Synthesizing the Gaseous State: Mapping the Geographic Convergence of Knowledge

ALEJANDRO MIESES
Pontifical Catholic University of Puerto Rico

Synthesizing the Gaseous State is an ongoing research effort that uses neural networks to map word relations, hierarchies, gaps, and foci of terms within large data sets of architecture and urban theory, as they are graphed geographically by epoch. It chooses to extend the usual application of algorithms from structural, physical, spatial, or mathematical purposes, to include theoretical concerns. The methodology employed in order to achieve this relies on each term stored according to its contextual relations (words directly surrounding the term in a sentence) that are referenced according to their place in multiple texts and occurrences within the same text. While this context is dynamic, as dependent upon the weighted relations traced through texts loaded upon the data set, it is balanced against a static or fixed context - the definition of these words, their author, place, and date. Relying upon the information acquired, each term behaves as an active agent, seeking out relations based upon new contexts and visually graphing them according to place, author, and date, as opposed to abstract or empty space. This comes with the purpose of visually revealing the movement of architectural thought through the terrain, while describing how it converges and diverges from certain concerns locally, as compared to global currents across time, as a Computer Aided Epistemology of architecture theory.

The fitness criteria used while evaluating the optimization of the software is its capacity to graphically seek out a determined proximity measure between associated terms, while still being constrained to a certain latitude-longitude range in position, associated with the author's affiliated institutions.

Currently the software uses a 1GB database of architecture theory taken from assorted periodicals (Architectural Design Magazine and Architectural Record), for proof-of-concept purposes. Next steps will include collaborations with Charles University, Prague, and the ICCIT University of Toronto, to synthesize 20+ years of PhD theses, in order to acquire a knowledge map of most related and commonly used terms, most discussed topics, authors cited by hierarchy, sites of intervention and least discussed subjects. The software allows any language to be entered and analyzed.

Applications for this can span various scenarios, such as: a) the need for aid in traumatized regions due to war and natural disasters where, acquiring/organizing information from the ground, can give a clearer and more rapid picture of the necessities of the people affected, as these are instantaneously built into a database b) an interface for exploring our collective knowledge, the most pressing stimuli upon our profession and our obscured topics as they are hierarchized through time, c) a synthesis of the diversity of public opinions upon the built/urban environment without sacrificing variables, or channeling opinions through surveys, d) ideological or economic movements through the terrain, such as protests, gentrification and educational inequalities, in data-substantiated ways, that can give further insight into other statistical analysis, and e) a dynamic graph that can connect, the user's interest, with particular places, authors, or historical facts, of the subjects he or she has queried.

Synthesizing the Gaseous State

Mapping the Geographic Convergence of Knowledge

Synthesizing the Gaseous State is an ongoing research effort that uses neural networks to map word relations, hierarchies, gaps, and foci of terms within large data sets of narrative information (currently architecture and urban theory), as they are graphed geographically by epoch. It chooses to expand the usual application of psychohistory from structural, physical, spatial, or mathematical purposes, to include theoretical concerns, with the purpose of visually tracing how knowledge builds, accretes and separates through maps and in abstract space. The methodology employed in order to achieve this relies on each term stored according to its contextual relations (words directly surrounding the term in a sentence) that are referenced according to their place in multiple texts and occurrences within the same text. While this context is dynamic, - dependent upon the weighted relations traced through texts loaded upon the data set - it is balanced against a static or fixed context - the definition of these words, their author, place, and date. Relying upon the information acquired, each term behaves as an active agent, seeking out relations based upon new contexts and visually graphing them according to place, author, and date, as opposed to abstract or empty space. This comes with the purpose of visually revealing the movement of architectural thought through the terrain, while describing how it converges and diverges from certain concerns locally, as compared to global currents across time, as a Computer Aided Epistemology of architecture theory. The aims of this research are twofold: a) to derive a working set of algorithms capable of tracing existing uses and relations across large data sets of theory, while b) creating a visual interface for users to explore, generate, and type in terms that demonstrate their usage and unforeseen combinations through the existing bodies of text scanned.

The database contains over 20,000 terms, directly gathered from text scanned, with over 900,000 citations. The software automatically scans the internet (dictionary.com) to gather the definitions of words, synonyms and antonyms. Each term is stored with all the citations which reference such term. Abstract or Word-Proximity Maps are built by in two phases: first a branching algorithm triangulates all terms related to the initial user-inputed (search query), for more than 1500 generations (iterations) this generates a primary lattice of proximities between directly related terms, after this initial phase, secondary terms are introduced by proximity to primary and are made to find balance by dynamically seeking out other related terms. The fitness criteria used while evaluating the optimization of the software is its capacity to graphically seek out a determined proximity measure between associated terms, while still being constrained to a certain georeferenced coordinate if it is viewed in terrain mode.

NEXT STEPS

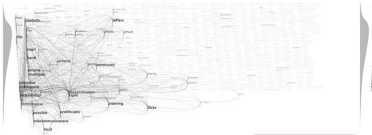
Currently the software uses a 1GB database of architecture theory taken from assorted periodicals (Architectural Design Magazine and Architectural Record), for proof-of-concept purposes. Next steps will include collaborations with Charles University, Prague, Department of Arts, and the Institute for Communication, Culture, Information and Technology, to synthesize 20+ years of PhD theses, in order to acquire a knowledge map of most related and commonly used terms, most discussed topics, authors cited by hierarchy, sites of intervention and least discussed subjects. The software allows any language to be entered and analyzed. The maps generated work in abstract space, as search queries, which give words by triangulated proximities to the ones entered, in addition to its geo-referenced display, which is still under development.

