



On the Information Overload Problem

Lilian Xie Mentor: Dr. Byron Gao

Wellesley College, Texas State University



ABSTRACT

There remains no universal, cross-disciplinary definition of **information overload** (IO). Colloquially, we understand information overload to be simply what occurs when there is **too much information**.

In computer science, IO is in no way an unprecedented phenomenon, but demands to be reexamined in the modern context of big data. We aim to **perform a review of the preexisting literature** on information overload in the fields of cognitive science, psychology, information science, and organizational science. We seek to **define information overload in the context of computer science** and **explore a case study of descriptive clustering**, a potential IO solution.

LITERATURE REVIEW

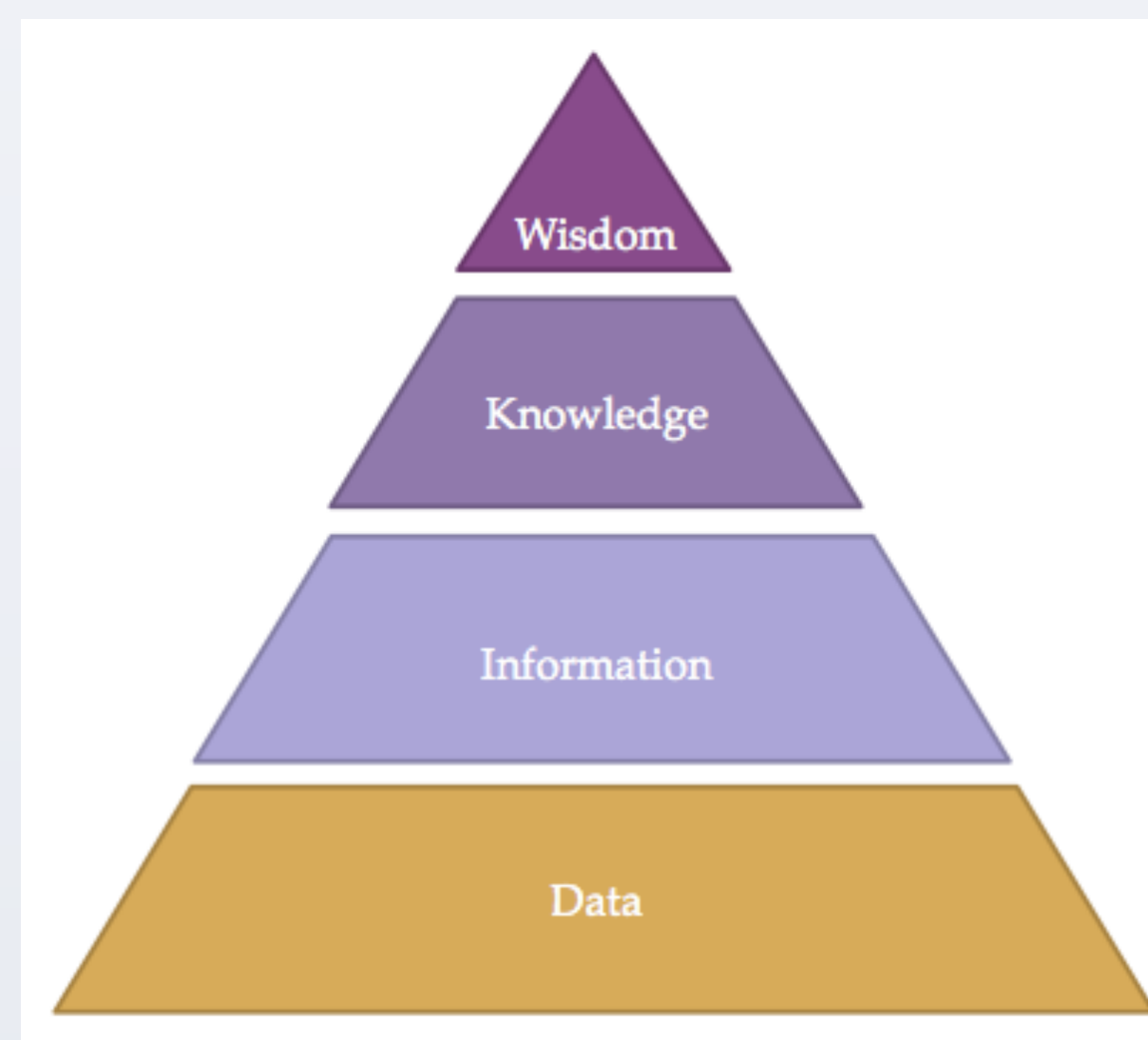
We performed a literature review from a number of different disciplines. From that survey, we have condensed the results into a table categorizing the various causes of information overload.

