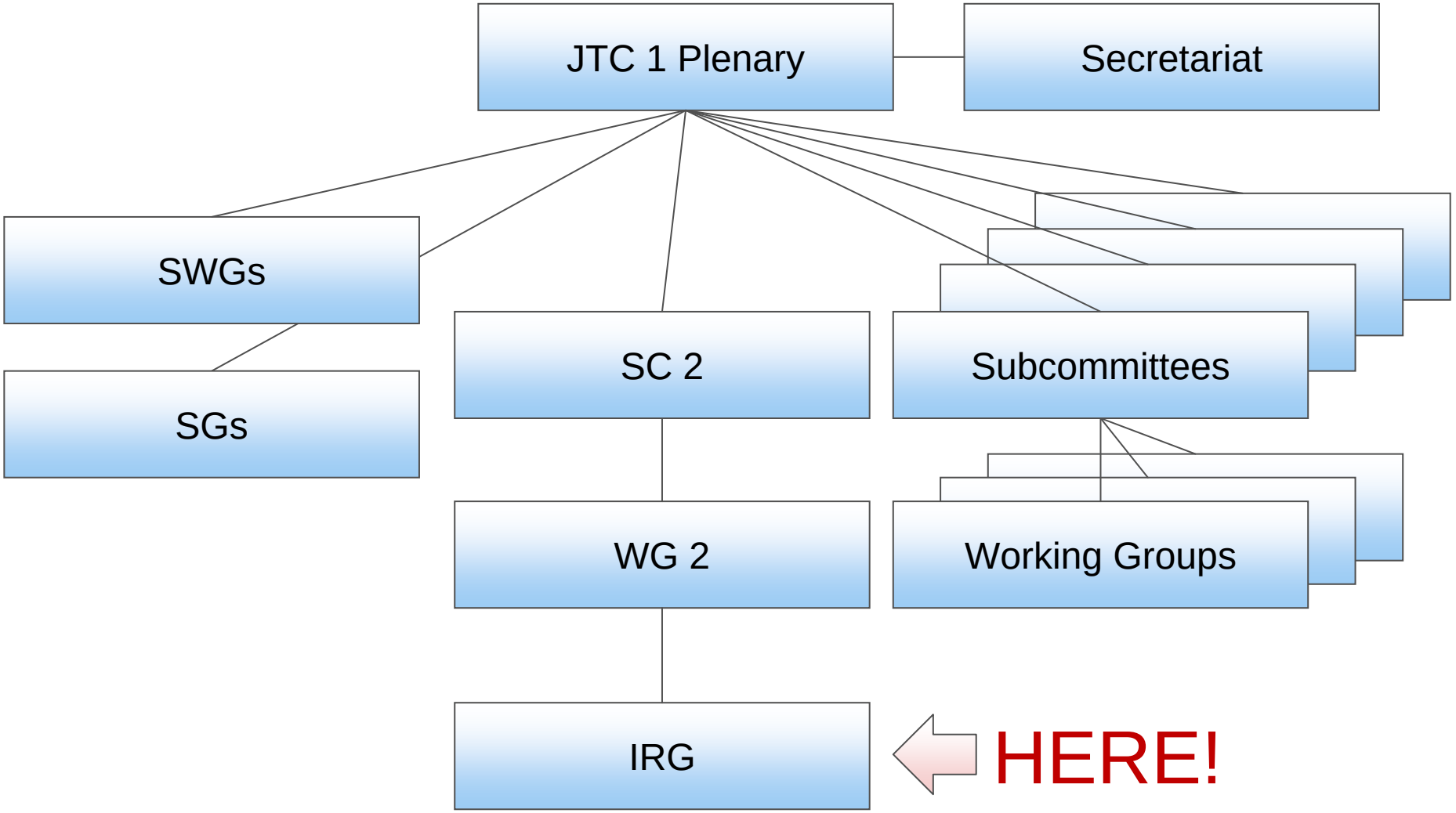
JTC 1 WGs and SCs

FUJITSU

WG 6	Corporate Governance of IT
WG 7	Sensor networks
SC 2	Coded character sets
SC 6	Telecommunications and information exchange between systems
SC 7	Software and systems engineering
SC 17	Cards and personal identification
SC 22	Programming languages, their environments and system software interfaces
SC 23	Digitally Recorded Media for Information Interchange and Storage
SC 24	Computer graphics, image processing and environmental data representation
SC 25	Interconnection of information technology equipment
SC 27	IT Security techniques
SC 28	Office equipment
SC 29	Coding of audio, picture, multimedia and hypermedia information
SC 31	Automatic identification and data capture techniques
SC 32	Data management and interchange
SC 34	Document description and processing languages
SC 35	User interfaces
SC 36	Information technology for learning, education and training
SC 37	Biometrics
SC 38	Distributed application platforms and services (DAPS)

IRG's Position

FUJITSU



Member Bodies

FUJITSU

- *Members* of ISO, IEC, or ITU are *countries/states*
  - Sometimes there are politically complicated issues on “What is a country/state” or “Is one particular *entity* a country/state?”.
  - We ask United Nations for consultation in the case.
- Designated national standards bodies or government agencies represent the member countries.
  - E.g., JISC (Japanese Industrial Standards Committee), a government-funded standards organization, represents Japan in ISO and IEC.

Japanese National Body

FUJITSU

- JISC（日本工業標準調査会）
  - “Japanese Industrial Standards Committee”
  - The official body funded by Japanese Government
- ITSCJ（情報規格調査会）
  - “Information Technology Standards Commission of Japan”
  - A part of IPSJ (Information Processing Society of Japan: 情報処理学会) that is primarily an academic organization
  - Consigned by JISC, representing Japan in JTC 1

Summary (2)

FUJITSU

- Three major international standards bodies: ISO/IEC/ITU.
- ISO and IEC jointly host **JTC 1** that works on Information Technology standards.
- Standards organizations have their hierarchical structures. Technical works are divided by areas.
- The *members* are countries/states, not companies nor individuals.

