A Correlation Technology Demonstration

Q: What is Correlation Technology?

A: Correlation Technology seeks to emulate the way the human brain acquires, stores and utilizes information.
Please note:

The Correlation Technology Platform (CTP) is not an end user application. The CTP is intended to support deployment of vertical market specific software applications which incorporate Correlation Technology. Some examples of vertical markets which have been identified as candidates for Correlation Technology solutions include Legal E-Discovery, Recruitment, and Market Research Qualitative Survey Analysis. Deployment of the CTP in each vertical market requires specialization and extension of the Platform. Specializations include customized analysis (called Refinement) of the Correlation result set (called the Answer Space), and customized GUI creation appropriate to the specific needs of each application.

The GUI and method of results presentation which follow will almost certainly never be used by any real-world application. The sole purpose of this presentation is provide a look “under the hood” of Correlation Technology, giving you a first introduction to this new, patented method of addressing data.
The first step is to ask the system if the terms of interest can actually be found in the “corpus” or collection of documents which make up the Correlation “domain”.

For this demonstration, the “corpus” is the entire contents of Wikipedia.
Wikipedia has about 3.5 M documents

Correlation Technology creates “Knowledge Fragments” by decomposing the documents in the corpus using a patented method. Our method does not rely on syntactical sentence decomposition, and Knowledge Fragments are not phrases.

We exclude documents with no data, or that are only used by Wiki editors.
An N-Dimensional Query asks the question, “Identify all the connections from an Origin term or concept to a Destination term or concept.” The Origin and Destination must represent a lexical and/or semantic disjunction. Although Origin and/or Destination can be “entities” (people, places, organizations, or things), for a standard English language expository corpus such as Wikipedia, other types of terms or concepts yield more interesting results. With an average article length of 724 words, Wikipedia is a “broad but shallow” corpus, meaning that many topics are represented, but not much content is present about the average topic. Like all information retrieval systems, Correlation Technology can not find connections from Origins to Destinations if no data about the Origins or Destinations is present in the corpus. Correlation Technology can not manufacture what is not there.
“Knowledge Fragments” which match or are related to the entered Origins are found. The "Rank" represents the number of unique Knowledge Fragments with the same base value. Only “checked” values will be used in the actual Correlation.

“Knowledge Fragments” which match or are related to the entered Destinations are found.

Although “Knowledge Fragments” which match or are related to the entered Contexts are found, none have been chosen to be used in this particular Correlation.

Note: “Contexts” are not binary filters. They expand rather than limit the answers returned, and "guide" the Correlation process.
These are the terms or concepts to be correlated using the Knowledge Fragments found in the decomposed Wikipedia corpus.

These are some governors used to manage the Correlation process. These governors are present for demonstration purposes only. In actual Correlation applications, the systems will typically be provisioned to run to exhaustion.
These messages show that thousands of correlations have been constructed connecting Origins to Destinations.
Correlation Technology results are not achieved by using a reasoner iterating over a graph. There is no graph. Every single successful correlation starts as a trial. Starting with each Origin Knowledge Fragment, all Knowledge Fragments in the Correlation Technology data store are examined for those that can be associatively linked – first to the Origin, and then to each iteratively linked Knowledge Fragment until either a Destination Knowledge Fragment is found, or no Knowledge Fragment can be associated. In the 30 seconds elapsed time permitted for this correlation, 2.4 billion trials were performed against the data store of 278 million Knowledge Fragments. Only about 70 thousand correlations were successful. These results would require the equivalent to several lifetimes of human effort.
In order to demonstrate some aspects of Correlation Technology, three views of the result set of correlations will be presented. The first is a simple “tree” hierarchical representation of the Answer Space of correlations. Once again, please note that this is not suggested as an end user GUI. Rather, this is equivalent to performing a query against a large RDBMS and scrolling through all the rows returned in the result set. There are four Origin Knowledge Fragments for which successful correlations have been constructed. Each “destination” is unique, meaning that at least one correlation has been found (hundreds or thousands of correlations can be constructed from each origin to each unique destination). By expanding the “trees”, some destinations and correlations can be displayed.

BTW: “Zanjan” is a province of Iran, and a center of Islamic scholarship.
For Origin Knowledge Fragment “1”, correlations have been constructed to 897 Destinations. The first 10 Destinations are shown. Correlation counts are shown. All these correlations will start with “average population density in Zanjan” and will end with the given Destination, such as 1.1 “links to terrorism”, or 1.2 “linked to terrorism”. Each correlation is unique, and may be constructed from completely different sets of Knowledge Fragments.
These are correlations showing just three of the thousands of paths linking “average population density in Zanjan” to “terrorism”. Our patents cover dozens of ways to associate Knowledge Fragments when constructing correlations.

