PERSONALIZING FINANCIAL INTERACTIONS ON THE NET: TECHNOLOGY AND SCOPE

Maurice D. Mulvenna[§], Tony de Bree[**], Sarabjot S. Anand[§]
Alex G. Büchner[§] & Matthias Baumgarten[*]

[§]MINEit Software
5a Edgewater Office Park, Belfast, Northern Ireland
[**]ABN AMRO Bank N.V.
Gustav Mahlerlaan, 10, Amsterdam 1082PP
[*]NIKEL
Faculty of Informatics, University of Ulster, Shore Road, Newtownabbey Northern Ireland

1. INTRODUCTION

The web is fast maturing into an important marketing medium that provides businesses and individuals alike with the ability to undertake one-to-one marketing and provide personalised services to their customers both from a traditional supplier as from a new self-service perspective.

In this paper we describe innovative Web Intelligence research in the CERENA project, that provides marketers and eventually customers and visitors in e-finance with access to data mining technology that can sift through large amounts of data collected automatically on customer interaction with the businesses web site. The technical objective of CERENA is to bring to light marketing knowledge that is mutually useful to the financial institution and to their customers.

Section 2 describes the background to the project, and places the research in context. Section 3 defines web intelligence, while section 4 describes web mining components within the framework of web intelligence. In section 5, the web interactional components are described. Section 6 discusses the use of web intelligence in the new financial services arena.

2. BACKGROUND

The late 1990‘s witnessed a thriving business arena, which has effectively moved business models online under the form of different mixed models (click-and-mortar) and channel-mix. Although the number of online customers is steadily increasing, electronic commerce (eCommerce) and electronic business finds itself in the phase of moving beyond static business models. Today’s leading financial services institutions face the challenge of having to adjust their method of customer communication to Internet standards while at the same time also supporting more traditional distribution channels at the request of the customer. National differences between the respective European countries are still considerable. These differences are related to amongst other things different channel-mix and different needs and requirements regarding service levels and product characteristics. Legal barriers, languages and different pricing policies are other issues preventing European leading institutions from marketing Europe as one single open market. The current challenge is to disband long and tedious product descriptions for the sake of rich content and provide personalised interaction experiences based on different data. These sources will include more and more metrics related to the actual behaviour of users of the site. Web-based business requires succinct information

and openness about sensitive business information such as pricing or interest rates. Dealing with a continuously fastidious customer-community, business has to be made for customers individually. Similar to the 'classic' market-situation, competition via customer-care follows establishment of the Internet as a business-platform. The core topic is to create customer choice and an environment promoting customer loyalty through service and personalisation, in order to maximise online customer experience and benefit.

	Traditional Business	Early Web	eCommerce
Value Proposition	Product	Information	Service
Speed	Weeks	Days	Minutes
Scope	Local, regional	National	Global, regional, local

Figure 1 New business drivers & value propositions

Figure 1 shows the rapid and significant changes that businesses, in particular financial services organisation such as banks, are undergoing. In the classic market situation, the value proposition focused on the product, with speed to market and time to close sales of months or weeks, and few businesses operating with a global reach. In the early days of the web, leading businesses in commodity industries modified their traditional business models, adding value to core product value proposition in terms of information either by internal transformation or by starting all over. But despite the improvement in reaction times to days, the scope of such re-engineering was primarily national. Now, at the changing of the millennia, as technical innovation accelerates the invention of new, sophisticated business models (Timmers, 1998), so the focus is on the provision of services, where the speed of provision and the reaction is measured in minutes and seconds, and the reach is global and direct to the individual or the corporate customer. However, the majority of business interactions is still taking place either through other existing electronic media like PC-banking, ATM's or call-centres. And depending on the culture, many people still prefer face-to-face contact especially for tailor-made advice.

