How to Introduce Artificial Intelligence to Link Analysts

How to Introduce Advanced AI Techniques to the Analytical Community without Actually Saying "Artificial Intelligence"

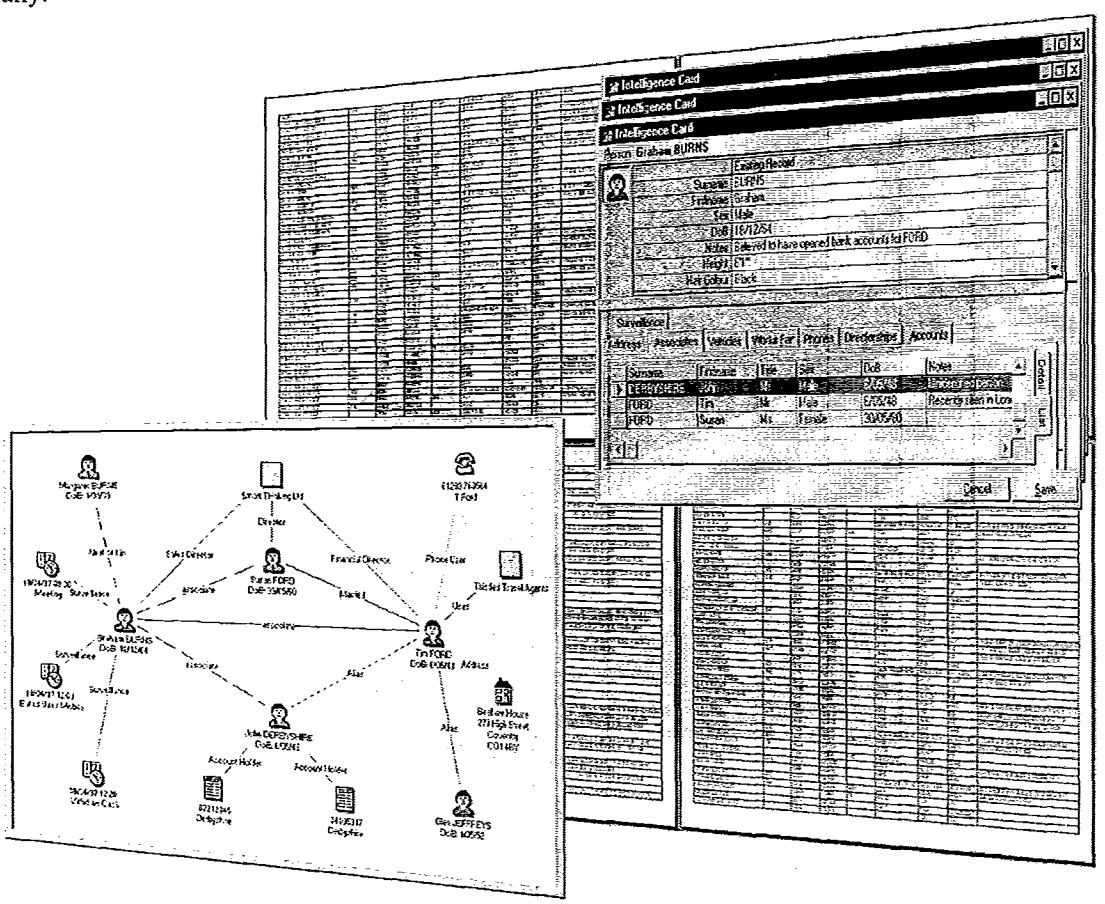
Mike Hunter
i2 Limited
Breaks House, Mill Court, Great Shelford
Cambridge
CB2 5LD UK
www.i2.co.uk

Introduction

i2 is one of the leading suppliers of Link Analysis (LA) tools with thousands of analysts using our products all around the world. As such, i2 is keenly aware of the problems facing analysts as they strive to cope with ever increasing amounts of information.

