

# **Biomimetic Approaches to Discovering Hidden Relationships in Data**

by K. Franks, M. Muldoon, R. Podowski  
SeeqPod, Inc.  
[info@seeqpod.com](mailto:info@seeqpod.com)

## Discovering Relationships

What two words come to mind when you hear the word “sky?”

For the vast majority of people, these words are “blue” and “cloud.”

These words are not synonyms. They are not semantically related. There is no direct way to relate “blue” and “cloud” to “sky.” The relationships can only be found by examining context.

Our system uses a biomimetic approach to discover:

- How objects are related
- How strongly they are related
- In what contexts they are related

## A Biomimetic Approach

The human brain is the best search engine on Earth. The best way to find relationships is to use approaches from the cognitive sciences and emulate how humans organize, search, and recall conceptually related information.

Our system auto-associates objects like a human might and scores their relationships based on proximity, frequency, and uniqueness. This system can be applied to any electronic form of data.

## The Process of Auto-Association

### *System overview*

A large body of electronic data is required as input. Meaningful relationships between objects are found when statistically scored associations are generated over thousands or millions of objects. Smaller collections do not contain enough associations to produce meaningful results.

The system partitions the input into frames of context (Pinker, S., 1997). An auto-association strategy (Xijin Ge, Shuichi Iwata, 2002) is applied to the objects that comprise each frame. Using the resulting data, the system scores how strongly related each two unique objects are. These relationship scores are stored in a collection of vectors, which the system saves as an associative memory module (AMM).

AMM data is represented by a vector space model (Raghavan, V. V., and Wong, S. K. M., 1986). Object similarity is estimated using relationship consensus scores, which are based on correlation coefficients of relevant AMM vector numbers.

## System implementation

### Context frames

Context frames are created by selecting a sequence of objects from the data. One frame is anchored on each object. The number of objects included in the frame is determined by how many lie within the distance threshold from the anchor. Each object must anchor exactly one frame, and may appear as a non-anchor in many frames.

### Relationship strength

Relationship scores between object pairs are based on a distance decay function utilizing the Fibonacci sequence (RA Dunlap, 1997). Decay functions based on exponential ( $e^{-\Delta x}$ ) and inverse distance ( $\Delta x^{-1}$ ) sequences were also considered (Fig. 1).

Given strength, defined as:

$$S_{ij} = \phi^{\Delta x}$$

Where:

- $S_{ij}$  = relationship score between object  $i$  and  $j$ ,
  - $\phi = 0.618034$  is the Golden Ratio component “phi”<sup>†</sup>, and
  - $\Delta x = |x_i - x_j|$  is the relative position between objects  $i$  and  $j$ .
- <sup>†</sup>  $\phi$  is the decimal component of the Golden Ratio  $\Phi = 1.618034$ .

The following decay curves result:

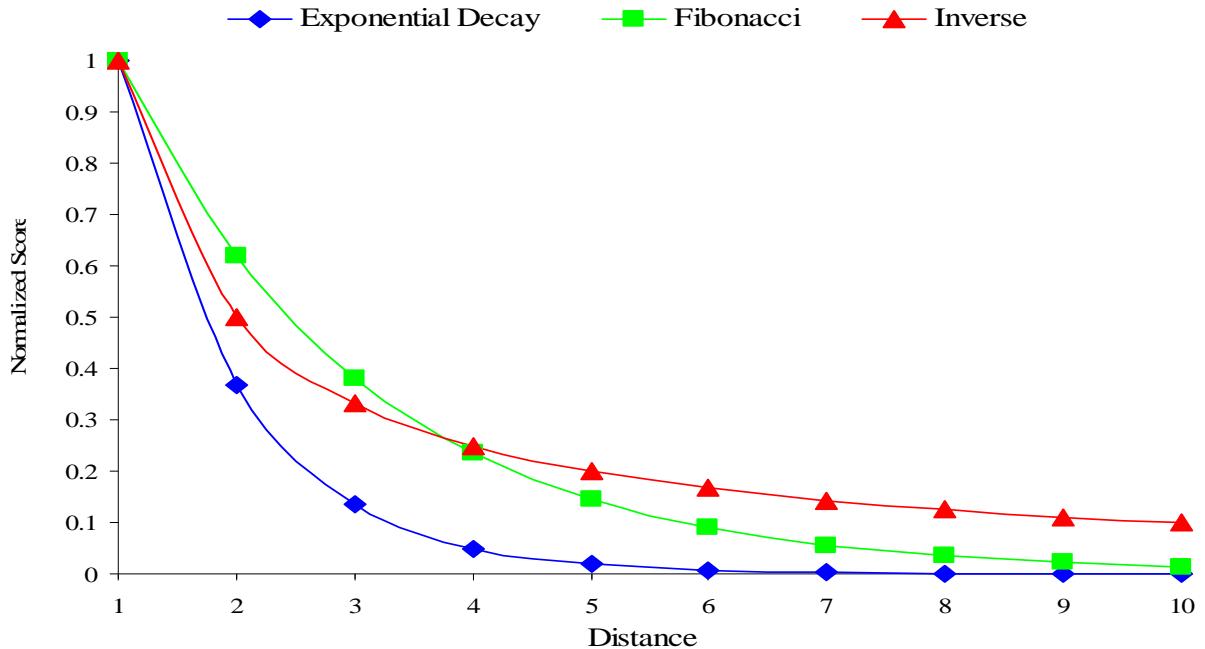


Figure 1. Distance Decay Curves. A comparison of three distance decay functions.

Exponential and inverse decay curves flatten out too quickly, resulting in poor differentiation among objects in the 6-10 distance range. The inverse decay curve does not approach zero, which also makes differentiation more difficult.

The decay curve using the Fibonacci sequence yields the most desirable result: it approaches zero at a rate that enables good separation of scores within the distance range of interest.

### Object attributes

For each unique object, a vector is used to contain its collection of attributes, where each attribute is a related object. Attributes have scores associated with them, which are the result of combining distance with other metrics. Vectors may contain tens of thousands attributes.

Significant relationships between objects are defined by both strong and unique connections between shared attributes.

### Associative memory modules

These object vectors are used to create AMMs. Relationships are inferred by considering the intersections of scored associative attributes of two or more object vectors within an AMM.

Subsets of AMM data can also be used as filters.

### Hierarchy-independent analysis

The basic elements of analysis are objects, which can be discrete elements of data, or pointers to collections of elements. The enables the system to operate in a hierarchy-independent manner.

Data collection identifiers are embedded in the framed context of each object, so queries can be traced to related groups of objects or sets of data sources.

### Process flow

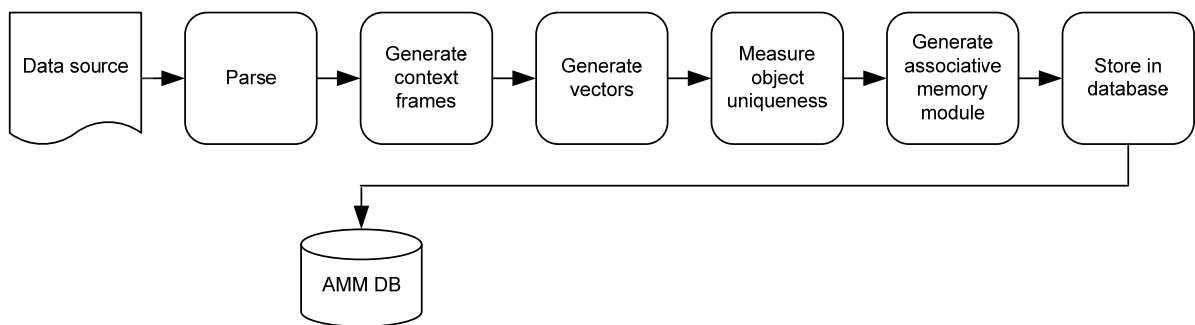


Figure 2. AMM generation.