FUJITSU

Coded character set Standard for CJK Ideographs

- What is CJK Ideographs?
- What is CJK unification
- Why CJK ideographs needs to be a part of International Standards?

On CJK Ideographs

FUJITSU

- CJK Ideograph is:
  - Abstraction of ideographic characters, called kanji (漢字) in Japanese, hanzi (汉字) in Chinese, etc.
  - A script (writing system) originally established in ancient China.
  - Adopted widely in East Asia, as Chinese culture influenced nearby regions.
- There are a lot of ideographs!
  - Around 2000-3000 ideographs for daily use. For example,
    2000 in 常用漢字表 in Japan
    3500 in 现代汉语常用字表 in China.
    Much more to cover proper nouns.
    And much much more to cover academic, historical, or special purpose usages.
- Sometimes various forms (字体) exist for a same ideograph
  - Sometimes called variants

CJK Unification

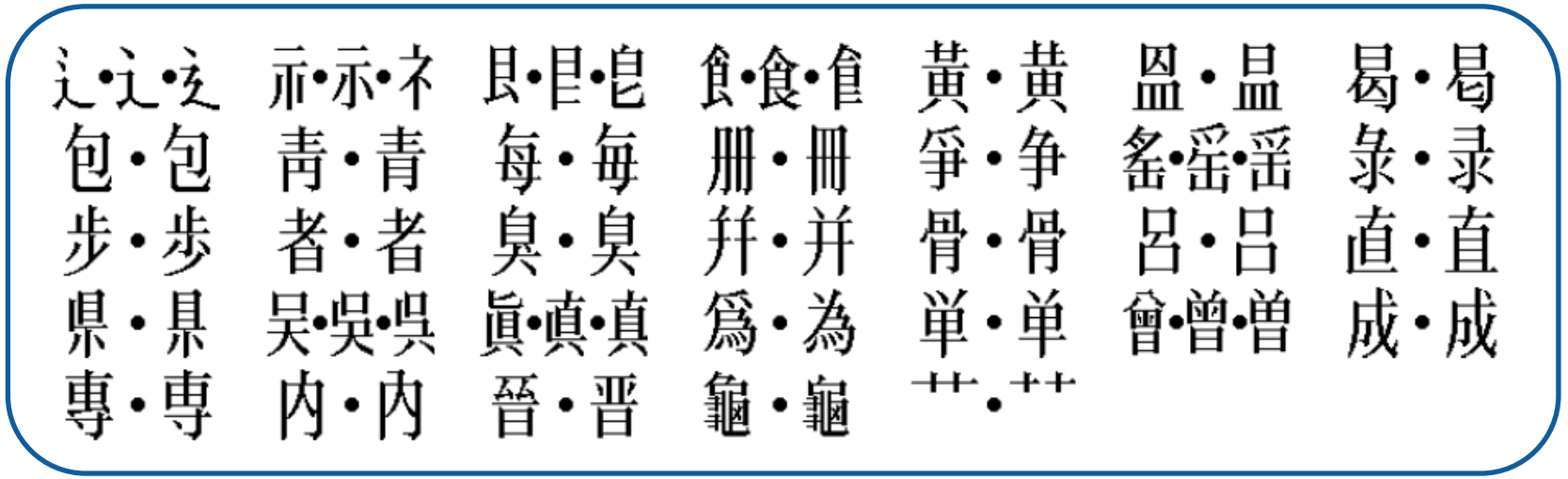
FUJITSU

- What is CJK unification?
  - Assign one code point to a same ideograph used in China, Japan, Korea, or other regions.
  - OK… But, what does “same ideograph” mean?
- Unification rules
  - Established a set of criteria to determine “same ideograph”.
  - A collaborative work by experts from various countries/regions, including China, Japan, Korea, US, …
  - The rule never unifies unrelated (called non-cognate) ideographs.
  - Variants having different structure are also excluded from unification.
  - Compares the shapes of the components.
  - Now a part of the international standard (ISO/IEC 10646)
  - Primarily compatible with pre-Unicode national standards.