Artificial Intelligence (AI) offers many different approaches that can empower the 'Analyst'. Neural Networks are being used to detect fraud, advanced graph matching algorithms to unravel large networks, and language comprehension systems to extract entities and links automatically.

If the full potential of AI is to be realized, researchers must consider carefully what role the AI tools should play and how best to introduce them to the user base. To illustrate this point I will discuss one particular application of advanced algorithms and how we at i2 chose to introduce it to our user base.



The Role of AI

Analysis is currently primarily a manual process. By this I do not mean that the majority of analysts do not use computers. Far from it; i2 has itself demonstrated that software is a key ingredient in a modern analytical tool-set. Rather I mean that the key processing of the source information still takes place in the mind, rather than the CPU. The software may collect, store, index, retrieve and display the information; but the human analyst is required to understand it, to realize its significance, to assimilate it into a 'mental model' of the scenario and to draw the appropriate inferences.

The problems that analysts address have proved difficult to automate for many reasons. They are usually ill defined, complex, create large volumes of data and differ widely from one example to the next (each requiring its own real-world knowledge that the analyst must learn rapidly). The source information is frequently of dubious quality (often including mis-information) and must at least be treated with the appropriate level of skepticism. It also arrives at the analyst in a whole variety of formats include a significant proportion of free text. These qualities do not lend the area easily susceptible to automated approaches.

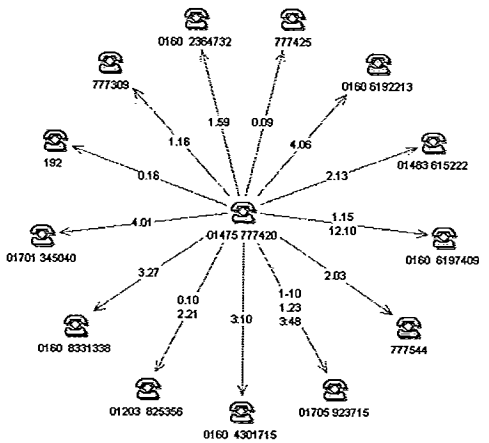
Successful Link Analysis software producers have to-date recognized this problem and concentrated on supporting the analyst's work rather than attempting to automate it. But it is clear that existing algorithms are of great utility. At a recent international analysis conference held in the UK I defined the term 'Assisted Analysis' as:

The processing of potentially large sets of data using a variety of automatic methods to identify specific areas of interest leading to subsequent manual analysis.

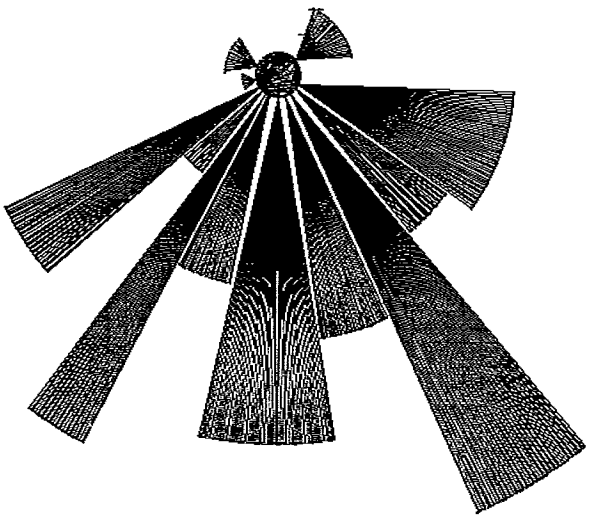
By this definition I wished to emphasize that to gain acceptance amongst the analytical community the new techniques should, at least initially, play a role subservient to that of the analyst - suggesting new avenues of exploration rather than dictating the next step. In the 1980's Prof. Donald Michie of Edinburgh University coined the phrase 'Intelligence Amplifiers' to describe the role that Knowledge Based Systems should play. I would suggest a suitable phrase to describe the best role of AI in LA might be 'Analysis Amplifiers'.

