CS101 – Програмчлалын үндэс

Лекц Суурь-03

Мэдээлэл

Профессор А.Эрдэнэбаатар

Лекцийн агуулга

- ▶ Мэдээлэл
- Мэдээллийн төрөл
- ▶ Мэдээллийн шинж, хэмжээ
- Мэдээллийн кодчилол
 - Текст
 - ▶ Бүхэл тоо
 - Бодит тоо
 - График
 - Дуу
 - Видео
- Мэдээллийн шахалт

Мэдээлэл

- Хүрээлэн буй орчны тухай хадгалах, боловсруулах, дамжуулах, ашиглах боломжтой өгөгдөхүүнийг хэлэх ба дохио, мэдээ, баримт зэргээр илэрхийлэгдэнэ
- Мэдээллийг олон шинжлэх ухаан судлах зүйлээ болгодог: мэдээллийн онол, кибернетик г.м.
- Хүн бүрийг янз бүрийн мэдээлэл хүрээлдэг ба түүнийг өөрийнхөөрөө тогтоож, хадгалах –ыг эрмэлэх нь хүнд байдаг шинж
- Хүний тархи мэдээллийг компьютертай адил хоёртын тоогоор хадгалдаг гэж үздэг, удаан хадгалах аргыг эрэлхийлж ирсэн
- Хүний өөр нэг шинж нь мэдээллийг бусадтай хуваалцаж дамжуулдаг, амьдралдаа хэрэглэдэг

Мэдээлэл

- Мэдээллийг дүрслэх хэлбэр, кодчилох ба хадгалах аргаар дараах төрөлд хувааж болно:
 - ▶ График хадгалах аргыг нь хамгийн түрүүн олсон төрөл (хадан дээрх зурагаас эхлээд фото зураг; цаас, даавуу, шаазан болон бусад материал дээрх схем, зураг)
 - Дуу бидний дуу хүрээлдэг, дууг бичих төхөөрөмжийг 1877 онд анх бүтээсэн. Дууны нэг хэлбэр эь хөгжмийн мэдээлэл ба нот бичлэгээр кодчилох аргыг бүтээсэн
 - Текст хүний яриаг тусгай тэмдэгтээр кодчилох арга. Хэл бүрт өөрийн багц тэмдэгт байдаг. Цаас онцгой ач холбогдолтой
 - Тоо хүрээлэн буй орчны обьект, түүний шинжийг тоогоор илэрхийлэх. Худалдаа, эдийн засаг мөнгөн солилцоо онцгой үүрэг гүйцэтгэсэн. Текстэй адил тусгай тэмдэгт ашигладаг
 - Видео мэдээлэл кино үүссэнээр хөдөлгөөнт дүрсийг хадгалах аргыг олсон
- Хадгалах аргыг нь олоогүй мэдээллийн төрөл байна: хүрэх, амтлах, үнэрлэх мэдрэмж

Мэдээлэл

- Мэдээллийг алс зайд дамжуулахдаа анх гэрлийн дохионы кодчилол ашигласан бол хочим нь цахилгаан дамжуулагч, радио долгионоор дамжуулах болсон
- Клод Шенноныг мэдээллийн онол, тооны холбооны үндэслэгч гэж үздэг ("A mathematical theory of communication", 1948)
- Орчин үеийн компьютерт мэдээллийг соронзон диск/лент, оптик диск, тусгай хагас дамжуулагч төхөөрөмж дээр хадгалж байна
- Глобаль сүлжээ "Интернет" дээрх мэдээлэл онцгой нэг төрөл юм

Мэдээлэл: Шинж

- Аливаа объектын адил мэдээлэл өөрийн шинжтэй:
 - ▶ Бодит байдал (objective) хүний ухамсраас ангид бодит ертөнцийн тусгал ("гудманд 22°", "гудманд дулаан").
 - Найдвартай байдал (reliability) байгаа зүйлийн илрэл. Бодит мэдээлэл ямагт үнэн. Үнэн мэдээлэл зөв шийдвэр гаргахад тусалдаг. Санаатай/санаандгүй байдлаар, эсвэл хэмжилтийн алдаанаас үнэн байдал алдагдаж болно
 - Бүрэн байдал (completeness) шийдвэр гаргахад хангалттай байх чанар
 - Унэн байдал (accuracy) бодит байдалд ойрын хэмжүүр
 - Ач холбогдол (timing) тухайн цагт хэрэгтэй байдал
 - Ашигтай байдал (relevance) хэрэгцээний нийцэл
- Дээрхийг хангасан мэдээлэл бол "Үнэ цэнэтэй" мэдээлэл