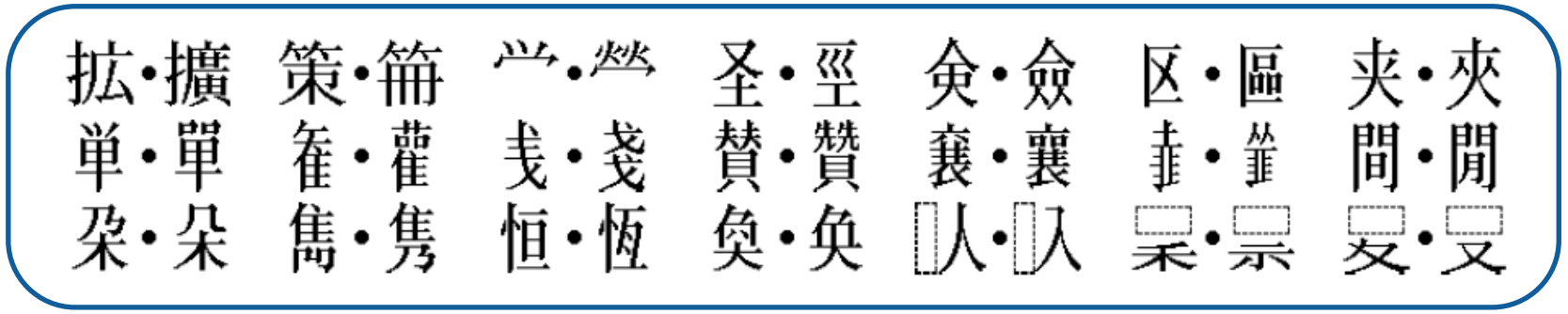
Examples

FUJITSU

Unified components, i.e., to consider as identical



Disunified components, i.e., to consider as separate ones

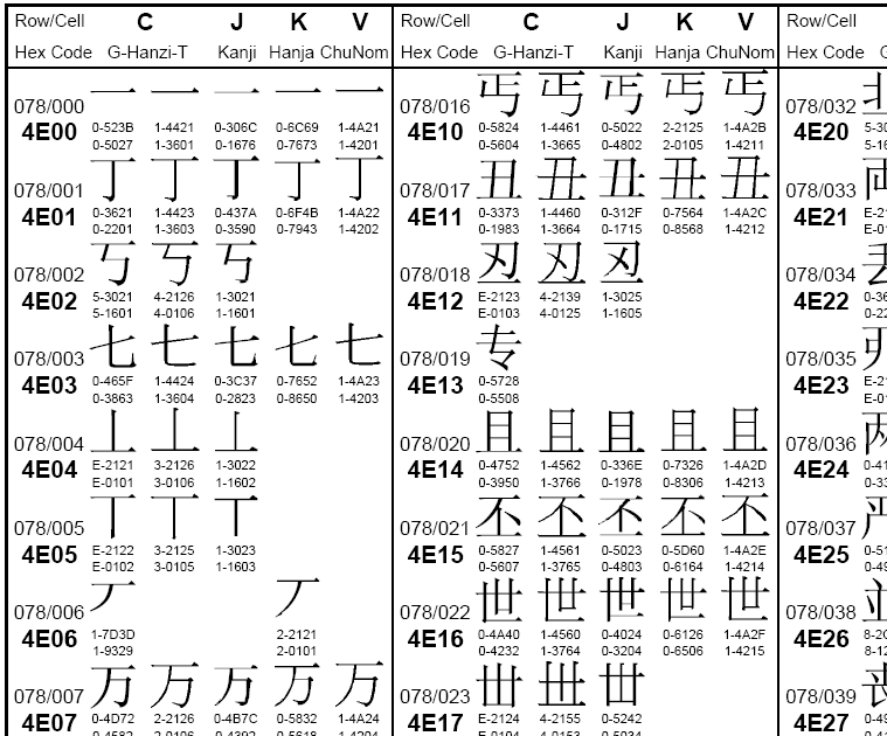


Code chart example

FUJITSU

- The first part of the code chart for CJK Unified Ideographs from ISO/IEC 10646
  - Shows up to 5 glyphs per a code point
  - Each column corresponds to a region
  - To show differences in preferred forms
  - Has gaps where not applicable (It doesn't mean the character is not used in the region)

ISO/IEC 10646:2003 (E)



Row/Cell	C		J	K	V
Hex Code	G-Hanzi-T		Kanji	Hanja	ChuNom
078/000	一	一	一	一	一
4E00	0-523B 0-5027	1-4421 1-3601	0-306C 0-1676	0-6C69 0-7673	1-4A21 1-4201
078/001	丁	丁	丁	丁	丁
4E01	0-3621 0-2201	1-4423 1-3603	0-437A 0-3590	0-6F4B 0-7943	1-4A22 1-4202
078/002	丂	丂	丂		
4E02	5-3021 5-1601	4-2126 4-0106	1-3021 1-1601		
078/003	七	七	七	七	七
4E03	0-465F 0-3863	1-4424 1-3604	0-3C37 0-2823	0-7652 0-8650	1-4A23 1-4203
078/004	丄	丄	丄		
4E04	E-2121 E-0101	3-2126 3-0106	1-3022 1-1602		
078/005	丅	丅	丅		
4E05	E-2122 E-0102	3-2125 3-0105	1-3023 1-1603		
078/006	𠂇			𠂇	
4E06	1-7D3D 1-9329			2-2121 2-0101	
078/007	万	万	万	万	万
4E07	0-4D72	2-2126	0-4B7C	0-5832	1-4A24

Row/Cell	C		J	K	V
Hex Code	G-Hanzi-T		Kanji	Hanja	ChuNom
078/016	丐	丐	丐	丐	丐
4E10	0-5824 0-5604	1-4461 1-3665	0-5022 0-4802	2-2125 2-0105	1-4A2B 1-4211
078/017	丑	丑	丑	丑	丑
4E11	0-3373 0-1983	1-4460 1-3664	0-312F 0-1715	0-7564 0-8568	1-4A2C 1-4212
078/018	刃	刃	刃		
4E12	E-2123 E-0103	4-2139 4-0125	1-3025 1-1605		
078/019	专				
4E13	0-5728 0-5508				
078/020	且	且	且	且	且
4E14	0-4752 0-3950	1-4562 1-3766	0-336E 0-1978	0-7326 0-8306	1-4A2D 1-4213
078/021	丕	丕	丕	丕	丕
4E15	0-5827 0-5607	1-4561 1-3765	0-5023 0-4803	0-5D60 0-6164	1-4A2E 1-4214
078/022	世	世	世	世	世
4E16	0-4A40 0-4232	1-4560 1-3764	0-4024 0-3204	0-6126 0-6506	1-4A2F 1-4215
078/023	丗	丗	丗		
4E17	E-2124	4-2155	0-5242		

Row/Cell	
Hex Code	G
078/032	丠
4E20	5-30 5-16
078/033	丒
4E21	E-2 E-0
078/034	丟
4E22	0-36 0-22
078/035	丣
4E23	E-2 E-0
078/036	丙
4E24	0-41 0-33
078/037	严
4E25	0-51 0-49
078/038	丠
4E26	8-20 8-12
078/039	丱
4E27	0-49

Why Kanji is a prt of I.S.?