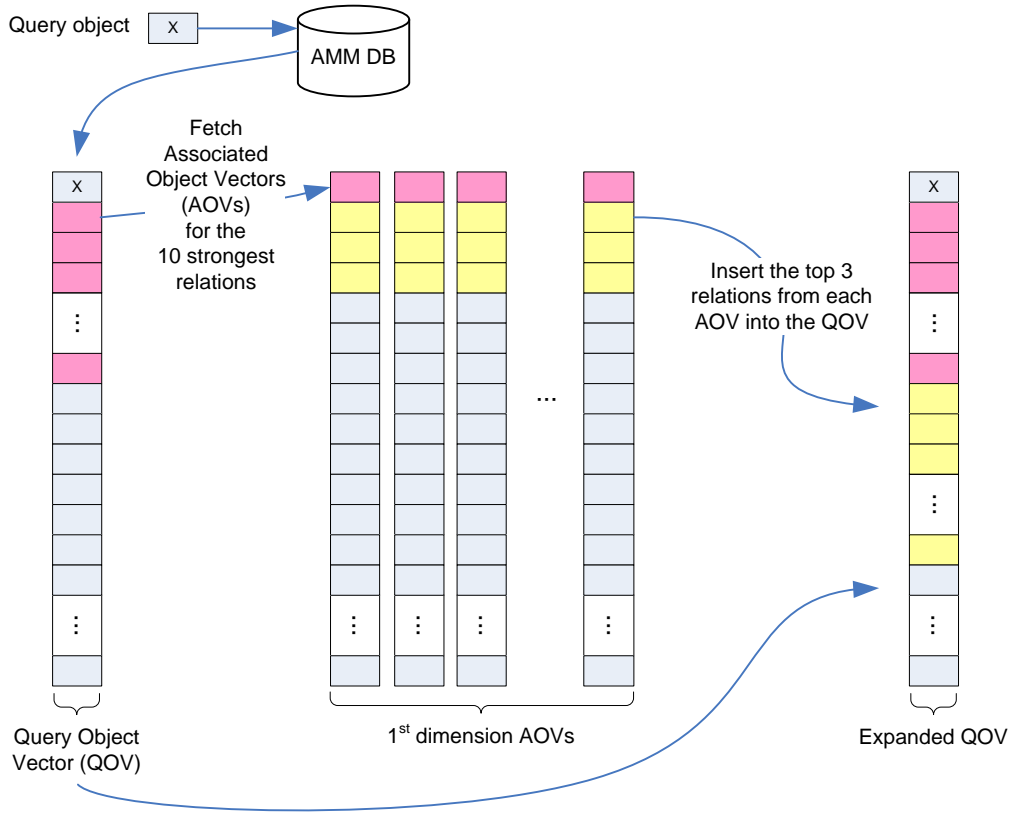


Figure 3. Query Object Vector (QOV) expansion.

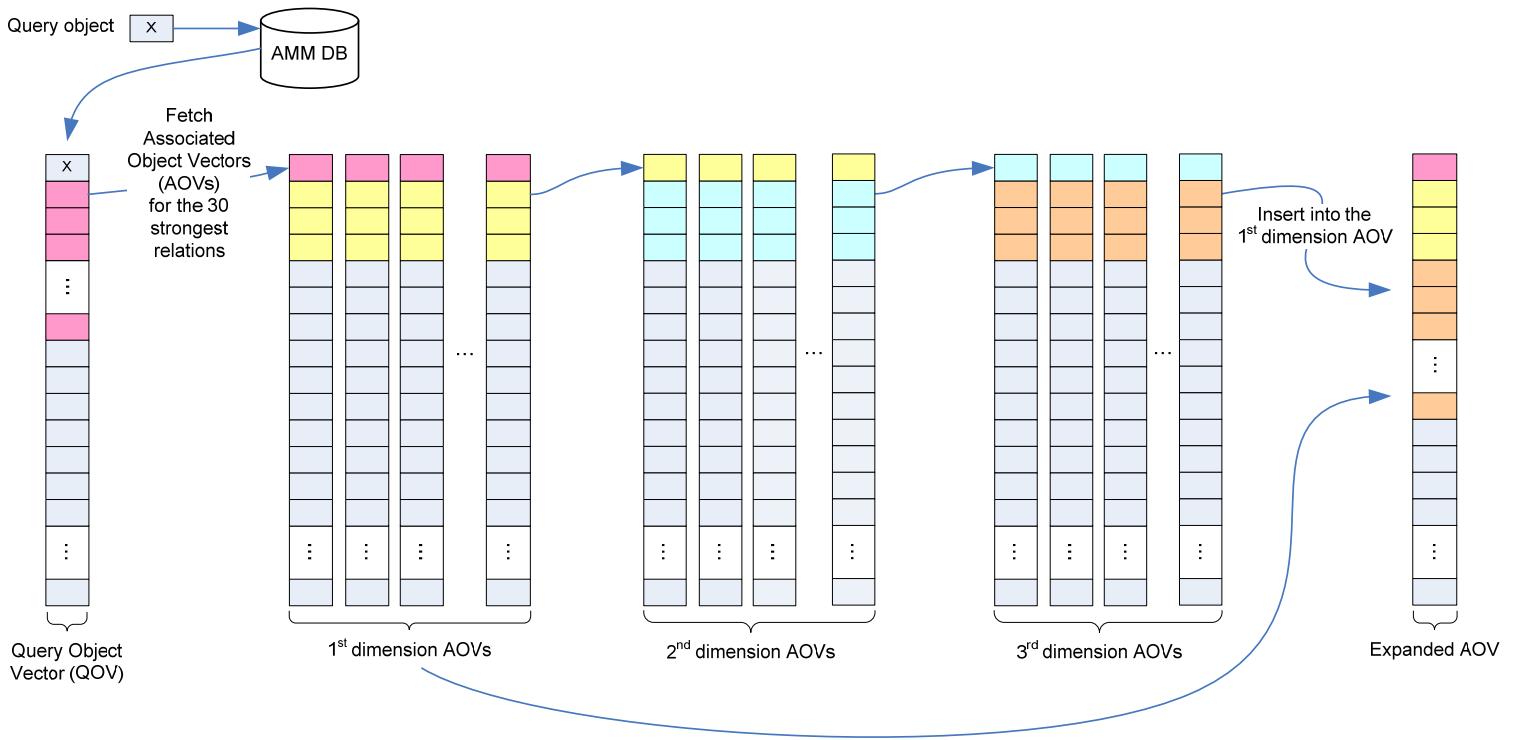
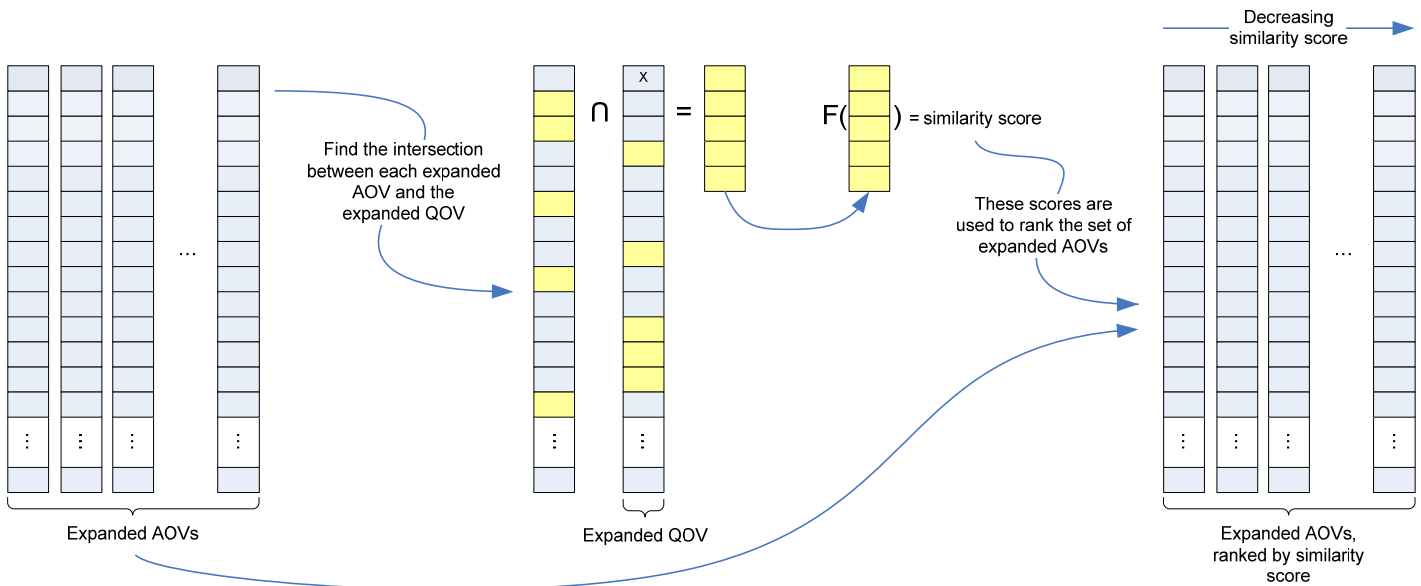