Мэдээлэл: Мэдээллийн хэмжээ

- Техникт мэдээллийн хэмжээ гэдгийг кодчилох, дамжуулах, хадгалах тэмдэгтийн тоог хэлдэг
- Компьютерт хэмжих нэгжээр бит (binary digit), байт (byte) г ашигладаг. Бит – хоёртын 0 эсвэл 1, Байт – 8 бит
 - 1 килобайт(Kb) = 2¹⁰b = 1024b
 - Arr 1 мегабайт(Mb) = 2^{10} Kb = 1024Kb = 1048576b
 - 1 гигабайт(Gb) = 2¹⁰Mb = 1024Mb = 1073741824b
 - 1 терабайт(Gb) = 2¹⁰Tb = 1024Tb = 1099511627776b г.м.
- Мэдээллийн онолд (К.Шеннон) мэдээллийн этропи гэдэг нь мэдээллийн санамсаргүй байдлын хэмжүүр бөгөөд ямар нэг анхдагч тэмдэгтийн гарч ирэх тодорхойгүй байдлыг илэрхийлдэг.
 п төлөвтэй санамсаргүй болох х үйлийн энтропи:

$$H(x) = -\sum_{i=1}^{n} p(i) \log_2 p(i)$$

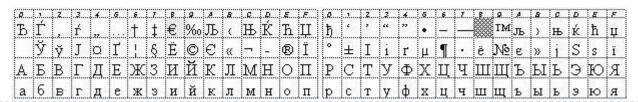
- Компьютерт хадалах, боловсруулах, сүлжээгээр дамжуулах аливаа мэдээллийг олон улсын стандартын дагуу хоёртын тоогоор дүрсэлдэг
- ▶ Текст мэдээлэл
 - Англи хэлтэй улсад том, жижиг үсэг (А...Z, а...z), 10 цифр (0...9), тусгай тэмдгүүд (.,:!"; '?()[]{}+-*?^% @ \$#&|\<>/_ г.м), зай тэмдэг зэрэг 100 тэмдэгт байдаг. Эдгээрийг кодчилоход 7 оронтой хоёртын тоо хангалттай (0000000...11111111 хүртэл нийт 128 тоо)
 - ▶ Ийм анхны кодчилолын хүснэгт нь ASCII (American Standard Code for Information Interchange) 1963 онд стандарт болж гарсан (хожим 8 битийн ISO/IEC 8859 ОУ –ын стандарт)

0	1	1 Z	: 3	4	6	6	7	В	g	A	8	C	D	Ε	F	0	1	Z	3	4	6	6	7	В	g	A	8	C	Ð	Ε	F
	0	0	*	٠	٠	±	•		0	0	8	9	ſ	J	✡	▶	4	\$		1	Ş		‡	1	1	\rightarrow	\leftarrow	L	\leftrightarrow	▲	¥
	!	"	#	\$	%	&	1	()	*	+	,	-		7	0	1	2	3	4	5	6	7	8	9		į,	<	=	>	?
(a)	Α	В	C	D	Ε	F	G	Η	Ι	J	K	L	Μ	Ν	0	Ρ	Q	R	S	Τ	U	V	W	Х	Y	Ζ	[١]	Λ	
7	а	ь	С	d	е	f	g	h	i	j	k	1	m	n	٥	p	q	r	s	t	u	٧	w	Х	у	z	{)	~	

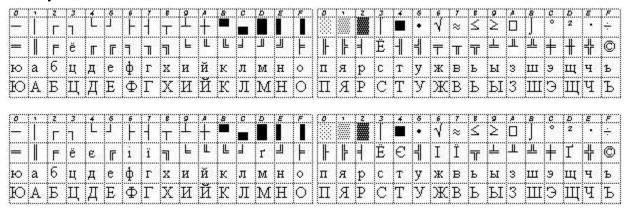
- ▶ ISO/IEC 8859 стандартаар ASCII хүснэгтийн сүүлийн 128 тэмдэгтийг өөр хэлэнд зориусан
 - ISO/IEC 8859-1:1998 Part 1: Latin alphabet No1
 - ► ISO/IEC 8859-5:1999 Part 5: Latin/Cyrillic alphabet