FUJITSU

- Users' view
  - Chinese is actually a world's major language.
  - Even Japanese is not only used in Japan.
    It is used world-wide (although not as a majority language.)
  - It's better that ICT devices available locally can handle Japanese language than that a user needs to import specialized ICT devices from Japan to use Japanese language.
    - Imagine PCs sold in Japan can only handle Japanese and we can't use English on them…

Why Kanji is… (cont.)

FUJITSU

- Industries' view
  - Global companies want to sell same products world-wide. Chinese/Japanese/Korean language processing feature should not be an exception.
  - It lowers cost, so consumers get benefits, too.
  - It does not prevent market-oriented customize.
  - A single international standard coded character set that covers world-wide language best suits their needs. It must definitely cover CJK ideographs.

Pre-Unicode Days

FUJITSU

- Before UCS/Unicode came (early 1990s), there were no such thing of “cover all” character set standard.
- Each country/region has its own standard coded character set to cover its own languages only.
  - Some multi-lingual countries such as India needed to use multiple domestic coded character set standards because of technical limitations.
- ICT devices needed to be customized for each market region to support the required languages before sold.

… UCS/Unicode changed it!

These Days

FUJITSU

- Yesterday, a brand-new phone was shipped worldwide, and they are exactly the same product on all markets.
- It was enabled by International Standards (at least partly)
  - Not just by UCS/Unicode, of course.

Summary (3)

FUJITSU

- CJK Ideographs is an abstraction of those characters for Japanese/Chinese/…
- Not all people in the world require ideographs, but it should be in the International Standard Coded character set UCS/Unicode. It facilitates both users and industries.

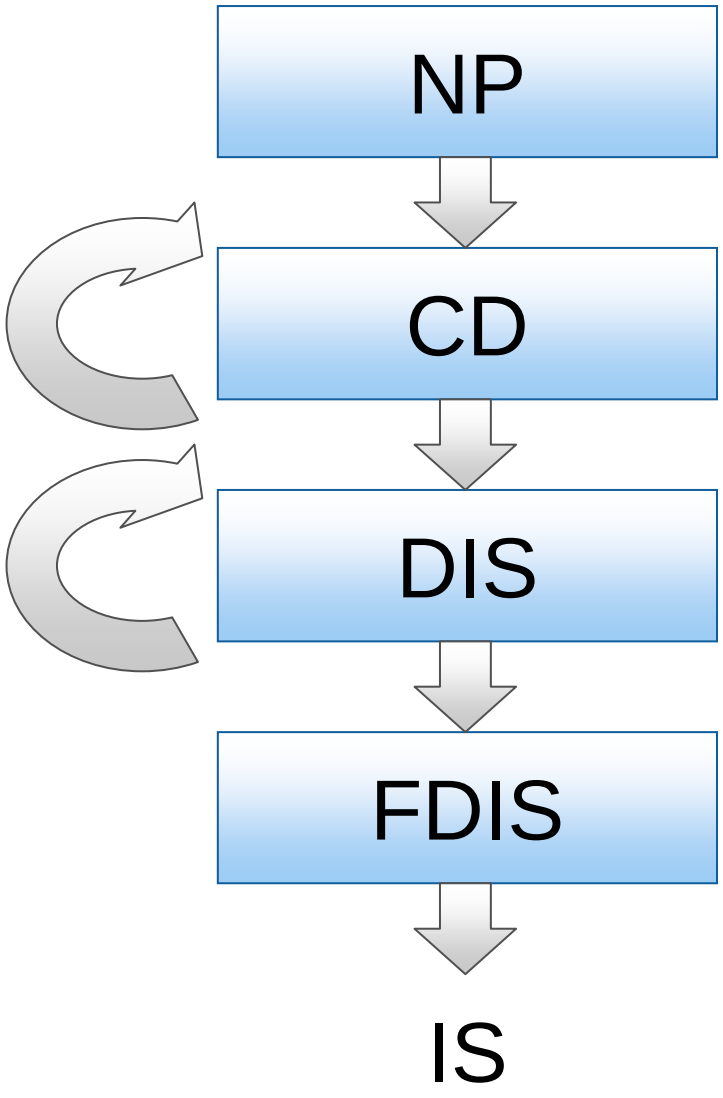
Standardization Processes

- JTC 1 Process
- Domestic Processes
- Who makes decision?