A Typical Problem

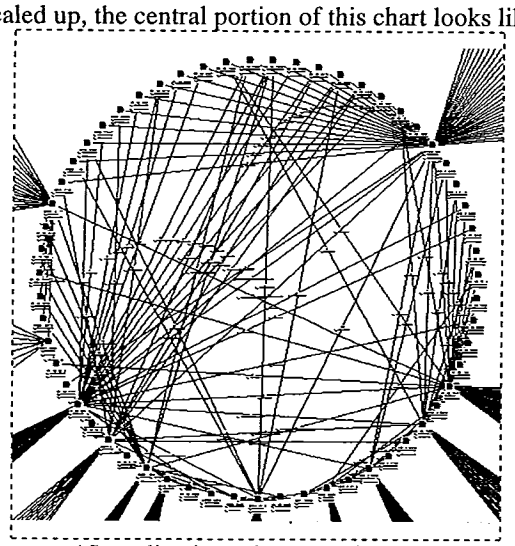
Link Analysis tools are frequently used to capture, visualize and analyze telephone call records. A simple set of data might be visualized as:



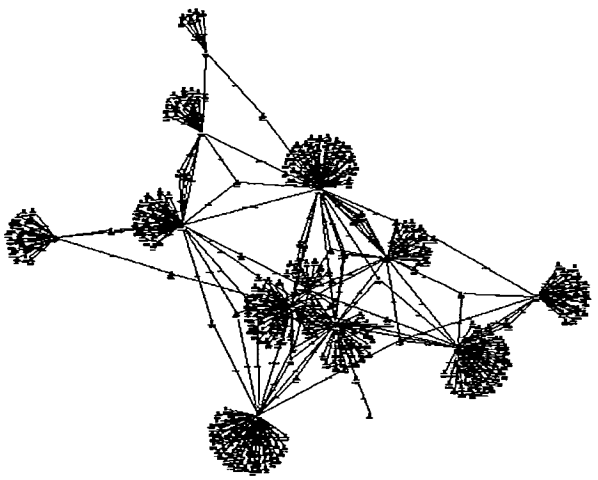
Unfortunately such simple data sets are rarely useful. The chart shown here is far more typical of something a tactical link analyst might assemble.



Scaled up, the central portion of this chart looks like:

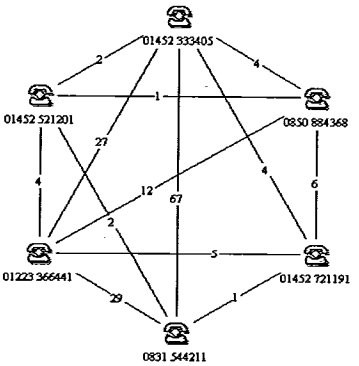


Clearly any AI applications that are addressing the area of LA must be scalable. Not only can the existing commercial LA tools, like i2's own, cope with large data volumes, but they offer various tools to assist with analysis. For example, they can efficiently re-organize the entities into layouts that reveal the underlying topology and key connecting entities:



Or they can examine a network and identify the most tightly bound clusters, which might in this case suggest a set of individuals acting as a group. For example, the most tightly bound cluster in the above chart, shown at the same scale, is:

or after re-arrangement:



Identifying such a tight grouping in a large network without automatic assistance is practically impossible. But the software cannot go onto explain the significance of the grouping; that is the task of the subsequent manual analysis.

These pictures are included to illustrate the kind of assistance already provided to modern analysts. What is clear is that the precise nature of the assistance required varies widely between analysis tasks. The challenge is to find a route by which such advanced functionality can be added to existing tools in a seamless and yet flexible manner.

An Architecture to Support AI Approaches

Analysts are now accustomed to commercial Link Analysis products, like i2's Analyst's Notebook, that are easy to use, robust, can integrate with existing databases and are capable of running on sensible platforms. If the new AI techniques are not to be stillborn, the prototypes must be constructed on top of existing proven platforms and tested on real pieces of work. The credibility of the AI techniques must not be endangered by experimental prototypes intended only for short-term use.