In this new environment, financial institutions and other providers of financial services recognise that the provision of high quality customer care is paramount in order to retain customers who no longer have the local branch 'lock-in' preventing them from switching. Moreover, a clear shift from transactions involving standard banking products towards personal finance and advice is also clearly visible. Indeed financial institutions are recognising that many new players are emerging from the melting pot of Internet business models and opportunities. These new players are characterised by a desire to win new high value customers, precisely the customer segments that provide many financial organisations with their profit. Furthermore, the new entrants are aggressive marketers, bundling financial services with other non-financial products and services. Many are new start-ups, and are free from the organisational and technological legacies that hinder the reaction times of larger established organisations.

Leading edge research such as CERENA engenders a paradigm shift in the field of customer satisfaction, often focusing on product attributes towards customer loyalty, driving revenue streams over a longer period of time. Financial institutions dealing with better informed customers who have more choices will make less money on each product they sell and each dollar/Euro of asset they service. Moreover, the shift in power will be accompanied by a shift in value, where more of the value in every transaction will end up with the

consumers, whether they enjoy that value in the form of lower prices, better products, superior service, or a combination of the three. In an ideal situation this leads to the consumer empowerment, giving them more benefit than they previously enjoyed: The customer-centric model. However, financial institutions and other companies in the marketplace and market-space are learning fast. And more and more financial institutions are implementing off-the-shelf software in support of their transition towards the mass-customisation business model.

The following sections concentrate on an outline of the components that enable the analysis of interactional and transactional information in the Internet channel with the objectives of combining the reduction of interaction costs and transaction costs within the financial institutions and with the customers and to enhance life-time-value through higher customer loyalty. Although the technology could in principle be used in combination with any web site for any internal or external user group, the focus will be on e-finance

3. WEB INTELLIGENCE

Web Intelligence is the application of business intelligence software and methods to Internet data. It consists of a three-tier architecture:

- Data based components: Data base / data warehouse; On-line Analytical Processing (OLAP) tools; Extract, Transform & Load (ETL) tools; Query optimisation, etc
- Web mining components: data mining tools & algorithms specially optimised for web data characteristics - including classification, segmentation, scoring, etc.
- Web Interactional components: knowledge representation schemas and mark-up languages, recommendation engines; rules based engines; collaborative filtering tools; packet sniffing toolsets; content servers, profiling engines.

The 3-tier architecture is necessary to cope with the extremely large volumes of data generated by the single channel of Internet interaction in eCommerce (Figure 2).

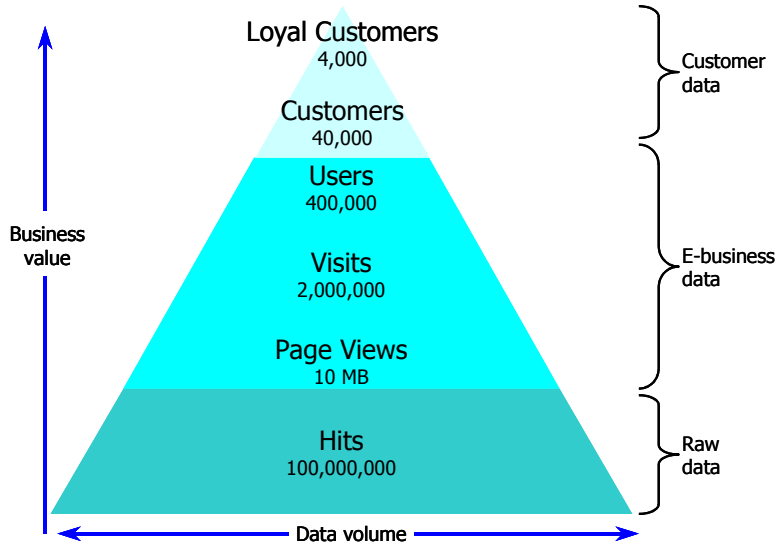

Figure 2 Business value: data perspective

In the present competitive environment, eCommerce organisations need to win and retain high-value customers to remain competitive. One technique that can be used to achieve greater loyalty from customers in Internet based retailing is to offer services that are predicated upon closing the marketing circle. These services include personalisation, recommendation, optimisation, etc. However for these services to succeed, "a library of rich visitor profiles must be present" (Schmitt *et al*, 1999). Web intelligence is the means by which eCommerce companies can build these rich visitor profiles, utilising all the data that is generated by visitors and buyers at Internet eCommerce sites. A detailed description of the data based components, including a description of web log data characteristics is covered in other articles by the authors (Mulvenna *et al*, 1998)), and will not be presented further. The following two major sections describe web mining and web interactional components respectively.