JTC 1 Normal Process

FUJITSU

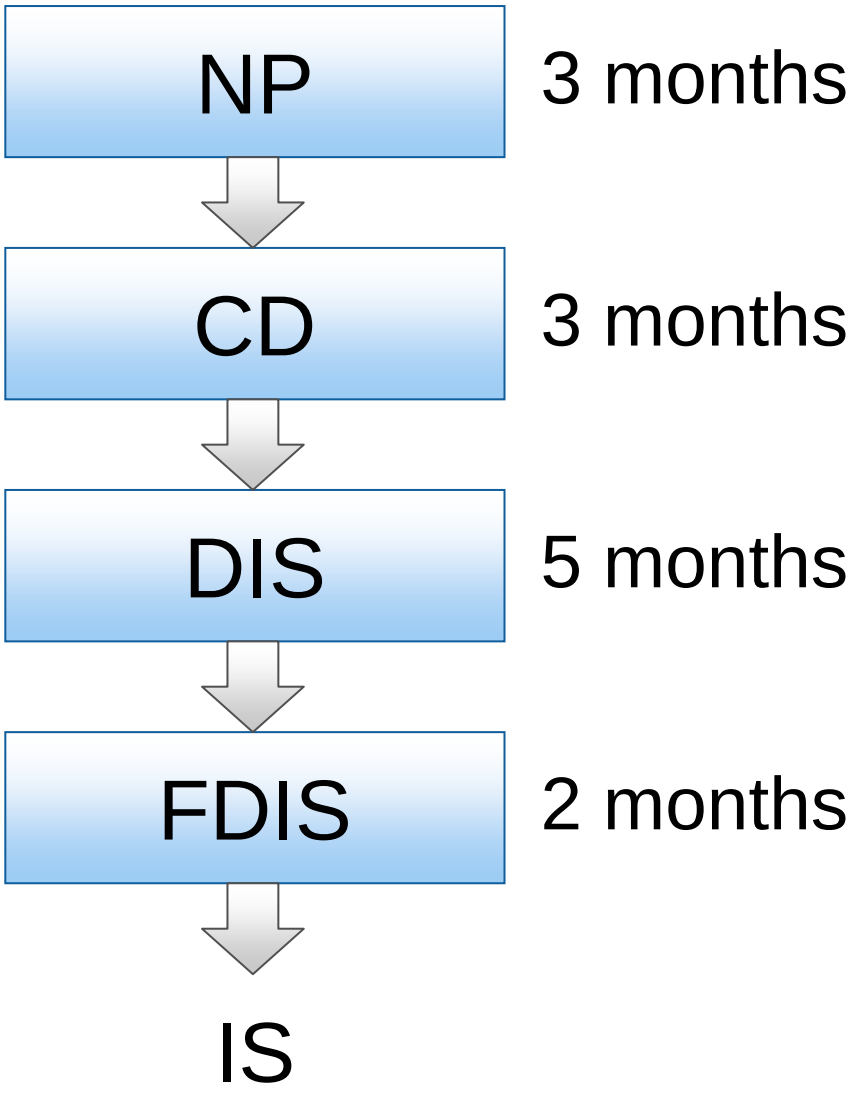
- Standards drafts advance in steps.
  - NP: New Proposal
  - CD: Committee Draft
  - DIS: Draft International Std.
  - FDIS: Final Draft Intl. Std.
- Member bodies vote on each step.
  - Upon failure, CD/DIS ballot may be repeated after modification to the draft text.



Ballot duration

FUJITSU

- Each ballot has defined duration.
  - So, we need at least 13 months without time for discussion.
  - Usually 2 or 3 months for discussion between ballots.
- 2 years or more is usual minimum.



What happens in a ballot

FUJITSU

- One ballot consists of:
  - Circulation of the document under ballot.
  - Member bodies examine the document and decides their positions.
  - Member bodies vote.
  - They also write any concerns on the document (called comments) and send it.
  - Secretary counts the votes after closing the ballot.
  - Usually the committee discusses the collected comments to improve the draft text.

Domestic Processes

FUJITSU

- Each member body may decide its position (voting) by its own way.
- In Japan, slightly different process takes place per each step. For an example, the process for the DIS voting is:
  - A technical committee to form a consensus.
  - ITSCJ to approve the position.
  - METI (Ministry of Economy, Trade, and Industry) to confirm the position.

Who makes decision?

FUJITSU

- Positions on technical issues are decided by technical committee members by consensus.
- Majority of technical committee members consists of experts from industries (i.e., who are working for commercial companies.) and researchers (e.g., university professors.)
- It is a part of the Government's bureaucracy, and we need some paperwork every time we vote on something…

How to participate?

FUJITSU

- You can always raise a proposal or make a comment as an individual.
  - Through your national body, through a liaison organization, or directly to the international committee.
- If you want to join the committee,
  - You need to contact an appropriate body.
  - In case of Japan (ITSCJ), we have some rules on the committee membership. The easiest way is through a company working in the area.
  - Please ask me details if you are serious.

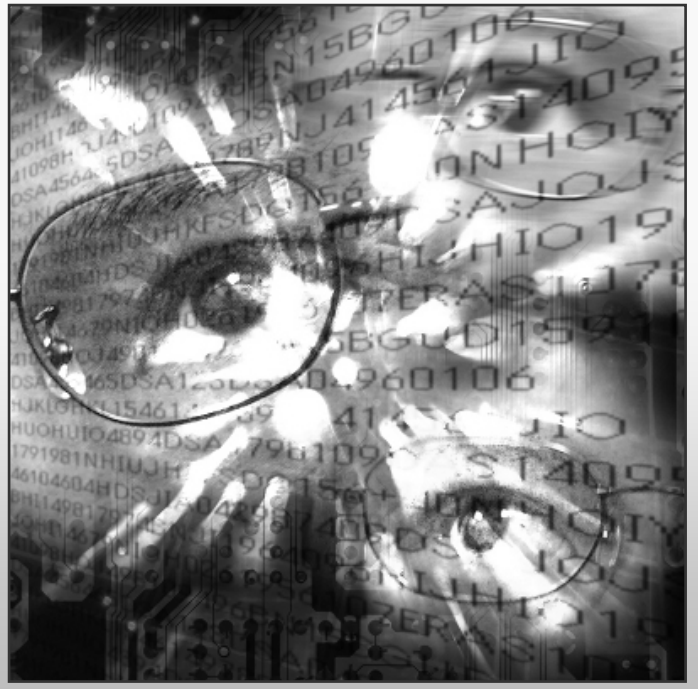
Summary (4)

FUJITSU

- There are defined processes to create International Standards
- ISO/IEC process is somewhat lengthy.
  - It requires about two years to establish a standard.
- Member bodies have their own rules on decision making.
- Ultimately, though, the International Standards are developed by consensus of motivated experts in the area.