Information Quantity	Amount of unprocessed information	Bawden, 2001; Herbig and Kramer, 1994; Jacoby et al., 1974; Jacoby 1977, 1984; Malhotra, 1982, as cited in Eppler and Mengis, 2004
	- Scale and acceleration of new information	Ho and Tang, 2001
Information Quality	- Ambiguity of information - Novelty of information - Complexity of information - Intensity of information - Overabundance of irrelevant information, i.e. data smog - Signal-to-noise dynamics - Information uncertainty - Information format and display	Schneider, 1987 Schneider, 1987 Schneider, 1987 Schneider, 1987 Tushman and Nadler, 1975; Ho and Tang, 2001 Edmunds and Morris, 2000 Tushman and Nadler, 1975 Eppler and Mengis, 2004
Environment	- Push technologies, such as email or IM - Notification systems - Speed and ease of access - Interruptions during complex tasks - Pressured work environments	Bawden, 2001 Edmunds and Morris, 2000 Schultze and Vandenbosch, 1998 Speier et al., 1999 Schick et al., 1990
Human Ability	- Cognitive processing bandwidth - Motivation, attitude, and satisfaction - Mood and personal factors, such as level of sleep - Experience with information processing	Miller, 1956 Owen, 1992; Hiltz and Turoff, 1985; Muller, 1984; Schneider, 1987; Swain and Haka, 2000; as cited in Eppler and Mengis, 2004 Owen, 1992 Eppler and Mengis, 2004

DEFINITIONS

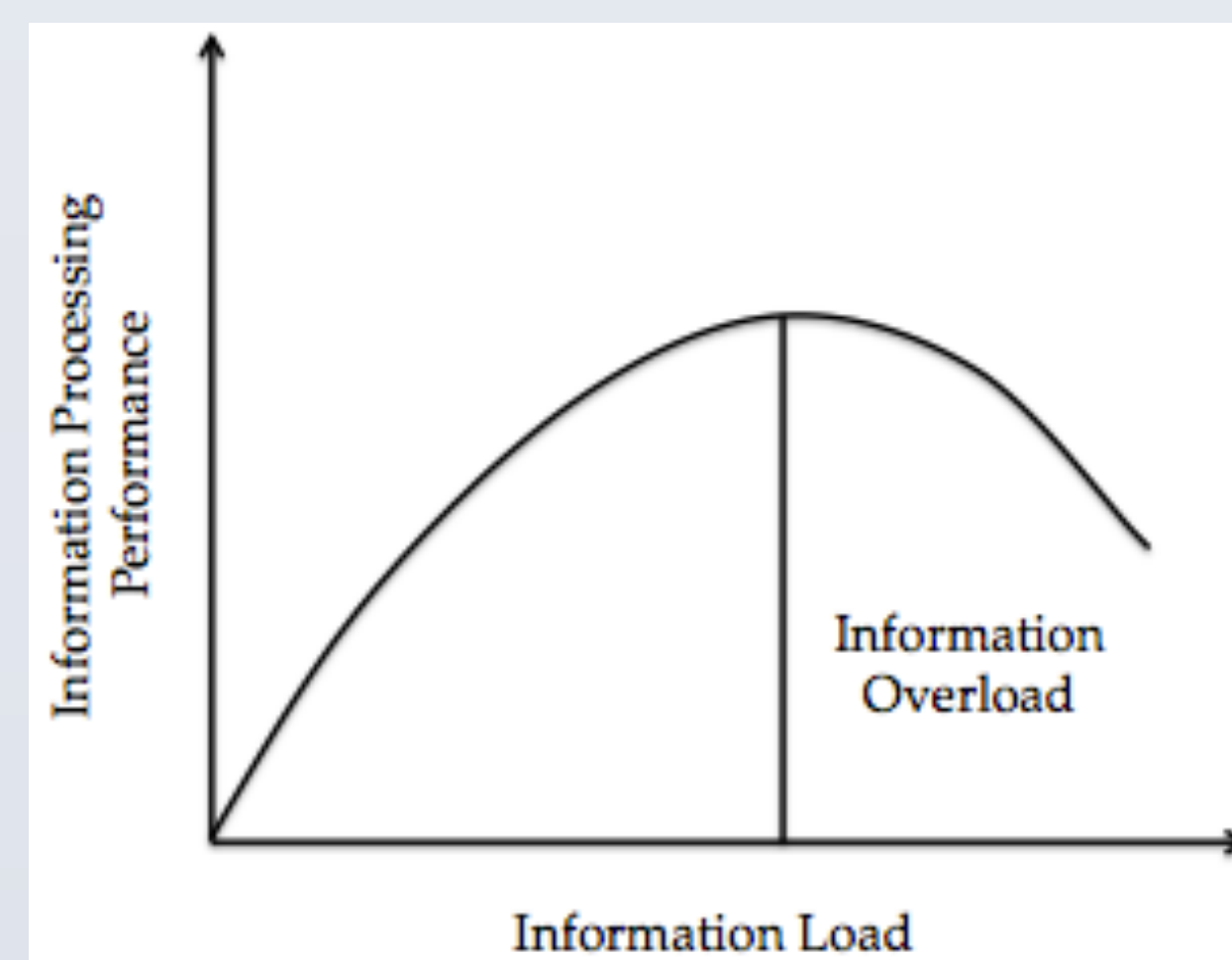
“Information”