Correlations can be suggestions, assertions, or statements of fact. A relation between Origin and Destination is required and proven, but the characterization of that relation, the relevancy of that relation, and the trusted-ness of that relation is dependent upon the actual Correlation Technology based application. The Refinement components of the CTP, which analyze and filter the contents of the Answer Space, are critical to applying business rules, integrating other processes, and achieving the objectives of vertical application specialization.
Although not actually required for “pure” Correlation, each Knowledge Fragment retains a bibliographic reference to the “resource(s)” – such as web pages or documents – from which the Knowledge Fragment was extracted.

When an identical Knowledge Fragment is extracted from more than one resource, all the “contributing resources” are captured.
Some additional Destinations to which successful correlations from Origin “1” have been constructed. Obviously, all 897 Destinations can not be displayed for this presentation.
More Correlation examples. Notice that every possible permutation of pathway from the Origin to the Destination is captured.
For Origin Knowledge Fragment “2”, correlations have been constructed to 58 Destinations. The first 10 Destinations are shown. All these correlations will start with "average population density of people" and will end with the given Destination, such as 2.1 “state to counter terrorism”, or 2.2 “state provide counter terrorism”.
As with the Origin “1”, not all Destinations can be displayed for this presentation.
Correlations linking “average population density of people” to “age of terrorism.”
Notice that *any* resource can contribute Knowledge Fragments to a correlation. Correlation Technology is not a document-centric solution. Each document is subjected to a one-way, exhaustive transformation into a collection of Knowledge Fragments which represent the entire “Knowledge Payload” of the original document. Each resulting Knowledge Fragment is an independent knowledge object, and each is “equally eligible” for Correlation.
For Origin Knowledge Fragment “3”, correlations have been constructed to 34 Destinations. The first 10 Destinations are shown. All these correlations will start with “average population density of area” and will end with the given Destination, such as 3.1 “United States backing state terrorism”, or 3.2 “United States attacked Islamic terrorism”.

Slide 18 - Execution Results Grouped 3rd Origin 1st 10 Destinations
Here is a correlation linking “average population density of area” to “United States fight terrorism”.
Here's the contributing resources for Correlation 3.32.1:

- [Area](http://blade04.8888/articles/dan/14/Denmark.html)
- [Area handed United States](http://blade04.8888/articles/lin/America.html)
- [United States fight terrorism](http://blade04.8888/articles/pla/Pakistan_Army_6511.html)
Here are Correlations linking "population density of inhabitants" to "cities consumed terrorism".
Here's the contributing resources for Correlation 4.58.1:

- [Prisoners have terrorism](http://blade/648888/articles/AlhyF4hGhIne8.html)
- [Prisoners committed terrorism](http://blade/648888/articles/AlhyF4hGhIne8.html)
- [Africa is terrorism](http://blade/648888/articles/AlhyF4hGhIne8.html)
- [Africa paying terrorism](http://blade/648888/articles/AlhyF4hGhIne8.html)
- [Africa for terrorism](http://blade/648888/articles/AlhyF4hGhIne8.html)
- [Cities consumed terrorism](http://blade/648888/articles/AlhyF4hGhIne8.html)

Correlation: 4.58.1

- Population density of inhabitants
- Inhabitants of city
- Cities consumed terrorism
This is a sample of Refinement. By analysis of the Answer Space (in this case a simple frequency distribution of contributing resources), a ranked listing of the contributing resources most significant to this particular N-Dimensional Query can be achieved. This type of listing is generated by the patent-pending Correlation Search Engine (CSE). Because this Correlation execution was not permitted to run to exhaustion, some anomalies show up in the list here. From the contents of the list, however, it should be clear that a properly provisioned full term execution of the NDQ would accurately identify and rank only those Wikipedia articles most significant to the query terms.

From 3.5M Wikipedia documents, only 1,946 are proven to have relevance to the complex N-Dimensional Query relating population density to terrorism.
Here is just another simple way to review the contents of the Answer Space.
The Business of Correlation Technology

- Make Sence, Inc. controls the world-wide licensing of Correlation Technology. Inquiries regarding licensing of Correlation Technology for any vertical market application should be addressed to Carl Wimmer c.wimmer@correlationconcepts.com

- Make Sence Florida, Inc. is the US-based research and development arm of Make Sence, Inc. and is responsible for all technical aspects of the Correlation Technology Platform. Address all technical inquiries to Mark Bobick m.bobick@correlationconcepts.com

For Business Inquiries:
Contact: Carl Wimmer
carl@makesence.us
Mobile: (702) 767-7001

For Technical Inquiries:
Contact: Mark Bobick
m.bobick@correlationconcepts.com
Mobile: (702) 882-5664