Luckily a variety of component based architectures now exist that allow products to be constructed in a compartmentalized manner, and for blocks of functionality to be added to existing applications incrementally. These include COM, CORBA and Java/JavaBeans. To illustrate the approach I will discuss how the latest version of i2's own core product has been implemented.

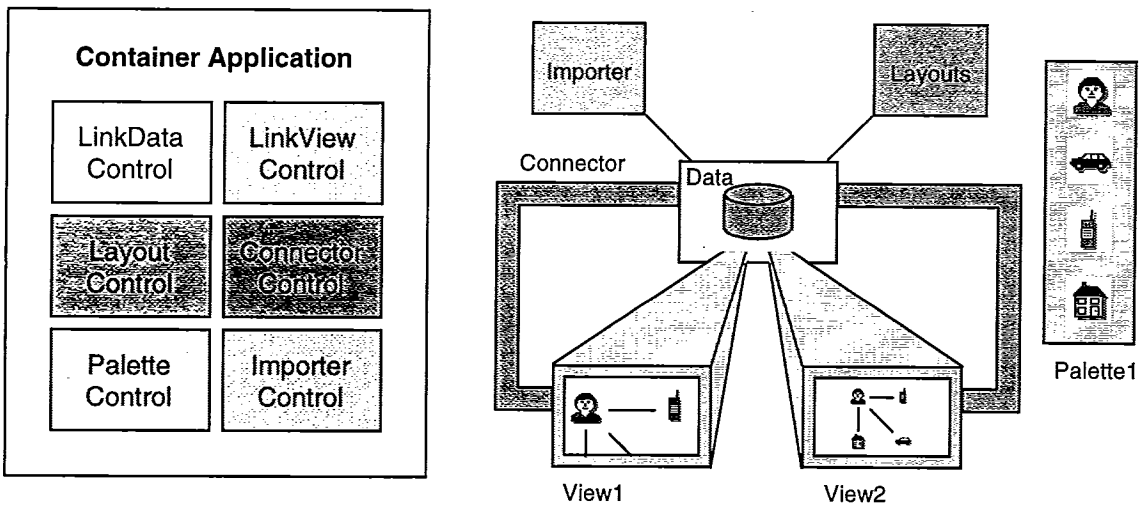


Figure 1 Link Notebook COM Architecture

Version 5 of i2's Link Analysis product, the Link Notebook, consists of a C++ wrapper containing a set of Microsoft COM components.

Each of the six components provides different functionality. The LinkData control, for example, manages the network of entities and links whilst the LinkView control displays it. The layout control includes several algorithms for arranging the network in different ways.

Component-based approaches can support several different product architectures. In the case of i2's Link Notebook, for example, there are three different options:

1. An external program running as a peer with the Link Notebook, communicating with it through the COM Interface (see Figure 2)

2. An additional 'plug-in' component, written in Visual Basic, C++ or any other ActiveX compliant development platform, that runs within the Link Notebook process but extends the Notebook's functionality (see Figure 3)

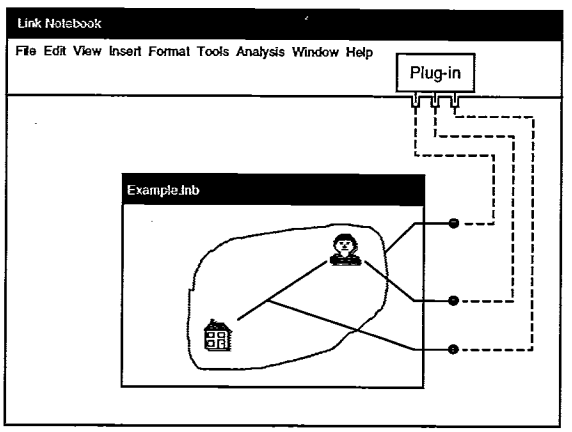


Figure 3 Link Notebook Plug-in Component

3. An external application that incorporates a subset of the Link Notebook's own components to provide Link Analysis technology within that application (see Figure 4).