- ► ISO/IEC 8859-5:1999 Part 6: Latin/Arabic alphabet
- ▶ ISO/IEC 8859-5:2003 Part 7: Latin/Greek alphabet
- ▶ ISO/IEC 8859-5:1999 Part 8: Latin/Hebrew alphabet
- MS Windows гарч ирснээр ASCII хүснэгтэд кодын хуудас (CP Code Page) гэдэг ойлголт гарч ирсэн. Ж: кирил үсгийн CP1251



▶ Linux YC –д орос, украйн кирил үсгийг кодчилохдоо KOI8-R, KOI8-U хүснэгтийг ашигладаг



- ▶ Монгол улсын дотоодын стандартаар KOI8-U –н Є –г Ө, Ї -г Ү болгосон байдаг
- Бүх улсын кодыг нэг хүснэгтэнд оруулах зорилгоор 1991 онд Unicode Consortium байгуулагдаж 16 битийн UTF-16 стандарт гарсан

Unicode –н эхний 256 код ISO 8859-1 –тэй адил

▶ Кирил үсгийн Unicode дахь байршил: 0400₁₆ – 04FF₁₆

	040	041	042	048	044	045	046	047
		A	P	a •===	p		(£)	Ψ,
	Ë	Б	C	б	C	ë	W	Ψ
S	T)	В	T	B	T	ħ	Ъ	0
The court	ŕ	Г	У	Г sanş	y	ŕ	\$ ***	9
	€	Д	Ф	Д	ф	6	E	V
8	S	E	X	e	X	8	I€ mas	V 10%
2000	I	Ж	Ц	来	Ц	i	A	V Name
Ž	Ï	3	Ч	3	Ч	ï	A.	V Der
	J	И	Ш ==255	H ans	Ш	j	IA.	Oy
11000	Љ	Й	Щ	Ŭ ang	Щ	Љ	12A	OV
	Ь	K	Ъ	K	Ъ	њ	X.	0
SCS-83	ħ	Л	Ы	Л	ы	ħ	X.	0
N	Ŕ	M	Ь	M	Ь	Ŕ	HX.	(5)
100		H 1210	Э	H	Э		HX.	රා
Section	ÿ	0	Ю	O	Ю	ÿ	ž	(Ē)
18	Ħ	П	R	П	R	ĮĮ.	XXX	W

8	048	049	04A	04B	04C	04D	04E	04F
	ς	Γ	К	¥ 	I	Ă	3	ÿ
L	ς 0431	L,	IC on	¥	來	ă	3	ÿ
	≠	F	H	X 9892	滋	Ä	Й	ý œ
ſ	75 0437	£	H,	X,	R	ä	Й 0453	ű "
	^> 0454	Б	H	Щ	15	Æ	Й	Ÿ
Ī	5	5	H	Щ		æ	й	ÿ
	٥ ••••	Ж	Щ	Ч		Ĕ	Ŏ	
XIIII		X,	П _Б	Ч, ₀₄₈ г	Ĥ	ě	Ö oer	
SILLIN		3	Q	Ч •••	Ĥ	Q	0	Ы
SULL		3	Q	प्		9	0	й
XIIII		K	Ç	h		Ë	ĕ	
Alle		K,	Ç	h see	Ч	3 ****	ë	
SULL		K	Т	e	ч	米		
SULL		K	Т	(e)		滋		
Silling		K	Y	ę		3	ÿ	
Alle	1111	k	Y	ę		3	ÿ	

▶ Бүх кодыг ижил урттай кодчилох нь мэдээллийг хадгалах, дамжуулахад тохиромжтой бус байсан тул хувьсах урттай UTF-8 хөгжсөн

