Information Gathering: Unobstrusive Methods

CSE 4407

Md. Bakhtiar Hasan

Assistant Professor
Department of Computer Science and Engineering
Islamic University of Technology

June 30, 2025

Sampling: The What and Why

- What: Systematically selecting *representative* elements of a population
  - Goal: Analyze the sample to understand the *whole* population (documents, people, data, etc.)
  - Analyst's Challenge: Which documents? Which people?
- Why
  - Costs: Examining *everything* is expensive! $\rightarrow$ Time = Money
  - Speed: Faster insights by looking at less (but relevant) data
  - Effectiveness: Deeper analysis on a smaller set; better chance to follow up on details
  - Bias: Avoid skewed perspectives from non-representative individuals or data points

Designing Your Sample: The Blueprint

- Overview: To get meaningful results, your sample needs a plan!
- Steps
  - Data: What *exactly* are you trying to find out or describe? $\rightarrow$ Identify variables, attributes
  - Population: What is the *entire* group your sample should represent? (e.g., last 6 months reports? All managers? Website users?)
  - Sample Type: How will you select the elements?
  - Sample Size: How many elements do you *need*?

Types of Samples: Picking Your Players

- Non-Probability: Selection isn't random
  - Convenience: Easiest, but riskiest! (e.g., "Anyone free Tuesday morning?")
  - Purposive: Based on judgment (e.g., picking "knowledgeable" people)
- Probability: Every element has a known chance of selection
  - Simple Random
    - Equal chance for all (like drawing names from a hat)
    - Needs a numbered list
  - Complex Random
    - Systematic: Every kth item (e.g., every 10th invoice) $\rightarrow$ Watch out for hidden patterns
    - Stratified: Divide population into subgroups first, then sample from each (e.g., sample managers, analysts, programmers separately) $\rightarrow$ Ensures representation
    - Cluster: Select groups assumed to be representative (e.g., pick 2 out of 20 help desks to study)

Sample Size: How Many Are Enough?

- What: It's the *absolute number* in the sample that matters most, not just the percentage of the population
  - Analogy: Tasting 1 spoonful of soup tells you about the pot, whether it's a small pot or a huge cauldron
  - Sampling 20 people can be effective whether the population is 200 or 2,000,000 (if done right!)
- Factors Influencing Size
  - Cost and Time (Practical constraints)
  - Desired Confidence Level (How sure do you want to be?)
  - Acceptable Interval Estimate (How much error can you tolerate?)
- Two Scenarios
  - Sampling Quantitative Data (e.g., forms, records) $\rightarrow$ Use statistical formulae
  - Sampling People (e.g., interviews) $\rightarrow$ Use guidelines and judgment

Sample Size for Data: The Formula Approach

- Goal: Determine n (sample size) needed to estimate a proportion (e.g., % of forms with errors)
- Steps
  - Attribute: What are you measuring? (e.g., 'Order Form Error')
  - Data Source: Where is this info? (e.g., 'Order Database'/'Past 6 months forms')
  - Estimate p: Guess the proportion with the attribute (e.g., Estimate 5% or 0.05 have errors) $\rightarrow$ Use 0.5 if unsure, but this leads to large samples
  - Acceptable Interval i: How much error is okay? (e.g., ± 0.02)
  - Confidence Level: How sure? (e.g., 95%) Look up z-value
  - Standard Error: $\sigma_p = i/z$
  - Sample Size: $n = \frac{p \times (1-p)}{\sigma_p^2} + 1$

%	z
99	2.58
98	2.33
97	2.17
96	2.05
95	1.96
90	1.65
80	1.28
50	0.67

Example: A. Sembly Company Orders

- Problem: Find the percentage of orders with errors
- Steps
  - Attribute: Orders with errors (name, address, quantity, model #)
  - Source: Order forms from the past 6 months
  - Estimate p: Analyst estimate ~5% contain errors
  - Interval i: Analyst decides ± 0.02 is acceptable
  - Confidence z: 1.96
  - Calculate σ_p: $i/z = 0.02/1.96 = 0.0102$
  - Calculate n: $\frac{p \times (1-p)}{\sigma_p^2} + 1 = \frac{0.05 \times (1-0.05)}{(0.0102)^2} + 1 \approx 458$
- Conclusion: Need to sample 458 order forms
- Note: Higher confidence or smaller interval would require a *much larger* sample

Sample Size for Interviews: It's About Time!

