Lecture 13: Heterogeneous Data & Heterogeneous Users

CMSC 25910 Spring 2024 The University of Chicago



Interfaces and data diversity

Data validation

- Are we considering all possible formats of the data?
- Are we forcing one answer when more than one applies?

Ethics	s20 Project Preferences				iQ Score: Great	Publishe	ed C	hange Question Type	
	ault Question Block				B	lock Options 🗸	Te	Text Entry	~
♥ Q1	Your name:							 Single Line Multi Line Essay Text Box Form Password 	
Q2	Your UChicago email address (e	.g., blase@uchicago.edu)	:	<i>i</i> e			Force Res	alidation Options Force Response Response Minimum Length	_
□ Q3 ¢	How excited or not excited are y	vou to do your class projec Very excited to work on it	ct in each area below? Excited to work on it	Somewhat ex work or		t excited		Maximum Length Character Range Content Validation Custom Validation	

Data validation

• Are we properly surfacing data errors?



Handling names in computer systems

- What policies should we enforce for names?
 - Which fields are offered? Which are required?
 - What minimum/maximum lengths do we allow?
 - Which characters are permitted?
 - What is our validation logic?

Interfaces forcing an identity

• Facebook required a binary gender upon initial registration

First name	Last name
Mobile numbe	r or email
New password	
Birthday	
May 🚽 4 🚽	1994 Why do I need to provide my birthday?
Female	Male

Facebook's first attempt to do better

	Gender	😫 Friends 👻
	Ge	
	Gender Fluid	
	Gender Variant	the first states
	Genderqueer	\$
	Gender Questioning	e.
	Gender Nonconforming	
	Agender	
Birthday	Bigender	- ~ 2 전 시 것은 문제 내 전에 관
	Cisgender	
	Cisgender Female	
	Cisgender Male	

Image taken from: Catherine D'Ignazio and Lauren F. Klein, Data Feminism, The MIT Press, 2020

Facebook's improved attempt (free text)

Gender	Friends 🔻
	-
What pronoun do you use?	
Neutral: "Wish them a happy birthday!" \$	
Your pronoun is OPublic. Learn more.	
Your pronoun is Public and can be seen anyone.	by

Image taken from: Catherine D'Ignazio and Lauren F. Klein, Data Feminism, The MIT Press, 2020

Facebook's current registration interface

facebook

Connect with friends a around you on Facebo

First name)	0	Last na	ime		Jr.com
Mobile nur	nber or	email				Ø
New passv	word					Log in
irthday 🕜						
Mar	~	30	~	2021	~	Forgot Password?
ender 🕜						
Female	0	Male		Custom	0	reate New Account
y clicking Sign	Up, you ag	ree to our Terms	s, Data Poli	cy and Cookies F	Policy. You	

Facebook's current registration interface

facebook

Connect with friends a around you on Facebc

Please select your pronoun.

Sign Up It's quick and easy.	
First name	
Mobile number or email	ur.com
New password	
Birthday 🕜	
Mar • 30 • 2021 •	Log In
Gender 🕐	
Female O Male O Custom	
Select your pronoun	
Select your pronoun	reate New Account
She: "Wish her a happy birthday!"	
He: "Wish him a happy birthday!"	
They: "Wish them a happy birthday!"	ge for a celebrity, band or business.
Sign Up	

Facebook's current registration interface

Connect with friends a around you on Facebo

Sign Up t's quick and easy.	×
First name	
Mobile number or email	r.com
New password	Ø
irthday 🕜	
Mar ~ 30 ~ 2021	► Log In
Female O Male O Custom	Forgot Password?
Select your pronoun	9~
our pronoun is visible to everyone.	reate New Account
Gender (optional)	
By clicking Sign Up, you agree to our Terms, Data Policy and Cookies Police nay receive SMS Notifications from us and can opt out any time.	y. You ge for a celebrity, band or business.