4. WEB MINING COMPONENTS

In the context of web intelligence, web mining may be defined as the application of data mining techniques to Internet data. Web mining has been sub-divided into web structure, web usage, and web content mining. Web structure mining is the application of data mining techniques to web site structures. In many cases this may be the entire web, and research in intelligent search engines and intelligent agents is described in many articles. In our research, we define web structure mining as the mining of Internet data, *together* with data about the structure of the site. This may be thought of as enriching the efficacy of the data mining process with domain knowledge. The application of domain knowledge is further discussed in the analytical process section.

Web usage mining is the application of data mining to Internet usage data, which is described in detail in other papers (Baumgarten *et al*, 2000; Büchner *et al*, 1998). Web usage mining forms the core of our research in web mining for web intelligence, and log data provides the foundation data for visitor analysis. This type of analysis of the visitors to a web site can be subdivided into technographic and psychographic analysis[1] (Schmitt *et al*., 1999).

A Technographic analysis focus on what is known about the visitor's technical platform, i.e., operating system, browser, plug-ins, user language, and cookie information. On its own, this information is not a rich source of discriminatory data for visitor profiling but in conjunction with the homogenous data sets available after extract, transform & load operations to data warehousing, it contributes significantly. Psychographic analysis is the examination of what we know about the behavioural patterns of web site visitors and other users. This includes the routes taken by visitors through a site, the time spent on each page, route differences based on differing entry points to site, aggregated route behaviour, general click stream behaviour, etc. This is the information of most use to web marketers, and is equivalent to marketing intelligence about where shoppers enter the store, where shoppers go in the store, where they leave the store or eMarketplace, what they look at but don't buy, what they buy and how quickly, etc.

Web content mining is the application of data and text mining algorithms and techniques to the contents of web pages, usually written in HTML. At its simplest, this entails the extraction of text between HTML tags for headings and titles, or the extraction of the HTML Meta tag content. Our research is based upon XML and RDF-based[2] data schemas that help to ensure correctness and proper context.

[1] http://www.forrester.com

[2] http://www.w3.org/TR/PR-rdf-schema

The following sections briefly describe the various techniques and technologies available for web mining, used primarily in the context of usage mining.

4.1. Segmentation

A starting point for traditional marketing is the segmentation of the customer base into smaller, more manageable groups of customers that have similar interests or other characteristics with respect to their interaction with the business. In the context of a traditional retailer and in the absence of more customer data, this generally implies customers who buy similar products. In the case of an e-business, the web logs provide a large source of additional information about customers at no additional cost.

Segmentation based on navigation behaviour is carried out on a large number of attributes (the number of pages on the web site). Generally, two kinds of clustering may be undertaken: session clustering based on pages visited and the average time spent on the pages.

4.2. Sequences & Click Streams

Navigation of a web site is temporal in nature. Therefore, one of the basic forms of knowledge that needs to be discovered from data collected in web usage is navigational sequence knowledge. This describes the most commonly tread pathways through the web site, where a pathway is defined as based on a threshold value of sessions that follow the pathway, referred to as support. Two types of sequences may be discovered: Open *sequences* and *click streams*. A sequence is a list of web page accesses ordered by the time of access within a visitor session or across visits for a particular visitor. An open sequence is not necessarily a contiguous navigation of the web site. This means that an open sequence of the form `<index.html, orderform.html>` does not imply that there is a direct link between the `index.html` page and the `orderform.html` page that was navigated by customers that support this sequence. Customers supporting this sequence may have taken distinct paths from index.html to `orderform.html`; however, none of the individual paths navigated by the customer have the required support value to be considered as interesting within their own right. A click stream is a special type of sequence where the pages accessed have contiguous navigation. Thus a click stream of the form `<index.html, orderform.html>` does imply that a direct link exists between the `index.html` and `orderform.html` page and that the customers navigated this link during a single visit.