Recent Topics

- Ideographic Variants
- *Emoji*



Ideographic Variants

FUJITSU

- CJK Unified Ideographs is after the CJK Unification rules. Some minor differences of actual shapes are ignored (not distinguished.)
- However, in areas such as legal handling of proper nouns, we sometimes need to handle such minor differences separately.

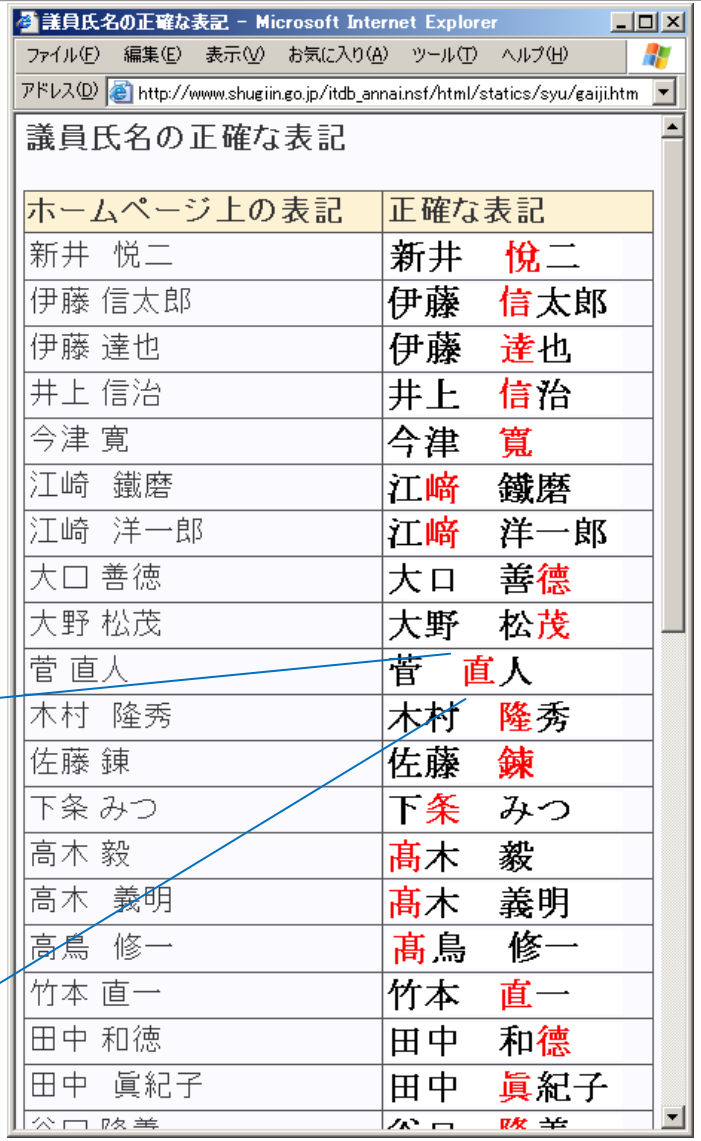
- Some sort of a gap between the standard and the real world

From the Diet web site

FUJITSU

- A page from the web site of the Japanese Diet, showing “Correct notations of some representatives’ names”
  - Many small differences are noted
- The Diet is not an exception; Many commercial companies do similar thing in low mandated disclosure through Internet, for example.

直 直



議員氏名の正確な表記

ホームページ上の表記	正確な表記
新井　悦二	新井　悦二
伊藤 信太郎	伊藤　信太郎
伊藤 達也	伊藤　達也
井上 信治	井上　信治
今津 寛	今津　寛
江崎　鐵磨	江﨑　鐵磨
江崎　洋一郎	江﨑　洋一郎
大口 善徳	大口　善徳
大野 松茂	大野　松茂
菅 直人	菅　直人
木村　隆秀	木村　隆秀
佐藤 錬	佐藤　錬
下条 みつ	下条　みつ
高木 毅	高木　毅
高木　義明	高木　義明
高鳥　修一	高鳥　修一
竹本 直一	竹本　直一
田中 和徳	田中　和徳
田中　眞紀子	田中　眞紀子

http://www.shugiin.go.jp/itdb_annai.nsf/html/statics/syu/gaiji.htm

From an Ordinance

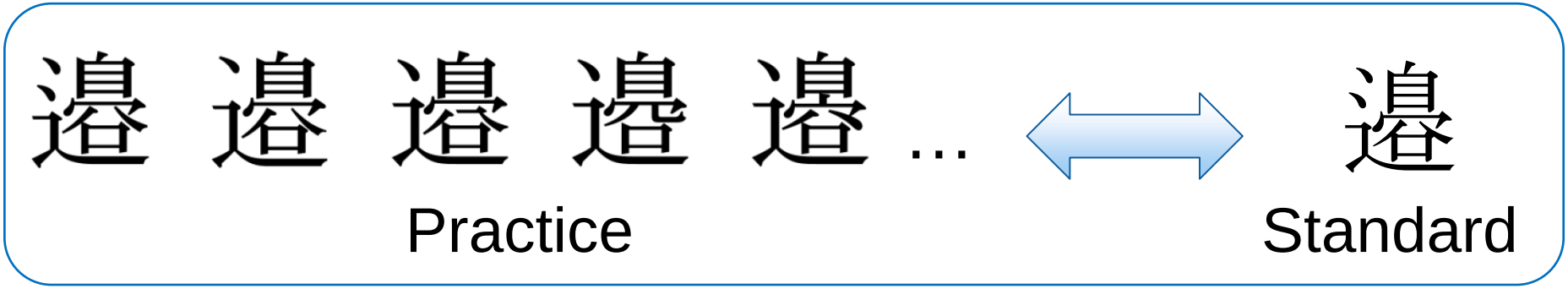
FUJITSU

- **Examples from Appended Table of Supplementary Provisions of the Ordinance for Enforcement of Family Register Act: List of permissible forms (戸籍法施行規則附則別表: 許容字体表)**
  - In practice, the list is often considered as a guideline for personal name handling.