- No Magic Formula: Determining interview sample size is more **art** than science
- Key Factor
  - Time commitment (yours and the interviewee's)
  - In-depth interviews are time-consuming
- Rule of Thumb
  - Interview **at least 3 people** from *each* organizational level involved (e.g., execs, managers, operational staff)
  - Interview **at least 1 person** from *each* functional area affected (e.g., Sales, Marketing, Production, IT)
- Stratification: A well-designed stratified sample means a relatively small number can represent the whole organization effectively
- Remember: You do not need huge numbers if you choose wisely $\rightarrow$ Quality over quantity

Analyzing Quantitative Documents: Playing Detective!

- What: The act of discovery and analysis
- Analyst Role: Like Sherlock Holmes - seeking clues in hard data
- Why
  - Offer insights unavailable elsewhere (historical performance, documented procedures)
  - Include: Reports, Records, Forms
  - Show where the organization *has been* and where it *thinks* it's going

Quantitative Docs 1: Performance Reports

- Purpose: Compare **Actual** vs. **Intended** Performance
  - Assess the *gap* between goals and reality
  - Identify trends: Is the gap widening or narrowing?
- Examples: Sales reports (revenue vs. target), Production reports (output vs. quota, defect rates), Service reports (delivery times vs. goals, satisfaction scores vs. benchmarks)
- Analyst Focus
  - Is performance measured for key areas?
  - Is the measurement adequate?
  - What story do the trends tell?

Week	Number of Batches Produced	Number of Batches Rejected	Percentage Rejected	Amount Away from 5% Goal
2/2	245	19	7.8	2.8
2/9	229	19	8.3	3.3
2/16	219	14	6.3	1.3
2/23	252	13	5.2	0.2
3/2	245	13	5.3	0.3
3/9	260	13	5.0	***
3/16	275	14	5.1	0.1
3/23	260	13	5.0	***
3/30	260	13	5.0	***
4/6	244	12	4.9	***
4/13	242	11	4.5	***
4/20	249	11	4.4	***
4/27	249	11	4.4	***

*** indicates met or exceeded the <5% goal

Quantitative Docs 2: Records

- Purpose: Provide periodic updates on business happenings → Offer rich detail if updated accurately and timely
- Analyst Focus
  - Errors: Check calculations, totals, data entry mistakes
  - Design Flaws: Is the record form confusing? Could it be improved?
  - Transaction Patterns: How many? What types? Frequency?
  - Automation Opportunities: Can calculation or data manipulation be computerized?

PROJ. NAME OAK. FC # 562 KEY SIGNATURE _____

RENT POTENTIAL				1175/0		81299		DEPOSIT POTENTIAL					PRORATE		15.50 121.12
Base Rent	Refrig erator	Furni ture	A/C	Util.	HMSR	T.V.	Maid	Total Rent	Securi ty	Clean ing	31175/0	81299	31700 Tax	Days	Daily Rate	Totals 5.20
855		55						910			H/S dep.	H/S rent		4	30.33/~~1.30~~	910 39
									200	115					Deposits 31.83	340

PAYMENT RECORD: Tot. 31175/0 + 81299 + Rent = 910 TOTAL INITIAL PAYMENT REQUIRED: 1430.52

Memo Only	Date Due	Date Paid	Receipt Number	Paid to Noon	Total Rent	Securi ty	Clear ning	31700 Tax	31175/0	81299		Other		Amount Paid	Balance Due
										Dates	Amt.	Descr.	Amt.		
TV 10/1 MO!	8/28	8/28	106642	9/30	1031.32	202	115	44.20	25			414.82	15	1430.52	~~0~~
	10/1	10/3	107503	10/31	910									910	~~0~~
	11/1	11/1	10935	11/16	485.28									485.28	~~0~~
C1H/S9-16	11/17	11/6	11200	11/23	212.31									212.31	~~0~~
Bill 1 MO	11/24														
Prorated															
H/S should be															
created toward															
refund deposit															

Orig. Move-in Date 8-28 d same Exp. X # 1

BLDG. # NAME Kendall 1st

Quantitative Docs 3: Data Capture Forms

- Purpose: Understand the *input* side of the information flow
- What to Collect
  - All forms: Official and unofficial (“bootleg”), paper and electronic (web forms, fillable PDFs)
  - Blank copies + instructions
  - Examples of filled-in forms (use sampling!)
- Cataloging Steps
  - Collect: Gather examples of every form
  - Note Type: How is it produced/used (printed, handwritten, web, etc.)?
  - Document Distribution: Who is supposed to get it?
  - Compare: Who actually gets it? (Reality vs. Plan)