Кодын дүрслэлийн хязгаар	UTF-8	Тайлбар
000000 – 00007F	0xxxxxx	Эхний бит 0, бусад нь 7 битийн ASCII
000080 – 0007FF	110xxxxx 10xxxxxx	Нийт 2 байт. Эхний байт 110; 2 дахь байт 10 –р эхлэнэ
000800 – 00FFFF	1110xxxx 10xxxxxx 10xxxxxx	Нийт 3 байт. Эхний байт 1110; 2, 3 дахь байт 10 –р эхлэнэ
010000 – 10FFFF	11110xxx 10xxxxxx 10xxxxxx 10xxxxxx	Нийт 4 байт. Эхний байт 11110; 2, 3, 4 дэх байт 10 –р эхлэнэ

- ▶ Тоон мэдээлэл (бүхэл тоо)
 - Бүхэл тоо, бодит тоон (хөвөгч таслалтай) дүрслэл байдаг
 - Тэмдэггүй (эерэг) бүхэл тоо (unsigned): 8, 16, 32 дараалсан битээр дүрслэнэ. Дүрсэлж болох нийт тоо 2⁸ 1= 255(8 бит), 2⁸ 1= 65 535(16 бит), 2³² 1= 4 294 967 295(32 бит)
 - ▶ Тэмдэгтэй (сөрөг) бүхэл тоо (signed): 8, 16, 32 дараалсан битээр дүрслэхдээ эхний бит 1. Сөрөг тоог дүрслэхдээ шууд бус хоёртын гүйцээлтээр дүрслэнэ. Ө.х.
 - **2^N A**, Үүнд: N дүрслэх нийт битийн тоо (8, 16, 32 г.м.) А тооны шууд дүрслэл (эерэг утга)
 - 8 битийн дүрслэлд -1 нь 11111111, -2 нь 11111110 болно
 - Дүрсэлж болох нийт сөрөг тоо 8 битийн дүрслэлд
 [-128(10000000) 127(11111111)], 16 битэд [-32768 32767],
 32 битэд [-2147483648 2147483647]

- ▶ Тоон мэдээлэл (бодит тоо)
 - ▶ Бодит тоо X –г дүрслэхдээ нормаль хэлбэрт оруулах хэрэгтэй:

```
X = \pm M \cdot N^P, Yүнд: M — мантис(бутархай хэсэг), N — суурь, P — зэрэг
```

- ▶ 10 –тын системд $X = \pm M \cdot 10^P$, 2 –тын системд $X = \pm M \cdot 2^P$ болно. Ж: 22.22_{10} нь $\pm 0.2222 \cdot 10^2$ болно
- ▶ Иймд бодит тоог дараах байдлаар дүрсэлнэ (S тэмдэг)



 Бодит тооны өөр төрлийн шинжийг IEEE-754-1985 стандартад тусгасан

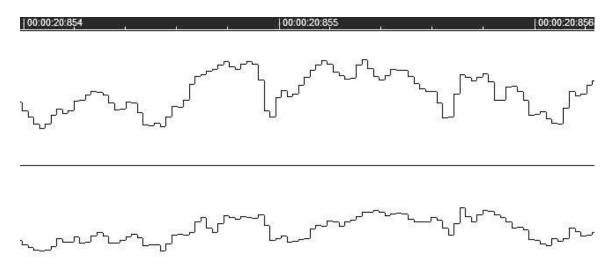
- ▶ Тоон мэдээлэл (бодит тоо)
 - ▶ Бодит тооны өөр төрлийн шинжийг IEEE-754-1985 стандартад тусгасан

Тороп	Хэмжээ,	Дүрслэлий	йн хязгаар	Нарийвчлал,	Алдаа		
Төрөл	бит	МИН	макс	цифр			
single	32	-3.4·10 ⁻³⁸	3.4·10 ⁻³⁸	6	1.192·10 ⁻⁷		
double	64	-1.7·10 ⁻³⁸	1.7·10 ⁻³⁰⁸	15	2.221·10 ⁻¹⁶		
Long double	80	-3.4·10 ⁻⁴⁹³²	3.4·10 ⁻⁴⁹³²	19	1.084·10 ⁻¹⁹		

Тоон мэдээлэл (график)