In the domain of technology, we rely on the **DIKW hierarchy**, borrowed from the field of information science, to define and differentiate between data, information, wisdom, and knowledge. Based on this hierarchy, we define information as **data that is interpreted and packaged to be communicated**.

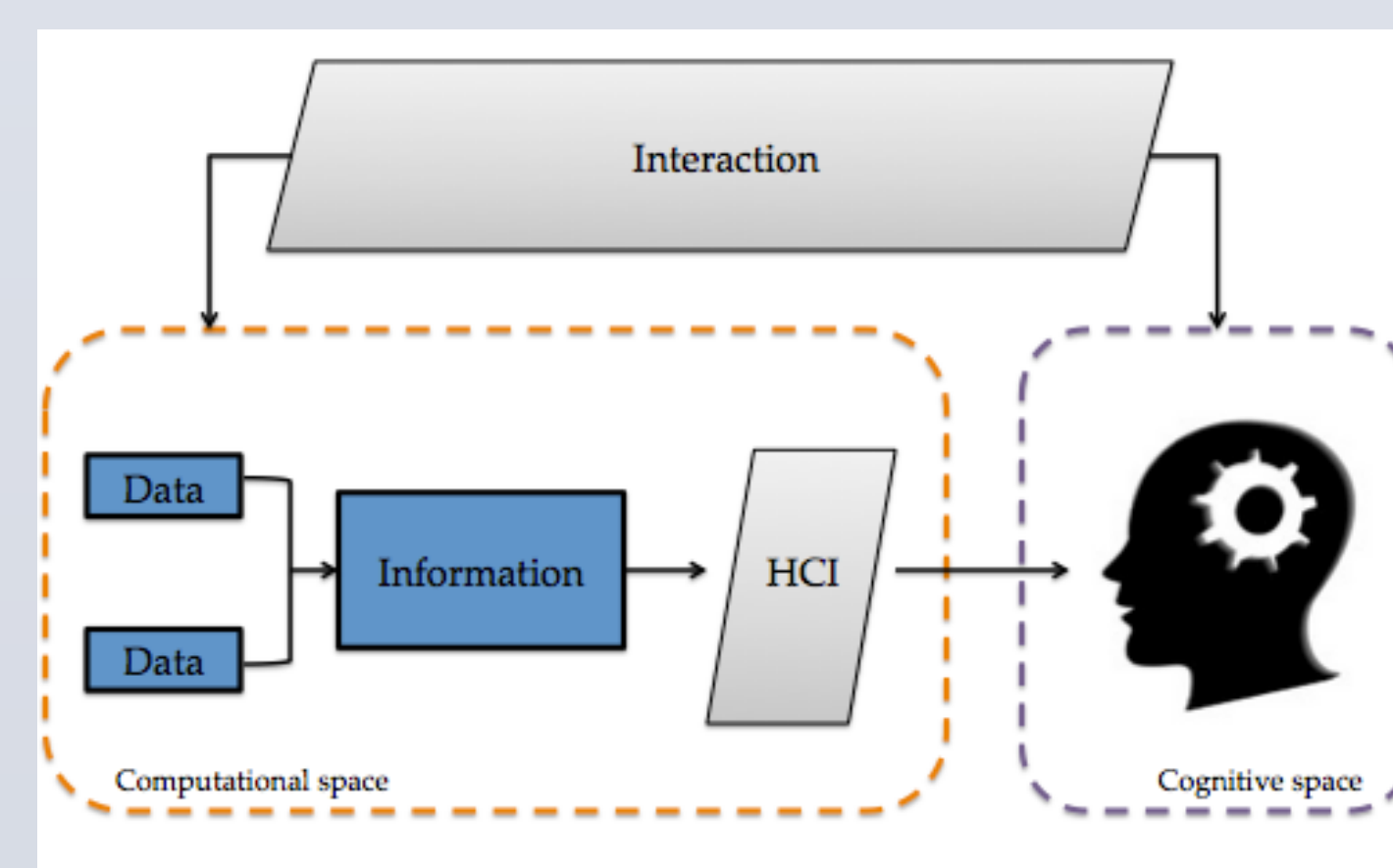


“Overload”

Historically, scholars have represented information overload schematically as a **function of information load and processing performance**, often represented by the inverted U-curve.