Analyzing Data Capture Forms: Official vs. Bootleg

- Key Questions
  - Completeness: Is it fully filled out? Which fields are skipped? Why? (Poor design? Unnecessary data?)
  - Usage: Are some official forms ignored? Why? (Bad design? Redundant?)
  - Circulation: Do forms reach the right people/files? Are online forms accessible? (Workflow/permission issues?)
  - Paper vs. Online: Compare completion rates if both exists (Usability differences?)
  - Bootleg Forms: Why do they exist?
    - Simplify a complex official form?
    - Address missing needs?
    - Indicate process problems or workarounds?
    - Sign of political issues? (Control over data input)

Analyzing Qualitative Documents: Reading Between the Lines

- What: Written materials revealing expectations, beliefs, values, and human interaction aspects $\rightarrow$ Emails, Memos, Signs, Web Pages, Manuals, Policy Handbooks
- Why
  - Rich source of detail on organizational culture
  - Reveal user expectations and attitudes towards technology
  - Uncover the 'affective, emotional, and motivational aspects' of HCI
- Scared? $\rightarrow$ Learn the systematic approach to decode these documents

Decoding Qualitative Docs: Systematic Guidelines

- Focus: Uncovering underlying meanings, values, and relationships
- Guidelines
  - Metaphors: Look for key or guiding metaphors (e.g., “We’re a family,” “It’s a machine,” “This is war”) $\rightarrow$ Language shapes behavior
  - Insiders vs. Outsiders: Identify an “us vs. them” mentality (e.g., department rivalries, management vs. staff)
  - “Good” vs “Evil”: Note terms used repeatedly to describe positive/negative aspects (e.g., “efficient,” “innovative” vs. “bureaucratic,” “legacy”)
  - Messages/Graphics: Analyze visuals and text on websites, bulletin boards $\rightarrow$ What themes are emphasized?
  - Humor: Recognize its presence and type $\rightarrow$ Indicates subcultures, morale, attitudes

Qualitative Docs: Memos and Signs

- Memos
  - Analyze content using the 5 guidelines
  - Reveals ongoing dialogues values, attitudes, beliefs
  - Note sender/receiver patterns (Who talks to whom?) $\rightarrow$ Often downward/horizontal flow
- Signs and Posters
  - Subtle reinforcers of official culture/values
  - Examples: “Quality is Job One,” “Safety First,” “Think Customer”
  - Give clues about the organization’s stated priorities

Memo

To: All Night Call Desk Staff
From: S. Leep, Night Manager
Date: 2/15/2023
Re: Get Acquainted Party Tonight

It’s a pleasure to welcome two new 11-7 Call Desk staff members, Twyla Tine and Al Knight. I’m sure they’ll enjoy working here. Being together in the wee hours makes us feel like one big happy family. Remember for your breaks tonight that some of the crew has brought in food. Help yourself to the spread you find in the break room, and welcome to the clan, Twyla and Al.

Qualitative Docs: Corporate Websites

- Scope: Examine B2C and B2B sites
- Analysis Dimensions
  - Content
    - Apply 5 guidelines
    - Clarity of message?
  - Design: Use of color, graphics, animation, layout
  - Technical: Interactivity? Accessibility? Security? Customization/Personalization options?
  - Managerial
    - Discrepancies between site and stated goals?
    - Does it reflect reality?
- Relevance
  - Even if not redesigning the site, it impacts the overall system view
  - Shows the company's external face
  - Ask for usage metrics if relevant

Qualitative Docs: Manuals and Policy Handbooks

- Manuals
  - Describe the **ideal** way things *should* be done
  - Apply the guidelines to language and tone
  - Reality Check
    - Are printed manuals up-to-date?
    - Are they actually used?
    - Compare ideal procedure to actual practice observed elsewhere
- Policy Handbooks
  - Cover broad employee/corporate conduct
  - Focus: Policies related to IT - computer use, access, security, data handling, service charges
  - Reveal guiding corporate values, attitudes, and beliefs regarding technology and information