Three kinds of domain knowledge may be used within the discovery of sequences. These are navigational templates, network topologies and concept hierarchies. Navigational templates are used to tailor the sequences discovered from the log data to the users needs. Using these templates, goal-driven navigation pattern discovery is possible through the specification of start, end, as well as middle pages for sequences that are of interest to the user. A typical start locator is the home page, a customer support page, or a URL providing information about a special marketing campaign. A typical end page is a purchase page or a page for requesting more information or a link to an actual person on a telephone. The second type of taxonomical domain knowledge is that of *network* topologies, which is useful when the topology of web site or a sub-network of a large site is of interest to the user for the discovery of sequences. This domain knowledge is used to include or exclude parts of a web site from analysis. Finally, concept hierarchies may also be specified and used to reduce the granularity of the discovered sequences in a similar way as their use within segmentation.

4.3. Bayesian Networks

Bayesian networks are a popular technique for representing uncertain knowledge and making probabilistic inferences. Nodes within the network represent variables while arcs connecting the nodes represent interdependencies between the variables. As most domains contain some uncertainty, Bayesian networks have been receiving a lot of attention with a number of applications of the technology in industry appearing in the last ten years.

Online behaviour analysis is an excellent candidate for the application of Bayesian networks due to the high levels of uncertainty in visitor online behaviour. While Bayesian techniques can be applied to tasks such as classification of online visits, it is the application of Bayesian networks to the prediction of online navigational behaviour that make it such an attractive proposition in the field of web mining. Bayesian networks are often interpreted as representing causal knowledge and techniques to extract such knowledge have been developed in literature. As an online visitor navigates through the web site, each page that they visit represents evidence that points to the destination of the visitor in the web site. Using Bayesian inference, it can be inferred as to whether the visitor is about to abandon the web site or stay online and if so, where the visitor is heading. Such inferences provide a head start to personalisation engines (see Section 5.3) that can use this knowledge to formulate ways of enriching the experience of the visitor, thereby increasing their loyalty.

The tight coupling of the knowledge representation and inference engine make the deployment of the Bayesian nets even more attractive in web intelligence.

4.4. Content Mining Algorithms

An online visitor on an e-finance web site provides digital imprints or clues to their interests and objectives in navigating a site. Organisations that leverage these clues, collating them into ‘partial profiles’ of their visitors will clearly be better equipped to provide the visitor with a more enriching online experience. One aspect of the visitor's online behaviour is the content of the pages that are more prominent (for example, based on the time spent on the page) in a visit by the visitor. Content mining algorithms attempt to extract the essence of the text document to provide a more concise profile of the users interests.

For example, if a visitor on an online investment web site that also provides newsfeeds from REUTERS, views a number of new stories on each visit, text & content mining may reveal that the visitor always reads news feeds on corporate tax schemes in Italy - a much more concise profile that a set of text documents that were accessed by the visitor. This profile may then be used to extract ‘similar’ documents for the visitor and may even lead to opinions on certain companies deemed as potentially interesting to the visitor being provided to the visitor using the deployment techniques as described in section 5.3.

4.5. The Role of E-metrics

Little attention to date has been paid on evaluating the success of web sites in terms of their utility to the users. Server statistics provided by web log analysis tools provide metrics for evaluating the success of the server in serving pages to users. However, no insight is given on how useful the content provided was to the user. Evaluation of the effectiveness of a web site must be based on the business goal of the e-business and metrics would provide the basis of evaluating the success of personalisation activities. Web Intelligence tools based on web mining have an important role to play in the development of these e-Metrics.

E-metrics can be summarised as having the following characteristics:

- Quantitative measures of online success;
- Aligned with business objectives (business performance);
- Valuations of corporate web sites;
- A common language to describe e-business results.

Diane Casey, Managing Partner of Financial Services at Grant Thornton says, "Banks need to view the Internet as one more marketing tool. The Internet is nothing more than another channel, and they need to put performance metrics in place before they start". Common e-metrics include:

- Visitor metrics: for example, visitor information, visitor-customer conversion rates;
- Customer metrics: for example, recency-frequency-monetary value (RFM), loyalty, profitability, acquisition costs, life-time-value and –size-of-the-wallet;
- Promotional metrics: for example, banner ad click-through, propensity to purchase from click-through;
- Site metrics: for example, page hits, page views, site focus, stickiness, and slipperiness.