In the context of computer science, we found that IO is less a function of information load than *perceived* information load. Thus, information overload is affected by the delivery of information in the **computational space** as well as the processing of information in the **cognitive space**.



CASE STUDY

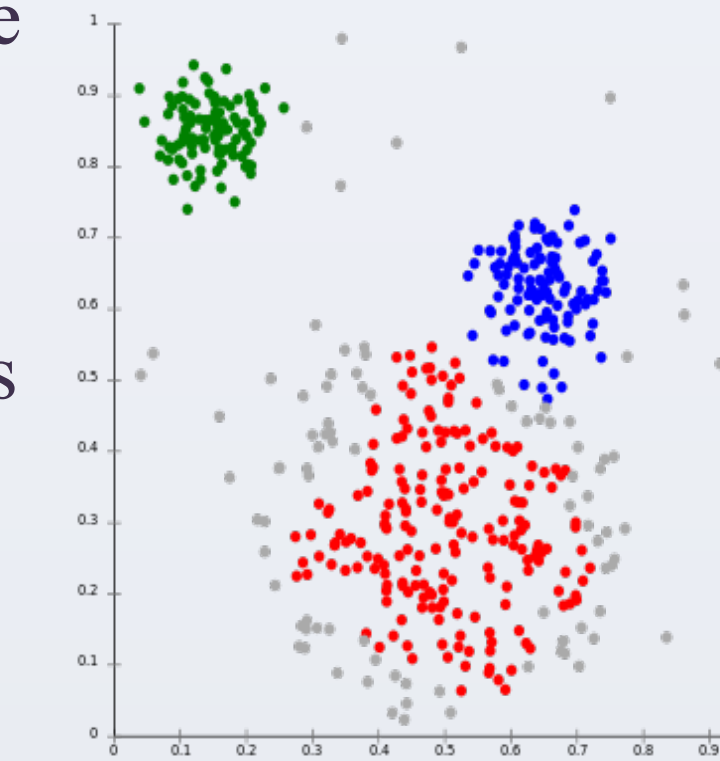
Descriptive Clustering

Descriptive clustering combines information retrieval, text mining, and traditional clustering to create clusters described with **meaningful, comprehensible, and compact text labels**.