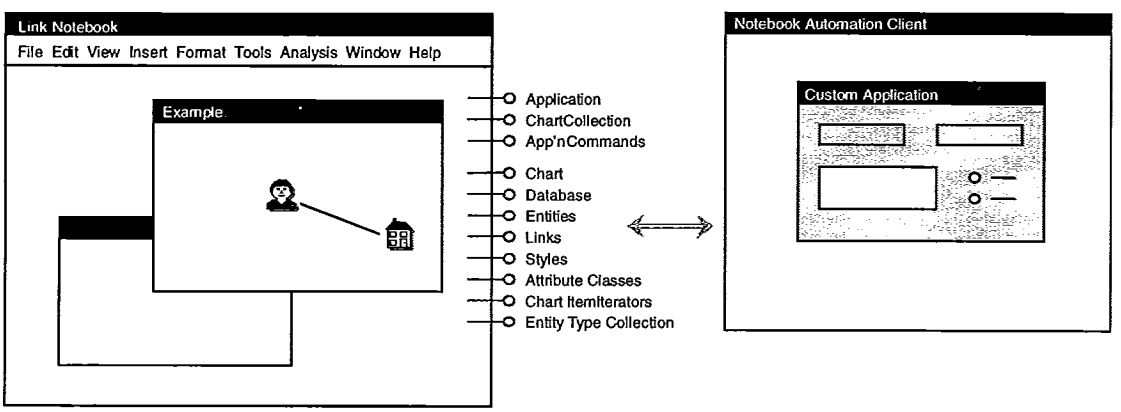


Figure 2 COM Interface Link with the Link Notebook

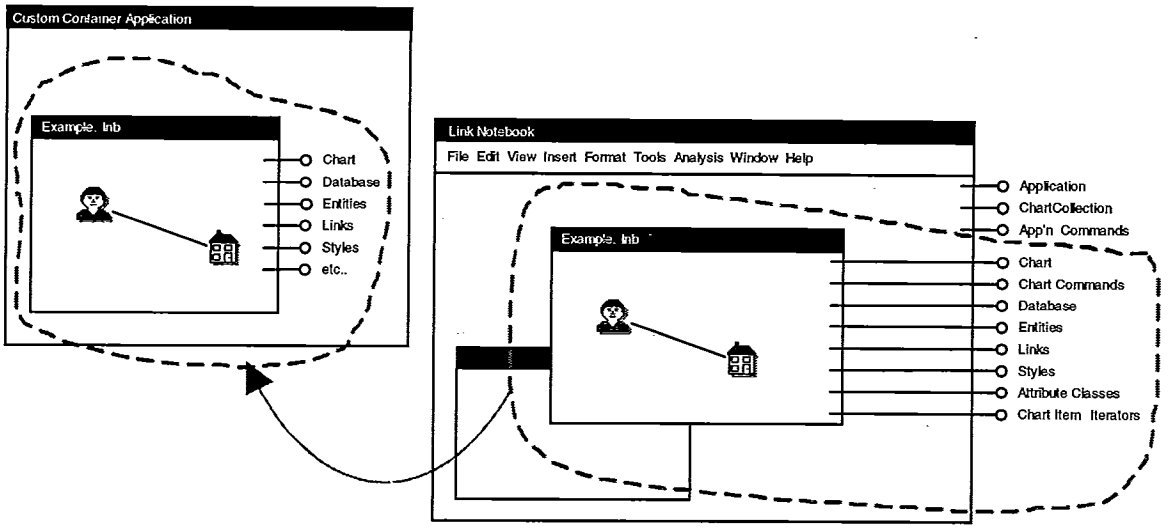
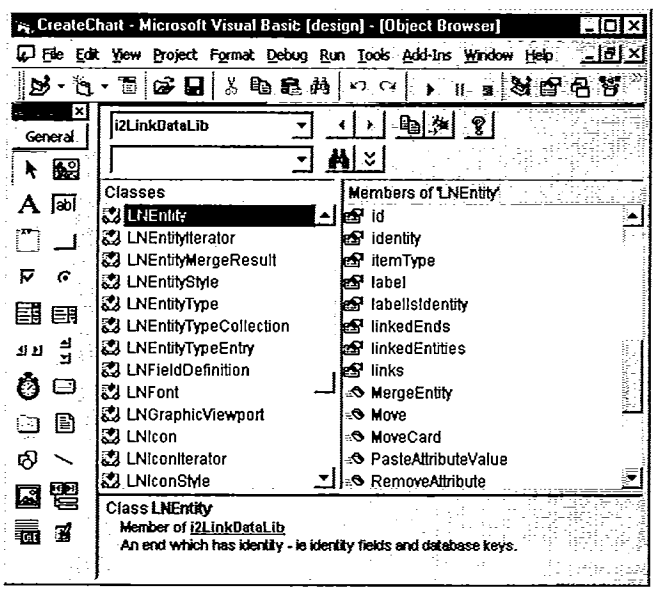