- ▶ График мэдээлэл хоёр хэмжээст муж бөгөөд түүний цэг (pixel) нь координат, өнгө, тунгалагжилт г.м. шинжээр илэрхийлэгдэнэ. Цэгүүдийн олонлогийг растер гэх ба вектор (шулуун, тойрог, тэгш өнцөгт г.м.) дүрслэлтэй байж болно
- ▶ Монитор болон бусад төхөөрөмж дээр босоо, хэвтээ чиглэлд дүрслэх цэгийн тоог нягтрал (resolution) гэнэ. 1920Х1080 г.м.
- Өнгийг RGB (Red-Green-Blue) загвараар 24 бит буюу 16.7 сая өнгөөр голдуу дүрсэлдэг (улаан <255, 0, 0>; ногоон <255, 255, 0>)
- Өнгийг бас CMYK (Cyan-Magenta-Yellow-black), HSV (Hue-Saturation-Value) загвараар дүрсэлж болно

- ▶ Тоон мэдээлэл (дуу)
 - Дууг тодорхой амплитут, давтамжтай багц синусиод долгион байдлаар дүрсэлж болно. Давтамж дууны өнгийг, амплитут чанга сулыг илэрхийлнэ. Дууны долгин 20 – 20000Hz



 Дууны өнгийг дууны карт 16 битээр илэрхийлвэл 65536 түвшин гарах ба хэмжилтийг голдуу 8 – 48КНz давтамжтай хийдэг

▶ Тоон мэдээлэл (видео)

- Видео бол хадгалах, боловсруулах, үзүүлэхэд хамгийн төвөгтэй мэдээллийн төрөл. Хөдөлгөөнт дүрсийг анх кино хальсанд 24кадр/сек хурдаар буулгаж байсан.
- Одоо телевизэд 25кадр/сек тэй PAL, SECAM систем, 29.97 кадр/сек тэй NTSC дирслэлийн системийг хэрэглэж байна.
- ▶ Видиео мэдээллийг кодчилохдоо MPEG (Moving Picture Experts Group) –ын анхны MPEG-1 стндарт 1992 онд гарсан. Одоо MPEG-7, MPEG-21
- Видео мэдээлэл өөрөө маш их орон зай эзэлдэг тул кодчилолын дээрх стандартуудыг шахалтын алгоритм гэж бас ойлгож болно

Мэдээлэл: Мэдээллийн шахалт

- Мэдээлийн хоёртын тоон дүрслэлд ерөнхийдөө тодорхой хэмжээний илүүдэл байдаг. Ижил бит, бүлэг бит цуварч давтагдах нь элбэг. Задгай мэдээллийг хадгалах, дамжуулахад орон зай, хугацаа их шаарддаг.
- Шахалт бол мэдээллийн технологид онцгой ач холбогдолтой
- Шахалтын шинж:
 - Шахалтын зэрэг анхдагч, үр дүнгийн харьцаа
 - Шаалтын хурд
 - Шахалтын чанар үр дүнд чанар буурах эсэх
- Шахалтын төрөл
 - Алдагдалтай буцаан задлахад анхдагч гарахгүй байх
 - Алдагдалгүй голдуу текст мэдээлэлд хэрэглэдэг

Мэдээлэл: Мэдээллийн шахалт

- Мэдээлийн хоёртын тоон дүрслэлд ерөнхийдөө тодорхой хэмжээний илүүдэл байдаг. Ижил бит, бүлэг бит цуварч давтагдах нь элбэг. Задгай мэдээллийг хадгалах, дамжуулахад орон зай, хугацаа их шаарддаг.
- Шахалт бол мэдээллийн технологид онцгой ач холбогдолтой
- Шахалтын шинж:
 - Шахалтын зэрэг анхдагч, үр дүнгийн харьцаа
 - Шаалтын хурд
 - Шахалтын чанар үр дүнд чанар буурах эсэх
- Шахалтын төрөл
 - Алдагдалтай буцаан задлахад анхдагч гарахгүй байх
 - Алдагдалгүй голдуу текст мэдээлэлд хэрэглэдэг

Мэдээлэл: Мэдээллийн шахалт

- Нийтлэг алдагдалтай шахалтууд: JPEG, TIFF, GIF, PNG фото зураг, MP3 – дуу, MPEG-4 – видео
- ▶ Нийтлэг алдагдалгүй шахалт: ZIP, BMP
- Сонгодог шахалтууд голдуу давтамж дээр суурилсан Хоффман болон Лемпел-Зив-Велчийн (LZW – Lampel-Ziv-Welch) алгоритмыг ашигладаг