Research Problem

How do we give **semantic meaning** to **mathematically optimized** data clusters in order to minimize information overload? How should we change our approach to clustering such that clustering:

- Creates sensible cluster labels
- Explains patterns
- Summarizes data
- Identifies topics



Methods

Frequency-based offline document clustering

During the past weeks Europe and America witnessed t of important international conferences, which coul hand guarantee the peace and stability of the world, lead deteriorate the conditions. The conference in M h dealt with a ban on demand for light arms, and the ence which had a much wider and extended agenda, had atures between them. The first was that large and n s of dissidents were opposing the wishes and interes but powerful group. The second feature was that bot ended without effective result, and the decisions t lutions adopted there lacked conclusiveness and were nal. The Bonn Conference however, which dealt with t

- MPQA Opinion Corpus (2002) of ~400 news articles

“LOCATION”: [“New York”, “New York Conference”, “The”, “conferences”, “them”, “NP”: [“America”, “New York”, “TI onference”, “The conference”, “The first”, “The second”, “a ban”, “a much wider”, “a much wider and extended ag relative success”, “a small”, “a small but powerful gr enda”, “an exception”, “both conferences”, “conclusiven mand”, “deteriorate”, “discussion”, “discussion and exci dissidents”, “effective result”, “exchange”, “important ional conferences”, “interests”, “it”, “large”, “large i ous bodies”, “light arms”, “numerous bodies”, “powerful “resolutions”, “stability”, “the Genoa Conference”, “the k Conference”, “the arrangements”, “the condition”, “the ns”, “the decisions”, “the global environment”, “the on

[["America", "00.16.28-8800"], ["Europe", "00.16.28-8800"], ["New York", "00.16.28-8800"], ["New York Co nference", "00.16.28-8800"], ["America", "00.16.28-8 800"], ["New York", "00.16.28-8800"], ["New York Con ference", "00.16.28-8800"], ["arms", "00.16.28-8800"], ["conferences", "00.16.28-8800"], ["America", "00 .16.28-8800"], ["New York", "00.16.28-8800"], ["Bonn Conference", "00.16.28-8800"], ["conference", "00.1 6.28-8800"], ["first", "00.16.28-8800"], ["second fe ature", "00.16.28-8800"], ["ban", "00.16.28-8800"], ["much wider", "00.16.28-8800"], ["much wider and extended agenda", "00.16.28-8800"], ["relative success

- Extract relevant noun phrases with Stanford Named Entity Recognizer (NER)
- Implement a greedy algorithm for set covering
- Each element represents a labeled cluster

Results

~40 descriptive clusters

[“UNITED STATES”, “CHINA”, “AFGHANISTAN”, “ISRAEL”, “RUSSIA”, “WASHINGTON”, “ZIMBABWE”, “ARGENTINA”, “MARCH”, “PRESIDENT”, “PEOPLE”, “NOVEMBER”, “POLICE”, “BETHLEHEM”, “KAZAKHSTAN”, “DEC”, “ENGLISH”, “GUANTANAMO”, “S”, “JAKARTA”, “NUMBER”, “JAMMU”, “SAUDI ARABIA”, “ASIA”, “AUGUST”, “UZBEKISTAN”, “TASS”, “SALO NICA”, “BANGUI”, “SOCIETY”, “EUROPE”, “STOIBER”, “HAJJI ISA SE BI”, “PALESTINIANS”, “SOUTH KOREA”, “PRIME MINISTER”, “COLOMBO”, “TURKMENISTAN”, “BC”, “CHINESE”, “COMPUTERS”, “ONE TRADER”, “MR_SUNIL MITTAL”, “MONDAY JANUARY”, “REPORT”]

CONCLUSION

Based on our research and the preexisting literature, we define information overload in the context of computer science as a **phenomenon that occurs when the sensation of information burden results in the inability to optimally process information**.



REFERENCES

Eppler M. and Mengis J. 1969. The Concept of Information Overload: A Review of Literature from Organizational Science, Accounting, Marketing, MIS, and Related Disciplines. The Information Society: An International Journal. 20(5). pp325-344

Miller G. 1956. The Magical Number Seven, Plus or Minus Two: Some Limits On Our Capacity for Processing Information. The Psychological Review. Vol. 63. pp81-97.

Misra S. and Stokols D. 2011. Psychological and Health Outcomes of Information Overload. Environment and Behavior. 44(6). pp737-759.

ACKNOWLEDGEMENTS

This research is funded by NSF REU award #1358939