The Problem

FUJITSU

- We somehow need to handle those variant separately on computer systems where laws or practices require it.
- The standard (UCS/Unicode) unifies those variants, giving only one code point.
- Use of an ad hoc technique breaks interoperability.



Possible Solutions

FUJITSU

- CJK Compatibility Ideographs
  - Allocate separate code points for ideographic variants, (in a sense) ignoring the CJK Unification.
  - Stable technology with some limitations.
- Ideographic Variation Sequences
  - Designate variation shapes for a CJK Unified Ideograph (a code point) with special auxiliary codes (variation selector).
  - We need further study how to use it in real applications. (At least in my opinion.)

IVD

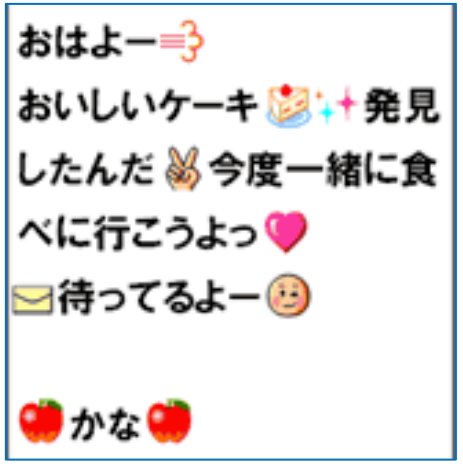
FUJITSU

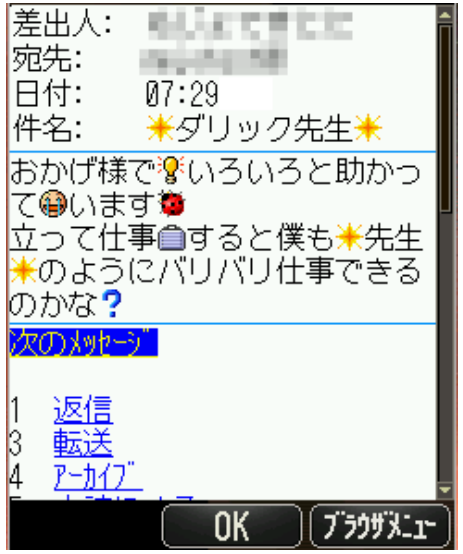
- Ideographic Variation Database
  - A central repository for variants.
  - Maintained by the Unicode Consortium through registration processes.
- IVS is based on IVD
  - Registering a variant assigns an ID.
  - An IVS for the variant is constructed using the ID.
- It is a part of UCS/Unicode specification

Emoji

FUJITSU

- Japanese mobile phones (ケータイ) support emoji (絵文字)
  - They are not part of the official standards and allocated as *private use* characters. Allocation is specific to network operator.
  - Emoji may be lost when sending to subscribers of different operators or to PCs.