The inability of e-businesses to measure their online success is a symptom of the fact that the industry is so new that there isn't a standard of how to measure success. E-metrics are progressively becoming crucial business tools through their increasing alignment with business marketing goals. The incorporation of e-metrics with Technographic, psychographic, navigational and transactional data is now recognised as being essential for businesses that aim to maximise their online endeavours.

5. WEB INTERACTIONAL COMPONENTS

In order for an eCommerce web server to utilise the marketing knowledge (the output of the models), the knowledge must be made available to the server in a machine-readable form, and the web server must be able to act upon the knowledge. These two problems are called *representation* and *deployment*.

5.1. Representing Mined Knowledge

The most suitable vehicle for the representation of marketing knowledge output is based upon XML. An important feature of XML-based representation is that it offers flexibility. It can be used by a web server and also transformed into a form easily understood by people. The Predictive Modelling Mark-up Language (PMML)[3] is a subset of XML, which is being developed by a consortium of data mining vendors, under the aegis of the Data Mining Group. Currently, PMML has representative Document Type Definitions (DTD) for associations, neural networks, clustering, classification, regression and decision trees.

In technical terms, what is required to close the marketing circle is the capture of the click request from a visitor *before* the content/action that has been requested is served back to that visitor. It is only by

[3] http://www.dmg.org/html/pmml_v1_1.html

performing this capture operation that the web server can decide 1) the appropriate PMML model, 2) the correct action to perform, 3) the content to source, 4) the dynamic construction of the page (e.g., banner ads, recommendations, personalisation, etc).

The capture operation can occur as a packet sniffing operation before the visitor request 'hits' the web server. However, most secure eCommerce sites make this a complicated operation, and the work-around (proxies, etc.) can add considerably to the server workload. Ultimately, this can result in reduced performance of the web server, and ultimately visitor disenchantment. One method that ameliorates this problem, is to pass-through the visitor request to the back-end content database (aka reverse proxy configuration), and process all the steps outlined above on either the content database server, or a third personalisation server. In this way, the web server architecture is unstressed, and optimisation of the personalisation server can tune-up the resulting system.

5.2. Deploying Mined Knowledge

Generating models with predictive capabilities is an important objective of web intelligence. One use of such predictive models to generate model output, which is in a form that is also machine-readable by the web servers that host eCommerce sites upon which the analyses and model building have been based. In this situation, the eCommerce sites can produce real-time, dynamically generated pages tailored to individual or group profiles. This is generally called personalisation, and is discussed in the following section. Various commercially available products[4] facilitate the deployment of the results of web mining.

5.3. Personalisation

Personalisation helps web visitors and customers to find individual solutions in their quest for the content or services that they seek within a web site. The power of the Internet as a two-way channel can be utilised by both the financial service provider and the end user. In terms of the fast emerging area of Customer Relationship Management (CRM), personalisation enables e-business providers to implement strategies to lock-in existing customers, and to win new customers.

Initial attempts at achieving personalisation on the Internet have been limited to *check-box* personalisation, where portals allow the user to select the links that they would like to be persistent on the selected page. However, this has limited use as it depends on the users knowing *beforehand* the content that is of interest to them. Arguably, collaborative filtering was the first attempt at using AI for achieving personalisation in a more "intelligent" manner. This allows users to take advantage of other users behavioural activities based on a measure of similarity between them. These techniques require the user to divulge some personal information on their interests, likes and dislikes, information that many web users would not necessarily wish to divulge.

In February 1999, Gartner Analytics[5] stated that, 'matching direct or inferred reader requests through content personalisation will be the most dramatic development in the Internet (…) through 2002, and will help differentiate the Web as a new medium'. A thorough discussion on personalisation can be found in Mulvenna *et al*., 2000.