APPLICATIONS

Applications for this research can span various scenarios, such as: a) the need for aid in traumatized regions due to war and natural disasters where, acquiring/organizing information from the ground, can give a clearer and more rapid picture of the necessities of the people affected, b) an interface for exploring our collective knowledge, the most pressing stimuli upon our profession and our obscured topics as they are hierarchized through time, c) a synthesis of the diversity of public opinions upon the built/urban environment without sacrificing variables, or channeling opinions through surveys, d) ideological or economic movements through the terrain, such as protests, gentrification and educational inequalities, in data-substantiated ways, that can give further insight into other statistical analysis, and e) a dynamic graph that can connect, the user's interest, with particular nuclei - places, authors, or historical facts, of the subjects he or she has queried.

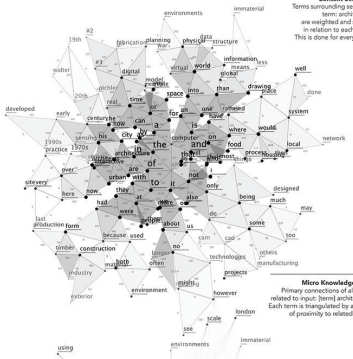
Zoomed View Knowledge Map
input: [term] computer.

Context Selection
Terms surrounding selected term: architecture are weighted and stored in relation to each other. This is done for every word.



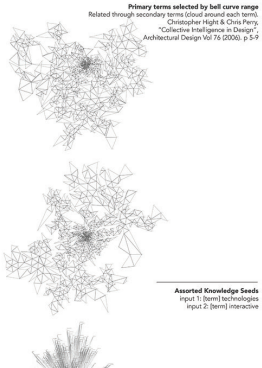
Database Visualization of Neural Network
Connections within single essay: Christopher Hight & Chris Perry, "Collective Intelligence in Design", Architectural Design Vol 76 (2006), p. 5-9.

Primary terms selected by bell curve range
Related through secondary terms (cloud around each term). Christopher Hight & Chris Perry, "Collective Intelligence in Design", Architectural Design Vol 76 (2006), p. 5-9

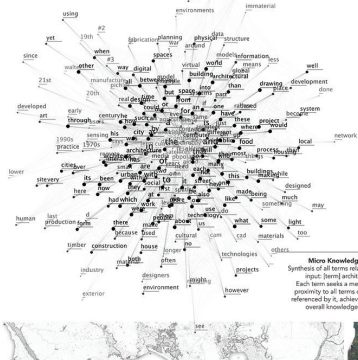


Micro Knowledge Seed
Primary connections of all terms related to input: [term] architecture. Each term is triangulated by a value of proximity to related terms.

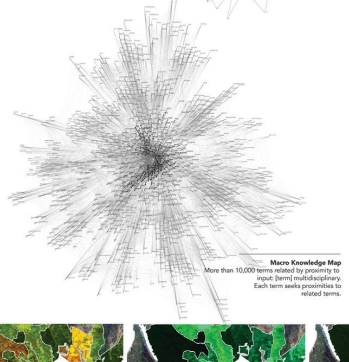
Macro Knowledge Seed
900 iterations, converging and diverging terms triangulated by proximity to: [term] architecture. 81,000 terms in total.



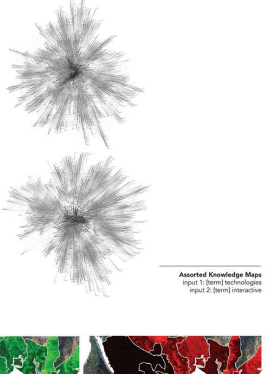
Assorted Knowledge Seeds
input 1: [term] technologies
input 2: [term] interactive



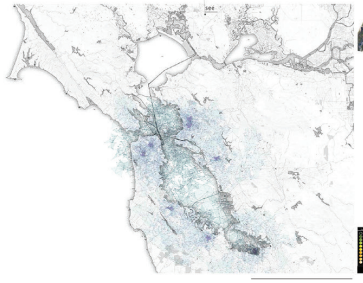
Micro Knowledge Map
Synthesis of all terms related to input: [term] architecture. Each term seeks a measured proximity to all terms directly referenced by it, achieving an overall knowledge cloud.



Macro Knowledge Map
More than 10,000 terms related by proximity to input: [term] multidisciplinary. Each term seeks proximities to related terms.



Assorted Knowledge Maps
input 1: [term] technologies
input 2: [term] interactive



Knowledge Map on Terrain: SF Bay
Localized provisionally on primary universities: UC Berkeley, UC San Francisco, University of San Francisco, UC Santa Monica & Stanford. This is in proof-of-concept stage.

Production of Knowledge by County • General
Gradient responds to words used by authors affiliated to institutions by county. Each color pertains to individual professions.

Production of Knowledge • Science & Technology
Terms related to science are green hued, technology are blue.

Production of Knowledge • Languages & History
Terms related to languages are yellow hued, history are green.

Production of Knowledge • Natural & Social Sciences
Terms related to natural sciences are red hued, social sciences are orange.

Database Visualization
Over 20,000 terms displayed, 900,000 citations, and primary of connections of most related/popular terms