Figure 4. Associated Object Vector (AOV) expansion.



**Figure 5. Ranking expanded AOVs by similarity to the expanded QOV.**

Figure 3 shows the first phase of querying the AMM database for a given query object. Figure 4 shows how objects related at the 3rd dimension are pulled into the 1st dimension set of AOVs. Figure 5 shows how each expanded AOV is measured for similarity to the expanded QOV and how that score is used to rank the set of expanded AOVs. The set of ranked, expanded AOVs represents the most common contexts for the query object. The top-ranked AOV can be used to infer the likely context of a particular query.

## Conclusion

The statistical scoring of object associations can be compared to how humans discover relationships via reinforced learning (Wenhuan, X., Nandi, A. K., Zhang, J., Evans, K. G., 2005) and the process of auto-association. Important relationships are strengthened when reinforced.

Our system uses these principles to create AMMs, which can be used to provide the following data:

- How objects are related.
- In what contexts they are related.
- How strong those relationships are.

*In silico* emulation of biomimetic object-association strategies has proven to be very effective in relationship discovery and mining, leading to new findings in numerous industries including life sciences, finance, search, and contextual advertising.

## References

Dunlap, Richard A. 1997. *The Golden Ratio and Fibonacci Numbers*. River Edge, NJ: World Scientific.

Pinker, Steven. How the mind works. New York: W. W. Norton & Company, 1997.

Raghavan, V. V. and Wong, S. K. M. *A critical analysis of vector space model for information retrieval*. Journal of the American Society for Information Science, Vol.37 (5), p. 279-87, 1986.

Wenhuan, X., Nandi, A. K., Zhang, J., Evans, K. G. (2005) *Novel vector quantiser design using reinforced learning as a pre-process*, Source Signal Processing, Volume 85, Issue 7.

Xijin Ge, Shuichi Iwata, *Learning the parts of objects by auto-association*, Neural Networks, v.15 n.2, p.285-295, March 2002.