Inferring identity

- Twitter guesses your gender if you don't provide it
 - From: https://reclaimthenet.org/twitter-assumes-gender-ads

If you haven't added a gender, this is the one most strongly asso account based on your profile and activity. This information won publicly.	
Female	(
Male	(
Add your gender	(
Fwitter's <u>reasoning for this</u> is that it wants to find use	Save
	U
interests" and characteristics – the interests that they	<i>really</i> have; not
just the ones they say they have.	

Who can access data in an interface?

Data access

- Slack: administrators *can* delete others' public messages
- Slack: administrators *cannot* access direct messages
- Zoom: administrators can request people unmute / turn on video
- Zoom: participants visible in meeting, but not in webcast
- Zoom: administrators can (retroactively) see participant times
- Unix OS: administrators can see everything

Stopping abuse

Anti-abuse

- Who chooses to block people?
 - On many sites, users can block others
 - Slack doesn't let a user block another

Retrospective editing

- Can you change what you've said in a platform after the fact?
 - Is there a marker? How do we balance accountability and privacy?

Retrospective editing

- Can you change what you've said in a platform after the fact?
 - Is there a marker? How do we balance accountability and privacy?

new post	/ Test 1 , a small edit.	[-] [deleted] 2 points 1 year ago	This is test 3
[deleted] ■ 2 Comments & Share …	This message has been removed.	 1) Brazil: 5 World Cups Germany: 4 WC, 3 Eur 	and 8 Co
(a) Deleted Reddit posts are denoted "[deleted]."	(b) Skype shows deletion with a trash can and edits with pencils.	(c) Reddit posts can be and changing the poster to "[de	-
Edit History	← Archive	a snap chat post	My post disappeared! Where'd it go? Threads expire and are pruned by 4chan's software at a
51 seconds ago This is test 1	Only you can see the posts		relatively high rate. Since most boards are limited to ten pages, content is usually available for only a few hours or days before it is removed. Most times, the post was probably pruned automatically, however in some cases it may have been removed by a moderator or janitor. We ask that users refrain from reposting material that they believe to have been deleted by the moderators.
26 seconds ago This is test 2		Image: Story Series Image: Story Series Image: Story Series	Can I retrieve an old post/image? No. Content that has expired is removed from our servers.
(e) Facebook provid a public edit history		(g) Snapchat allows dele- tion after a set time.	(h) 4chan deletes content automati- cally if there are no new replies.

Figure 1: Example retrospective mechanisms and their associated markers as currently deployed on popular platforms.

Image taken from: Günce Su Yılmaz, Fiona Gasaway, <u>Blase Ur</u>, Mainack Mondal. Perceptions of Retrospective Edits, Changes, and Deletion on Social Media. In *Proc. ICWSM*, 2021

What are your goals? What are your responsibilities? (Business vs. user goals)

Accessibility

Crucial considerations

- Cognitive impairments
 - Illiteracy
 - Memory difficulties
- Visual impairments
 - Sightedness
 - Color-blindness
- Dexterity impairments
- Hearing impairments

The importance of "alt" text

-
- This is what screen readers use!

The importance of video captions

Add your own subtitles & closed captions

Subtitles and closed captions open up your content to a larger audience, including deaf or hard of hearing viewers or those who speak languages besides the one spoken in your video. If you already have captions or subtitles, get help editing or removing existing captions.

- 1. Go to your Video Manager 🗹 by clicking your account in the top right > Creator Studio > Video Manager > Videos.
- 2. Next to the video you want to add captions or subtitles to, click the drop-down menu next to the Edit button.
- 3. Select Subtitles/CC.
- 4. Click the Add new subtitles or CC button.
- 5. Choose how you want to add or edit subtitles or closed captions to your video:

Create subtitles or closed captions	~
Upload a file	~
Transcribe and auto-sync (original video language only)	~
Use automatic captioning	~

The importance of fonts and text

- Websites / apps should still work when magnified to a large size
- Don't use color to convey meaning
 - Screen readers do not interpret it
- High-contrast color combinations
- Don't include placeholders in form fields
- Make form fields easy to find

Subtitles

Example from <u>https://www.youtube.com/watch?v=TjPp2TNqRgk</u>





Subtitles

SSA

- Speech-to-text software can generate initial (imperfect) subtitles
- Example file formats:
 - SRT (wide support)