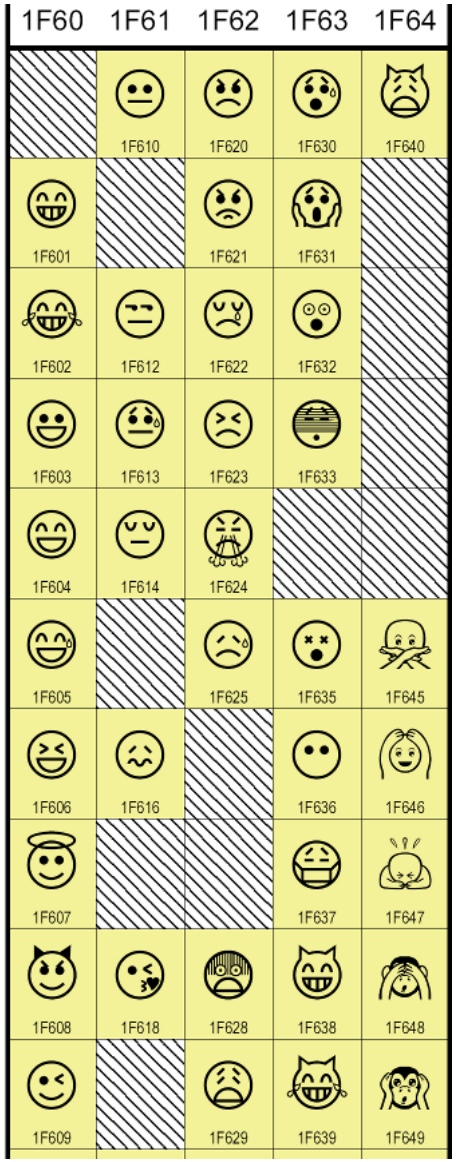


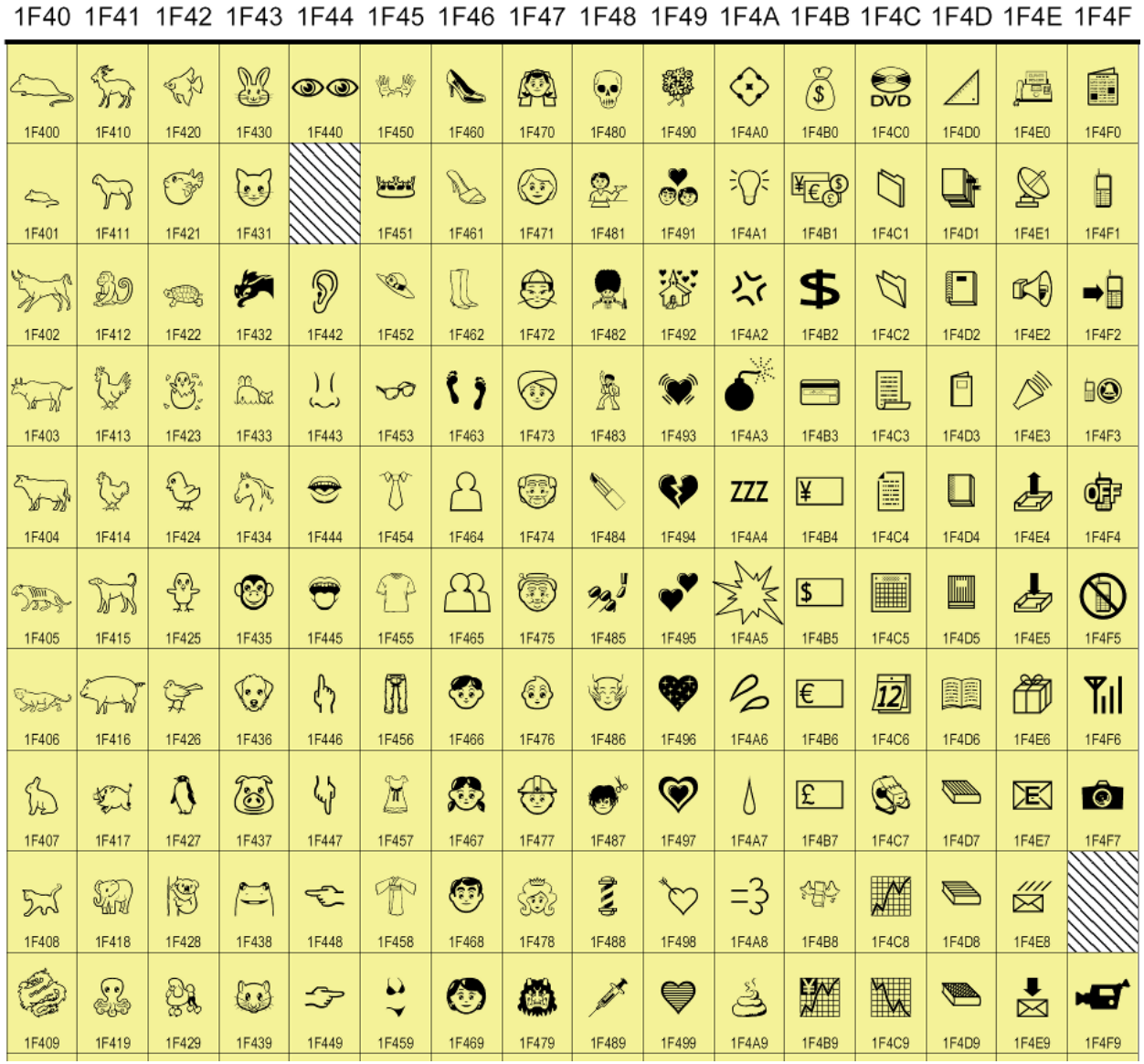
Emoji Addition to UCS

FUJITSU

- Emoji Proposal
  - Originally written by people working for Google and Apple
  - ISO/IEC JTC 1/SC 2/WG 2 finalized the draft at its last meeting (Apr. this year.)
- ≈ 600 new symbols to cover existing *emoji* set.
- ISO/IEC 10646:2003/Amd.8 containing emoji will soon be under its FDAM ballot
  - FDAM is an equivalent to FDIS step.

Examples





What happen next?

FUJITSU

- FDAM ballot
  - I expect it passes.
- Implementation
  - Google officially states that they will support UCS/Unicode emoji in its web services.
- Beyond Japanese market
  - Many people outside of Japan seems interested in using the emoji symbols.

Summary (5)

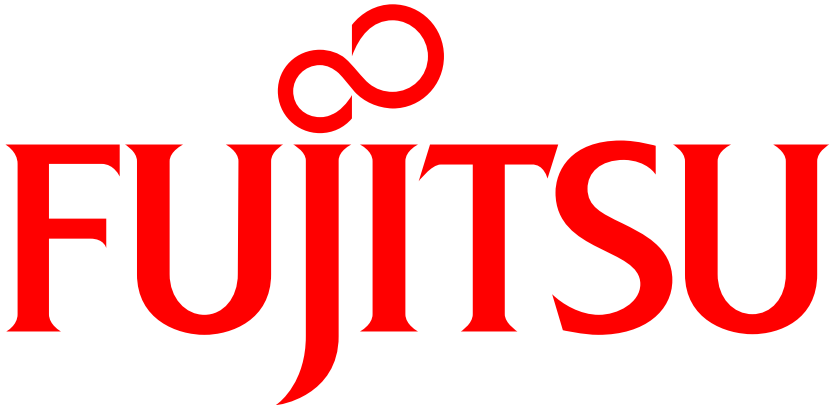
FUJITSU

- Ideographic Variants
  - Some applications require separate handling of variants that UCS/Unicode unifies.
  - IVS (Ideographic Variation Sequence) is a solution.
- Emoji (ケータイ絵文字)
  - Addition of several new symbols to support Japanese existing emoji set
  - Will soon be an official part of UCS/Unicode

Summary of Summaries

FUJITSU

- Three Major SDOs: ISO, IEC, ITU.
  - ISO and IEC jointly host JTC 1, working for Information Technology, including coded character sets.
- CJK Ideographs is a part of Intl. Standard.
- Standards bodies have process.
- Standards are developed by consensus of motivated experts.
- Two recent topics:
  - ideographic variants and emoji.



shaping tomorrow with you