Seed Fund Application

SECTION ONE

Executive Overview

1. Title of the project			
Network Analysis Commercialisation Project			
2. Applicant Name /R&E Contact			
Dr John Jones/David Wells			
3. Applicant Department			
Biological Sciences			
4. Date of Application (Year and month)			
Year	2011	**Month**	May
5. Amount of Investment requested from Seed fund(£)			
£19,800			
6. Summary of Project *Approx 50 - 100 words that give an executive overview of the project*			
Dr John Jones has developed a novel technique for analysing information networks. The technique has a number of potential commercial applications in high-value markets. Of these, the most commercially important are: • social media analytics: understanding online structure and content, particularly in formal and informal social networks • document and content classification, including document and web searching (information retrieval) • optimisation of computer hardware networks			

SECTION TWO
Project Description

1. Project Description

Approximately 500 words that describes the work to be carried out and the outcomes.

Background: information network clustering to identify meaningful groups

Dr Jones has already shown that his technique is capable of identifying meaningful groups of users on Twitter using the frequency with which they communicate among themselves via direct messages. Twitter users whose messages to one another form closely-linked networks are considered to be members of the same group. Manual inspection of the actual messages exchanged within such groups confirms that they indicate Twitter users with particular shared interests, for example pets or politics.
This initial analysis used a published network partitioning algorithm and is not in itself novel. Similar analysis has been used to identify social network groups in other concepts, for example in the blogosphere.

However, having identified such groups of closely linked social network users, linked by messages to each other and/or by linking to similar online resources, it is often impractical to characterise them by manual inspection, which is slow and expensive.

For example, the paper “Mapping Iran’s Online Public: Politics and Culture in the Persian Blogosphere” [April 2008: The Berkman Center for Internet and Society at Harvard University] used a team of Persian speakers to read and classify over 500 blog entries to characterise the groups of Iranian bloggers that it had identified. However, such resources are not always available.
Moreover, such human inspection has been shown to be unreliable as well as expensive.

Nevertheless, human inspection is still widely used in commercially important fields such as sentiment analysis (a branch of social media analytics dedicated to understanding the opinions of online users towards products, services, political policies, etc.), and document classification (including the classification of online content and news articles).

In other words, even such an expensive and unreliable technique is of commercial value.

Dr Jones’s technique, however, allows for classification to be automated, rather than relying on human interpretation.

Novelty: automatic characterisation of the meaningful groups

The novelty of Dr Jones’s technique is in using one or more automatic mechanisms to inspect groups that have been identified as closely-linked and characterise them based on the associated use of language. For the Twitter groups, for example, the messages between the Twitter users in each group were concatenated into one large file for each group. The file for each group, representing the use of language in the conversations between members of that group, was then compared with the concatenated files from all the other groups and the keywords identified that were most characteristic of that particular group.

Dr Jones's insight that combining these two known techniques is particularly powerful and valuable. A patent was filed in December 2011 and the international search results are encouraging.

Commercial applications

Such an understanding of the nature of the groups can be used either to interpret the dialogue and behaviour of each group (for "pull applications") or to understand how best to communicate to members of the group (for "push applications").

Pull applications include:

- Sentiment analysis ("is the opinion of this particular group important to us? If so, what kind of words are likely to be used to indicate approval or disapproval, given the nature of the group")
- Interpreting group recommendations ("what kind of products/videos/etc does this group like? If I am similar/dissimilar to this group, these might be good/bad recommendations for me").

Push applications include:

- Targeted advertising and other offers, based on knowledge of the group's likely interests and the type of language that is used in the group
- Making better recommendations based on known or inferred group membership ("as the current user appears to be a likely member of these group(s), recommend the products/services/content that are known to be of interest to these group(s)")

All of these are extremely commercially active areas. A considerable amount of research is being carried out into improved techniques to support them.

Dr Jones's technique also has applications in the related areas of:

- Document classification: automatically organising existing libraries of commercial, technical or scientific information and classifying new information appropriately
- Search (information retrieval): identifying information in a document repository or online that matches a particular requirement

2. Commercial Opportunity

a) Anticipated Outcomes

From this project and the ultimate project goal.

The outcome of this project is to produce compelling material to demonstrate the potential of the technique to analyse a real-world social network, suitable for presentation to potential licensees of the technology and/or R&D partners.

Although we already have slightly similar material, based on the Twitter analysis previously described, the technique has been significantly enhanced since that pilot project. Moreover, the data produced was tailored to the demands of presentation in a scientific paper describing the algorithm, rather than to sceptical potential commercial partners.

In this project we will produce a fresh analysis to group 100,000 users from the video website YouTube to illustrate how the technique could be used to support the applications previously listed.

The ultimate goal is to license the patent and, preferably, to collaborate with a licensee to further develop the technique via a new research group at RHUL.

b) Financial

Estimate the level and timeframe of the commercial opportunity

The technology has the potential to attract at least a six figure sum by combining licensing and joint R&D with one of the larger players in the industry, for example Google.

Timescales in social media analytics are rapid, and we would expect to attract the interest of a licensee before additional patent costs will be incurred in December 2011.
Timescales in other application areas, for example document classification and search, are slower, but we will work to the same deadline.

3. Describe the status of IP and IP investigations

a) What IP is associated with the project?

UK Patent Application covers the technique and a number of applications of it included those mentioned above. The initial patent search was received in April 2011 and is encouraging.

b) What new IP will be created during the project?

The project itself will generate IP in the form of:

- program code to download metadata from YouTube and pre-process it
- a repository (SQL Server database) containing information about c. 100,000 YouTube users
- analyses of that data using code already developed by Dr Jones

All of this IP and associated know-how will be the property of RHUL.

4. Long Term Opportunity

a) Potential for further work

Opportunities which may arise for research, consultancy?

The primary goal is to seek an R&D collaboration with a licensee that will fund additional staff at RHUL to further develop the technology.

b) Is there a wider context in which to consider the value of the work

Strategic development, Developing partnerships, reputation, social benefit?

Participation in a very fast-moving highly commercially-linked area of computer science such as social media analytics will enhance the reputation of the College for an ability to engage in such activities.

5. **Provide a description of how the investment will be spent**

For what? When needed? Milestones & timescales.

a) How will money be spent?

1) Egham Computer Consultants (ECC) has quoted against a detailed specification to implement the software needed for the YouTube analysis: £11,798 + VAT = £14,400.
A separate independent assessment by another software SME arrived at a very similar development price guideline (£12k + VAT).

2) Dr Jones an experienced software developer and manager, will carry out 5 days' work: 2 days to assist in developing the initial specification for ECC and to give feedback on their project proposal and 3 days to review the detailed technical specification and to chair a project kickoff meeting, a progress meeting and a wrap-up meeting. At £500/day + VAT this is £3,000.

3) *Dr Ray, a social media and web marketing expert, will carry out 4 days' work: 1 day to assist in developing the initial specification; 1 day to develop three specialised one-page covering letters describing the technique for three target markets (Google; other content owners (Facebook, Twitter); social media analytics software companies), 1 day to critique and design the presentation of the project result data, 1 day to identify appropriate initial contacts at Google, Autonomy, Alterian and IBM. At £500/day + VAT = £2,400.

b) Project milestones

Milestone/Stage	Time	Cost funded by this application (£)	Associated FEC(£)
1) ECC development, initial design of results presentation	1 month	£17,700	N/A
2) Process results, tune algorithms (Dr Jones)	2 weeks	N/A	N/A
3) Develop presentation of results with Dr Ray	2 weeks	£2,100	N/A
TOTAL			**£19,800**