Subtitle number Subtitle display timing 00:00:28,840 --> 00:00:34,460 My father went home - Subtitle text two hours ago yResx: 720 lavResy: 576 avpeptn: Timer: 100.0000 wrapStvle: 0 V4+ Styles Format: Name, Fontname, Fontsize, PrimaryColour, SecondaryColour, OutlineColour, BackColour, Bold, scalex, scaley, spacing, Angle, Borderstyle, outline, shadow, Alignment, MarginL, MarginR, MarginV, style: Dino, Arial, 30, &H0024DA4C&, &HffFFFFF&, &H009BD728&, &HffFF8080&, 0, 0, 0, 0, 100, 100, 0 Style: Texte, Arial, 50, &H009BD728&, &HffFFFFF&, &H0024DA4C&, &HffFF8080&, 0, 0, 0, 100, 100, [Events] Format: Layer, Start, End, Style, Name, MarginL, MarginR, MarginV, Effect, Text Dialogue: 0,0:00:00.00,0:00:15.00,Dino,0000,0000,0000,,{\pos(40,160)}{\p1+} m 36 51 1 42 54 1 48 48 93 39 87 36 b 78 54 b 81 51 78 45 78 42 b 93 51 102 51 108 48 b 96 24 63 27 51 27 b 39 27 39

Images taken from: https://www.checksub.com/subtitle/subtitle-formats-which-ones-to-use/

Readability

- Standard metrics for the grade level of text
 - <u>https://en.wikipedia.org/wiki/Readability#Popular_readability_formulas</u>
 - <u>https://www.online-utility.org/english/readability_test_and_improve.jsp</u>

Conducting accessibility studies

- Follow accessibility guidelines
 - <u>https://developer.gnome.org/accessibility-devel-guide/stable/gad-ui-guidelines.html.en</u>
- Don't forget to user test with diverse populations that accurately reflect all users
 - https://www.youtube.com/watch?v=t4DT3tQqgRM

Internationalization

	•	11	*	
	Public	Prieteni F	rsonalizat	
	۲		•	
	rați în legătură cu alții w you connect with people yo	ou know.	Schimbați setările	
-	ी कि से से 		58%	
Timeline	and Tagging			
Control wh timeline.	nat happens when friends tag	you or your content, or post on yo	our Schimbaţi setările	
Reclame	e, Aplicații și Website-u	ri		
Gestionea	ză-ți setările pentru reclame,	aplicații, jocuri și saituri.	Schimbaţi setările	
Limiteaz	ză audiența pentru pos	tările anterioare		
Limit the a	audience for posts you shared	d with friends of friends or Public	Administrează vizibilitate postărilor din trecut	



การทวีตสื่อ	😑 แสดงสื่อที่อาจจะมีเนื้อหาที่อ่อนไหว			
	回 ทำสัญลักษณ์ไว้ว่าสื่อของฉันนั้นมีเนื้อหาที่อ่อนไหว			
	หากคุณทวีตรูปภาพหรือวิดีโอที่อาจมีเนื้อหาอ่อนไหว โปรดเล [ื] อกซ [่] อง ตัวเลือกนี้เพื่อที่คนอื่นๆ จะได้รับคำเตือนก่อนรับชมสื่อของคุณ ศึกษา เพิ่มเติม			
ความเป็นส่วนตัวของ	🔲 ป้องกันทวีตของฉัน			
ทวีต	If selected, only those you approve will receive your Tweets. Your			
	future Tweets will not be available publicly. Tweets posted			
	previously may still be publicly visible in some places. Learn			
	more.			
การทำให้เป็นส่วน	Tailor Twitter based on my recent website visits			
บุคคล	Preview suggestions tailored for you (not currently available to all			
	users). Learn more about how this works and your additional privacy controls.			
	privacy controls.			
HTTPS เท่านั้น	🔲 ใช้ HTTPS เสมอ			
	ใช้การเชื่อมต่อแบบปลอดภัยเมื่อมีให้ใช้งาน			
รีเช็ตรหัสผ่าน	Require personal information to reset my password			
	By default, you can initiate a password reset by entering only your			
	@username. If you check this box, you will be prompted to enter			
	your email address or phone number if you forget your password.			