[4] http://www.mineit.com, http://www.netperceptions.com, http://www.mannainc.com, http://www.atg.com

[5] www.gartner.com

6. DISCUSSION

Table 1 shows the cost per transaction across several channels in the financial industry. This table highlights the cost of direct or semi-automated human response (branch platform / email reply / call centre / branch teller). Contrast this with the cost for automated response (voice-response / ATM / Internet). Clearly, for financial institutions to retain existing customers and remain profitable, they need to offer low cost transaction channels for the majority of customer transactions related to commodity products, while retaining channels with direct human response where necessary. The crux of the financial institutions' problem is that the majority of customer prefer and value the 'human touch' and might decide at any moment in time to use a specific channel. Consistency of information, offerings and pricing strategies across channels are therefore needed. In order to migrate the majority of commodity transactions to low cost channels, significant value must be added to these channels or price might be used to influence the transactional behaviour of customers and prospects.

Financial industry channel	*Cost per transaction ($)*
Branch platform	5.30
Email reply	4.78
Call centre	2.12
Branch teller	1.56
Voice-response unit	0.32
ATM	0.26
Internet	0.09

Table 1 Transaction channels: High-cost versus low-cost (Source: IBM Consulting Group)

Web intelligence provides some of the facilities required to add value in Internet based e-finance. In particular, its appropriate use provides customers with what they need, in terms of easily found information and services ("content"). A recent McKinsey report stated that 24% of the 6 million current users are "completely dissatisfied" with their experience on web sites (Figure 3) (Butler *et al*., 1997).

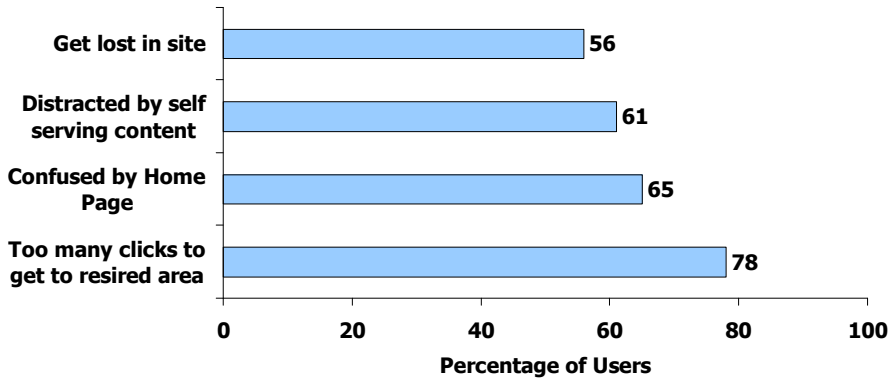

Figure 3 Common Complaints with e-business web sites

The use of web intelligence provide the financial industry with e-metrics based upon the behaviour of web site visitors and customers that provides direct, measurable marketing knowledge, which provides invaluable direct response information upon which financial services can be constantly fine-tuned or should be taken out of the value proposition all together. Information about preferred interaction channel can also be part of that acquired knowledge.

Perhaps the most important goal in digital marketing is to remain focused on a holistic approach that takes into account all the touch points through which customers interact with a financial institution (Figure 4).