Figure 4 Link Notebook Components within a Container Application

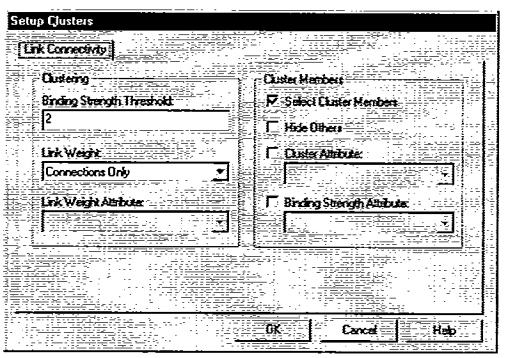
Whichever architecture is chosen, the object-oriented COM Interface is integrated into the standard programming environments.

The entity-clustering algorithm discussed above is an example of architecture 2. The algorithm examines the topology of the network to suggest tightly bound clusters. It is implemented as an additional, and completely separate, C++ COM component that is added to the core Link Notebook simply by registering its existence and declaring the functionality it offers. Once a component is so registered, the Link Notebook automatically detects its presence and offers it to the analyst like any other piece of functionality.



The user is unaware of the separate nature of the component. It is invoked from a standard menu and can interact with the link charts behind the scenes (via a full Object Oriented COM interface). It can add, remove, edit, move or in any other way interact with the elements in the link charts. The user can control the algorithm through a set or parameters edited in a familiar forms-based interface.

Architecture 3 allows new systems to be constructed by mixing and matching sets of existing and proven components. For example, all of the British Police forces will shortly begin using a new tool to run major investigations called HOLMES2. This investigative system utilizes a subset of the i2 components mentioned above to provide the users with link analysis capability within the



main application.

At i2 we are now working on a number of additional components performing a variety of tasks, some of which require AI approaches. As soon as each component reaches maturity it will be integrated with the core product as described above. This approach allows an unprecedented level of extensibility and flexibility. New components can use any of the latest AI techniques to interact with and manipulate the content of the link charts. And as the commercial tools support a documented interface, external researchers can write the components with little or no

support from the company offering the commercial platform.

Conclusion

If it is to be accepted, researchers need to consider carefully how they introduce AI techniques to the analytical community. They should look to:

- 'amplify' the role of the human analysts rather than threaten them with 'automation'
- utilize the existing, accepted analytical tool sets rather than seeking to replace them with immature tools
- work with real clients on real problems to ensure that the new techniques are truly useful and scalable.

Early adoption of the component based software architectures means that the existing commercial products can provide a steady platform without restricting the functionality of new AI components. Researchers can take advantage of these proven platforms to introduce the latest techniques to the analytical community in an incremental manner, as the techniques mature.

As the author's company, i2, has perhaps the first component based Link Analysis platform in widespread use, we would be happy to hear from organizations conducting such research that would benefit from exposure to real products, problems and users.