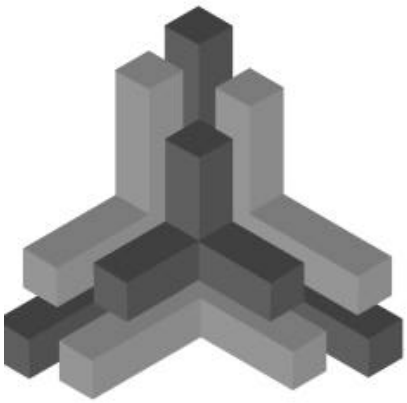
Future Edge



ice

intelligent classification engine

Functionality Overview Whitepaper

1 Introduction

The Intelligent Classification Engine (ICE) is a fully functional concept extraction engine. Rather than relying on the appearance of keywords within a document, ICE is able to learn about the subjects that its content base includes, relate them, and rank their importance. Because of this ICE is not only able to provide much more relevant results, but is also able to accept fully natural language queries in any language.

Integrating ICE into an application allows rich personalisation and community development features to be added quickly and easily with minimum ongoing maintenance, as the processes involved are largely automatic.

2 Content Segmentation

By integrating ICE with our sync family of information collation tools content can be broken into three separate areas:

1) **Internal Content**
   Internal Content is classed as that which is under the direct publishing control of the client. This is not necessarily limited to one server or domain.

2) **Approved Content**
   Approved Content is content from specified websites and databases that the client has reviewed and decided will complement the Internal Content. Approved Content adds richness to a user's experience, increasing the likelihood of user retention.

3) **Internet Content**
   Internet Content is content from the Internet at large. The client has no control over the content, and its relevance could be low. However, given that a client's Internal and Approved Content are likely to be on a particular set of subjects including references to Internet Content can increase the relevance of the content base to a larger number of users.

Content does not necessarily mean a ‘news article’. ICE is also able to index information about, for example, products, TV and radio listings, and even people. Because of the concept extraction technology, as long as there is a natural language description of something, ICE is able to understand how that object relates to all the others it knows about.

3 User Interaction

ICE provides a number of methods for users to discover the most relevant content from each of these areas:

1) **Natural Language Search**
   ICE allows a user to enter a query in full natural language rather than the older keyword and Boolean approaches. Results are provided with intelligent query-specific summaries.

2) **Automatic profiling**
   Once a user is identified, ICE reads along with them, learning the subjects that interest the user. Each user's experience is then highly personalised with no effort on their part. As this profile is generated automatically and in real time it is always up-to-date, and reflects the variety of interests that a user may have.

3) **Trackers**
   ICE allows a user to enter their interests in natural language. This allows a user to keep up to date with the subjects that interest them automatically. Rather than most present methods that rely on a user selecting very broad subjects from a list, ICE allows users to specify very clearly the subject that interests them. Not 'rowing' but 'Steve Redgrave', not 'Athletics' but '100m final'. ICE will also look across the complete content for relevant documents and looks at the document concepts, not just the keyword tags.

4) **Path of Interest**

Because ICE understands which concepts are more important within a document it is able to automatically link related information together. Rather than a publisher having to search for relevant related documents, and categorise that into a very limited set of subjects, ICE automatically and dynamically provides much richer related links across the complete range of concepts contained within a document.

4 Community Development

ICE builds up an understanding of a user's interests, by watching the documents they read, by noting the queries submitted, and by noting the tracker definitions. Because of this, ICE has a comprehensive understanding of the interests of that user, and their relative importance. From this, ICE is able to look at not only linking users to documents, but also linking users directly with other relevant users.

5 Content Publishing

ICE can also help on the publishing side:

1) **Automatic Related Documents**

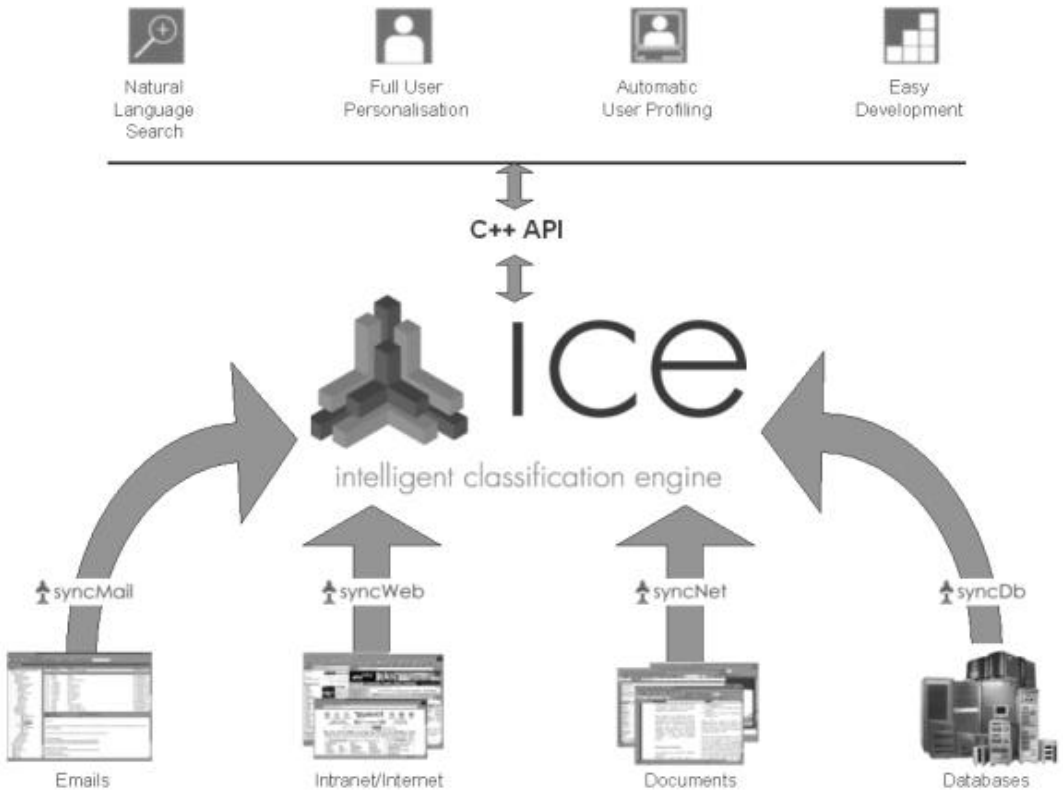
ICE allows publishers to save time by automatically and dynamically relating documents together. The client then saves the time spent manually linking and tagging documents, and the user gets a richer, more relevant experience. Taking BBC News Online as an example, which publishes around 1,500 articles each day, the 5 minute saving in manually specifying related links and keyword tags could result in a £0.75m saving per year.

2) **Automatic Categorisation**

Because ICE learns it is able to watch the type of documents published to different categories and develop an understanding of the concepts contained within those documents. ICE is then able to suggest relevant categories for the publication of new content. Taking into account ICE's ability to index from

a wide variety of sources this then means that relevant community forums, newsgroups, or even chatrooms can be included as relevant locations to publish to.

6 ICE Family Architecture



7 Further Information

For further information please contact **info@futureedge.co.uk**