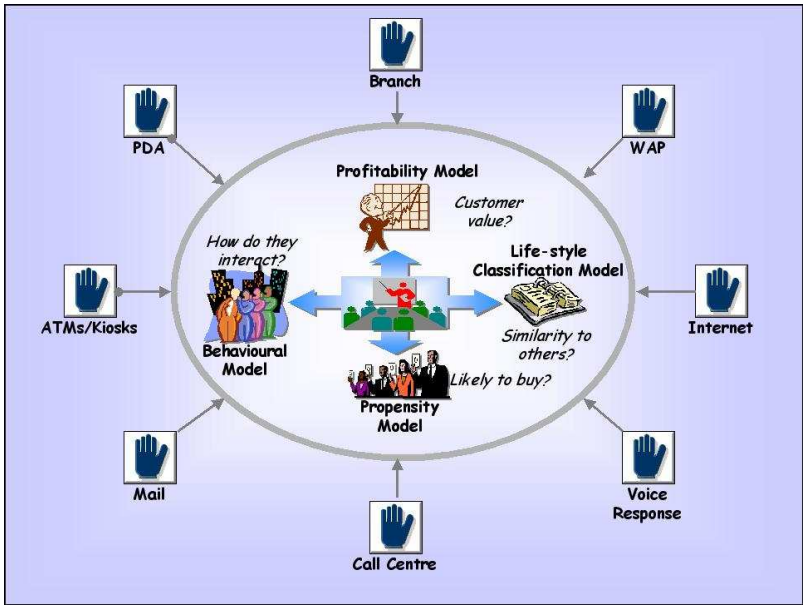

Figure 4 Information Model - modified from (Wells &Wolfers, 2000)

This holistic approach makes several assumptions, some technical and some business related. Firstly, the demands on the ICT infrastructure can be intimidating, both in terms of capital and recurring investment required, and in terms of sheer scale of development effort. Secondly, this approach places the greatest demands on the organisation, in terms of the re-engineering and other perhaps even more drastic organisational changes required in the business. In any case, the internal costs that any financial institution is making in order to supply products and services will have to be reduced at least to the level of costs enjoyed by the new players. The incumbents will not be able to continue to let the customers pay for their legacy organisation or their legacy ICT infrastructure.

In the near future, development in the financial services arena will concentrate on expanding the capabilities of web intelligence to manage digital market interactions through other digital 'touch-points', including WAP telephony, PDAs, and digital interactive TV while also using the technology to optimise internal customer-focused transactions and interactions. Because of the emerging Internet-technology of which web intelligence is a powerful example, the new organisation will need much less co-ordination and management layers than before. All non-customer value-creating activities are perceived by organisational economists as being non-value added work, to be suppressed as quick as possible.

7. ACKNOWLEDGEMENTS

Portions of this research have been funded through grant support from RADIANNE's I-banking, EU MIMIC (No. 26749), EU CERENA (No. IST-1999-10039), and UK/EPSRC NetMODEL (No. GR/N02986). Other sources cover the research carried within the context of the PhD thesis of Tony de Bree to be published fall 2000 (de Bree, 2000)

8. REFERENCES

Baumgarten, M., Büchner, A.G., Anand, S.S., Mulvenna, M.D., Hughes, J.G., (2000): "Navigation Pattern Discovery from Internet Data", In: Masand, B., Spiliopoulou, M. (eds.) *Advances in Web Usage Analysis and User Profiling*. Lecturer Notes in Computer Science, Springer-Verlag.

de Bree, T., (2000) Personal communications on unpublished PhD thesis, "The transformation of Financial Services companies in the Global Knowledge Economy. How to map and measure customer value created with relationship capital".

Büchner, A.G., Mulvenna, M.D. (1998): "Discovering Internet Marketing Intelligence through Online Analytical Web Usage Mining", *ACM SIGMOD Record*, ISSN 0163-5808, 27(4): 54-61.

Butler, P., Hall, T.W., Hanna, A.M., Mendonca, L., Auguste, B, Manyika, J.,
Sahay, A. (1997), "A revolution in Interaction", *McKinsey Quarterly, Number 1:* 4-23.

Coase, R (1937): "The Nature of the Firm", *Economia:* 386-405.

Mulvenna, M.D., Anand, S.S., & Büchner, A.G., (eds.), (2000): "Personalization on the Net using Web Mining", *Communications of the ACM*, ISSN 0001-0782, 43(8): 30-34.

Mulvenna, M.D., Norwood, M.T. & Büchner, A.G. (1998): "Data-driven Marketing", *Int'l Journal of Electronic Commerce and Business Media*, 8(3): 32-35.

Schmitt, E., Manning, H., Paul, Y., Tong. J., (1999): "Measuring Web Success", Forrester Report.

Timmers, P., (1998): "Business Models for Electronic Markets", *International Journal of Electronic Markets*, 8(2), http://www.electronicmarkets.com

Wells, N. & Wolfers, J., (2000): "Finance with a Personalized Touch", *Communications of the ACM*, ISSN 0001-0782, 43(8): 30-34.