Using Text Analytics: Making Sense of Unstructured Data

- Challenge: Organizations are flooded with Unstructured Qualitative Data
  - Examples: Emails, interview transcripts, reports, wikis, blogs, chat logs, social media posts (Facebook, TikTok, $\mathbb{X}$, etc.), customer reviews
- What: Software that analyzes this “soft” data to find patterns, themes, and insights
- Goal: Understand what customers, employees, and vendors are thinking, feeling, and saying $\rightarrow$ Discover motivations and values

The Power Of Text Analytics

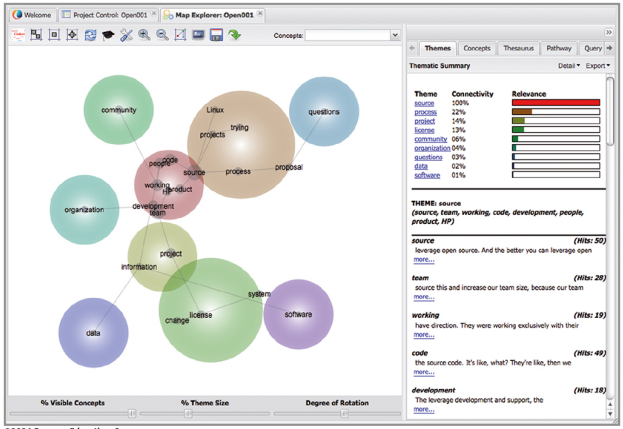
- Benefits
  - Gain valuable insights into customer/vendor/employee perspectives
  - Understand motivations, values, and opinions expressed freely
  - Identify emerging trends or issues mentioned in open text
- Contrast: Unstructured vs. Structured Data
  - Text Analytics: Analyzes Unstructured Qualitative Data (text, conversations) $\rightarrow$ Provides qualitative insights
  - Data Analytics: Analyzes Structured Quantitative Data $\rightarrow$ Uses queries/algorithms for quantitative insights
- Relevance: Crucial for companies with significant online presence or lots of text-based communication

Under the Hood: Text Analytics Basics

- General Process
  - Input: User submits documents (transcripts, reports, web data)
  - Preprocessing: User defines “stop words” (e.g., ‘a’, ‘the’, ‘is’) to ignore
  - Analysis: Software performs keyword counts, identifies co-occuring words $\rightarrow$ Concepts
  - Output: Generates visualizations and summaries
- Leximancer (Example Tool)
  - Focuses on rapid, visual, qualitative analysis
  - Identifies Concepts: Words that frequently appear together, forming themes
  - Creates a “thesaurus” of related terms based on context
  - Generates visual outputs like concept maps and ranked lists

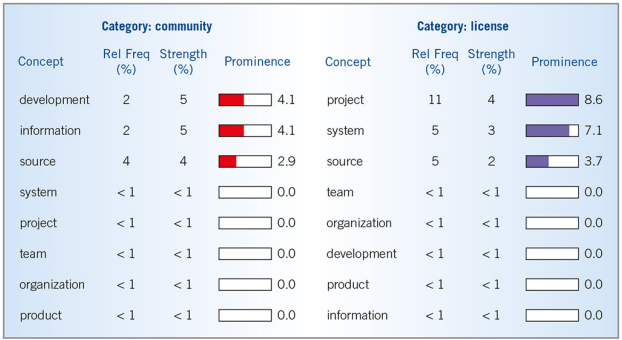
Making Text Visual: Leximancer Outputs

- Concept Maps
  - Show relationships *between* concepts visually.
  - Proximity and connecting lines indicate strength of relationship.
  - Size of concept ‘bubbles’ can indicate frequency/importance.



Making Text Visual: Leximancer Outputs

- Ranked Concept Lists
  - Show the most prominent concepts within specific categories or overall.
  - Often displayed as bar charts showing frequency or relevance scores.



Category: community				Category: license			
Concept	Rel Freq (%)	Strength (%)	Prominence	Concept	Rel Freq (%)	Strength (%)	Prominence
development	2	5	4.1	project	11	4	8.6
information	2	5	4.1	system	5	3	7.1
source	4	4	2.9	source	5	2	3.7
system	< 1	< 1	0.0	team	< 1	< 1	0.0
project	< 1	< 1	0.0	organization	< 1	< 1	0.0
team	< 1	< 1	0.0	development	< 1	< 1	0.0
organization	< 1	< 1	0.0	product	< 1	< 1	0.0
product	< 1	< 1	0.0	information	< 1	< 1	0.0