Translation Center вюд в	Forums			Need h	elp? - Blaze Cups -
Home / Twitter.com /			12		
All phrases	L	Intranslated Trans	slated/Voted Live	privacy	Search
 will not be shown publicly. You can change your privacy settings at any time. 1 translations with 0 votes 	121220 D			y identifying inf	
Twitter does not disclose personally identifying information to third parties except in accordance with our Privacy Policy. 2 translations with 1 votes	third partie	+ Add translation	i More information	vith our Privac	y Policy.
<a _blank"="" class="personalization-preview-
link" href="
{{web_personalization_preview_path}}
target=">Preview suggestions tailored for you (not currently available to all users). <a< th=""><td>Vote A Tr info eset Pos</td><td>0 votes ✓ Translation live</td></a<>	Vote A Tr info eset Pos	0 votes ✓ Translation live			
class="personalization-support-link" href="https://support.twitter.com/article target="_blank">Learn more about how this works and your additional privacy controls. 1 translations with 0 votes	A Twitter nem oszt meg személyes azonosításhoz használható információkat harmadik féllel, kivéve az Adatkezelési Nyilatkozatban leírt eseteket. Posted by Mazvazzeg on November 02, 2011			1 votes	

		Search		
Learn about cookies and Facebook.		Search translations	Q	
Szavazás Fordítás		Translation		
Tudj meg többet a sütikről és a Facebookról.	✓ X	Your Progress	-	
kapd be a faszt te kis buzi	✓ ×	0 translations	votes	
Tudj meg többet a sütikről és a Facebook-ról.	~ ×			
Tudj meg többet a cookiekról és a Facebookról.	¥ ×			
Tudj meg többet a facebookról és a "cookiekról"	✓ ×			
További 3 🔻				

Representing Written Language
Representing Text: ASCII (7 bits) ASCII TABLE

Decimal	Hex	Char	Decimal	Hex	Char	JDecimal	Hex	Char	Decimal	Hex	Char
0	0	[NULL]	32	20	[SPACE]	64	40	0	96	60	`
1	1	[START OF HEADING]	33	21	1	65	41	Α	97	61	а
2	2	[START OF TEXT]	34	22		66	42	В	98	62	b
3	3	[END OF TEXT]	35	23	#	67	43	С	99	63	с
4	4	[END OF TRANSMISSION]	36	24	\$	68	44	D	100	64	d
5	5	[ENQUIRY]	37	25	%	69	45	E	101	65	е
6	6	[ACKNOWLEDGE]	38	26	&	70	46	F	102	66	f
7	7	[BELL]	39	27	1.00	71	47	G	103	67	g
8	8	[BACKSPACE]	40	28	(72	48	н	104	68	ĥ
9	9	[HORIZONTAL TAB]	41	29)	73	49	1	105	69	i
10	А	[LINE FEED]	42	2A	*	74	4A	J	106	6A	j
11	В	[VERTICAL TAB]	43	2B	+	75	4B	K	107	6B	k
12	С	[FORM FEED]	44	2C	,	76	4C	L	108	6C	1
13	D	[CARRIAGE RETURN]	45	2D	- C	77	4D	М	109	6D	m
14	E	[SHIFT OUT]	46	2E	1.00	78	4E	Ν	110	6E	n
15	F	[SHIFT IN]	47	2F	1	79	4F	0	111	6F	0
16	10	[DATA LINK ESCAPE]	48	30	0	80	50	Ρ	112	70	р
17	11	[DEVICE CONTROL 1]	49	31	1	81	51	Q	113	71	q
18	12	[DEVICE CONTROL 2]	50	32	2	82	52	R	114	72	r
19	13	[DEVICE CONTROL 3]	51	33	3	83	53	S	115	73	S
20	14	[DEVICE CONTROL 4]	52	34	4	84	54	т	116	74	t
21	15	[NEGATIVE ACKNOWLEDGE]	53	35	5	85	55	U	117	75	u
22	16	[SYNCHRONOUS IDLE]	54	36	6	86	56	V	118	76	v
23	17	[ENG OF TRANS. BLOCK]	55	37	7	87	57	W	119	77	w
24	18	[CANCEL]	56	38	8	88	58	Х	120	78	x
25	19	[END OF MEDIUM]	57	39	9	89	59	Υ	121	79	У
26	1A	[SUBSTITUTE]	58	ЗA	1.00	90	5A	Z	122	7A	z
27	1B	[ESCAPE]	59	3B	;	91	5B	[123	7B	{
28	1C	[FILE SEPARATOR]	60	3C	<	92	5C	١	124	7C	
29	1D	[GROUP SEPARATOR]	61	3D	=	93	5D	1	125	7D	}
30	1E	[RECORD SEPARATOR]	62	3E	>	94	5E	^	126	7E	~
31	1F	[UNIT SEPARATOR]	63	ЗF	?	95	5F	_	127	7F	[DEL]
			-			-		_			