Text Analytics: More Than Just a Tool

- Your Role
  - Demonstrate Value: Show the organization how analyzing unstructured data yields actionable insights and predicts trends
  - Design Human Processes: Implementing the tool isn't enough! Design *how* people will use it, interpret results, and integrate insights into decisions
  - Guide Interpretation: Help users understand the *qualitative meaning* behind the visuals. It's NOT just about converting text to numbers.
  - Facilitate Action: Help translate insights from text analytics into well-reasoned recommendations and organizational improvements
- Key: Develop systems that *include* text analytics as a useful means to a qualitative end

Observing Decision Makers: Beyond the Interview

- What: Another *unobtrusive* technique for systems analysis.
- Purpose
  - Understand what decision makers *actually do*, not just what's documented or said
  - Observe interactions with their physical/ergonomic environment
  - See relationships with other organizational members firsthand
  - Assess Human-Computer Interaction (HCI) - how well systems fit the user

Capturing the Dynamic Nature of Decision Making

- Challenge
  - Managerial work is often fragmented - “a series of interruptions punctuated by short bursts of work”
  - Difficult to capture fully via interviews
- Value: Allows the analyst to see firsthand:
  - How managers gather information
  - How they process it
  - How they share it with others
  - How they use information and technology to accomplish tasks

A Humanistic Approach: The Playscript

- Concept: Describe observed actions like a script for a play
- Structure
  - Left Column: Actor $\rightarrow$ The decision maker being observed
  - Right Column: Script $\rightarrow$ Their actions, described using action verbs
- Example Verbs: Asks, Prints, Discusses, Reads, Compares, Inputs, Observes, Calls, Drafts, Sends, Rewrites
- Value
  - Provides an organized, systematic record of observed activities
  - Helps determine information requirements for decisions
  - Reveals dependencies
  - Shows information sources

Demo: Playscript Example

Observing the Physical Environment: Clues in the Workspace

- Concept: Decision makers influence, and are influenced by, their physical work environment (office)
- Purpose: Gain insights into human information requirements by observing the workspace
- Relationship to Other Methods: Observing complements interviews, questionnaires by confirming or contradicting the gathered narrative

A Framework for Observation: STROBE

- Analogy: Like a ‘mise-en-scène’ analysis in film criticism
  - Systematically analyzing elements within the frame to understand meaning
  - Can the environment confirm or contract the ‘dialogue’ (interview/survey/data)?
  - The ‘Room Rater’ phenomenon!
- Method: **STR**uctured **OB**servation of the **E**nvironment
- Goal: Systematically examine concrete elements of the office environment

STROBE: Key Observable Elements (1-3)

- Office Location
  - Who gets the corner office?
  - Are key decision-makers centrally located or dispersed?
  - Is it easily accessible?
- Desk Placement
  - Face the door? $\rightarrow$ Power
  - Face away? $\rightarrow$ Openness
  - Encourage interaction or create a barrier?
- Stationary Equipment
  - Large filing cabinets? $\rightarrow$ Personal storage
  - Minimal storage? $\rightarrow$ Reliance on digital/shared
  - What kind of equipment is present?

STROBE: Key Observable Elements (4-7)

- Props
  - Evidence of PC, smartphone, tablet use?
  - Personal items?
  - Tools fo the trade?
- External Information Sources: Presence of trade journals, newspapers, books? $\rightarrow$ Indicates seeking information beyond internal sources
- Office Lighting and Color
  - Bright, fluorescent? $\rightarrow$ Task-oriented
  - Warmer, incandescent? $\rightarrow$ Casual
  - Color usage - institutional or personalized?
  - Inviting or sterile?
- Clothing Worn
  - Formal suits? Business casual? Uniforms?
  - Reflects organizational culture and individual's positioning

Practical Application: STROBE Checklist

- Method $\rightarrow$ Anecdotal Checklist
  - Identify key themes/narratives from interviews/documents
  - Systematically observe the environment using the 7 STROBE elements
  - Compare observations to the narratives
  - Use shorthand symbols on a checklist to record the comparison
- Goal: Structured Analysis of how observations confirm, contradict, modify, or supplement the existing organizational narrative

STROBE Checklist Symbols

- Symbol Key
  - Tick: Confirms the narrative
  - Cross: Negates the narrative
  - Oval/Eye: Cue to investigate further
  - Square: Modifies the narrative
  - Circle: Supplements the narrative
- Example Use: Checklist compares narratives against observations in different STROBE categories

Demo: STROBE Example