https://simple.wikipedia.org/wiki/ASCII

Extended ASCII (8 bits)

128	Ç	144	É	160	á	176		193	⊥	209	╤	225	ß	241	±
129	ü	145	æ	161	í	177	·····	194	т	210	π	226	Г	242	≥
130	é	146	Æ	162	ó	178		195	F	211	Ш	227	π	243	≤
131	â	147	ô	163	ú	179		196	_	212	F	228	Σ	244	ſ
132	ä	148	ö	164	ñ	180	4	197	+	213	F	229	σ	245	J
133	à	149	ò	165	Ñ	181	4	198	þ	214	Г	230	μ	246	÷
134	å	1 <i>5</i> 0	û	166	3	182	-	199	⊩	215	₽	231	τ	247	8
135	ç	151	ù	167	0	183	п	200	L	216	ŧ	232	Φ	248	•
136	ê	152	_	168	ż	184	Ŧ	201	F	217	L	233	۲	249	
137	ë	153	Ö	169	_	185	╢	202	ᄟ	218	Г	234	Ω	250	
138	è	154	Ü	170	_	186		203	╦	219		235	δ	251	\checkmark
139	ï	156	£	171	1∕2	187	ה	204	ŀ	220		236	ω	252	_
140	î	157	¥	172	1⁄4	188	IJ	205	=	221		237	ф	253	2
141	ì	158	_	173	i	189	Ш	206	ł	222		238	ε	254	
142	Ä	1 <i>5</i> 9	ſ	174	«	190	F	207	⊥	223		239	\sim	255	
143	Å	192	L	175	»	191	٦	208	ш	224	α	240	≡		

https://www.webopedia.com/definitions/extended-ascii/

Unicode

- Variable-width character encoding
 - <u>https://en.wikipedia.org/wiki/UTF-8</u>

Availability

Availability

- Are you assuming a technology that people don't have?
 COVID-19 contact tracing
- Are we assuming constant internet connectivity?
 - Are we assuming anything other than periodic connectivity?
- Are we assuming a high-bandwidth connection?
 - Are we designing for low-bandwidth situations?
 - Does our system adapt to these low-bandwidth situations?
- Are we assuming constant access to electricity?

Mobile Money



How to get started with M-Pesa



Mobile Money



Good Intentions

One Laptop Per Child (OLPC)

- Project centered in the MIT Media Lab ~2005
 - Funding came from companies like AMD, eBay, Google, Marvell, etc.
- Goal: Transform education around the world in part by distributing electronic devices around the world



Failure of One Laptop Per Child (OLPC)

 US-centric worldview; expensive; focus on computers (to the detriment of other concerns); minimal availability of software; problems with maintainability; minimal training/infrastructure; over-adherence to FOSS; negative environmental impacts; bugs





Legacy Systems

Legacy systems and legacy code

- A shocking amount of code written in languages like Fortran and COBOL is still in use!
- A shocking number of mainframe systems are still in use
- How do we deploy modern protections and modern performance enhancements to these systems?
 - For example, what if we want to roll out two-factor authentication?