



## Proceeding of the International Conference on Advanced Science, Engineering and Information Technology 2011

Hotel Equatorial Bangi-Putrajaya, Malaysia, 14 - 15 January 2011

ISBN 978-983-42366-4-9



# Intelligent Interactive Multimedia by Converging the Intention of Spectator and Multimedia Creator

Seow Hooi Tan, Chee Onn Wong, Choo Yee Ting

<sup>#</sup> Faculty of Information Technology, Multimedia University

Cyberjaya, Selangor, 63100, Malaysia

Tel.: +603-83125268, E-mail: {[sh.tan](mailto:sh.tan@mmu.edu.my), [cowong](mailto:cowong@mmu.edu.my), [cyting](mailto:cyting@mmu.edu.my)}@mmu.edu.my

**Abstract**—In this research, we propose a new approach on how human and technology interact with each other. Here, by enhancing the current HCI framework, it will enable interaction between human and technology become more effective and ideally. The aim of this research is to create an Intelligent Interactive Multimedia by converging the intention of spectator and multimedia creator. Several methods are proposed to achieve the conception of Intelligent Interactive Multimedia. Digital Drawing Block is the interactive multimedia with the initial intention of multimedia creator and it forms an interaction with spectator. Spectator intention has been categorized into four common categories, additionally, five features of hand gesture recognition is proposed to deduce the spectator intention. All these five features will be captured by the web-cam during the spectator's interaction with the Digital Drawing Block. Moreover, captured features will be sent to the machine learning for analyzing. Proposed user models are to assist the machine learning to evaluate the most appropriate category of human behaviour which matches the spectator actual intention. Lastly, graphic that represents spectator intention will be generated together with the initial intention of multimedia creator. The new creation from spectator and multimedia creator will be displayed through the Digital Drawing Block. The conception of Intelligent Interactive Multimedia can represent as 70%'s effort of Multimedia Creator + 30%'s effort of spectator.

**Keywords**— Human Computer Interaction, Interactive Multimedia, Intelligent, User Modelling, Bayesian Network, Human Behaviour

## I. INTRODUCTION

The emerging world has transformed the way human interact with the machine and also the way machine interact with the human. In early time, machine helped to solve the preformulated problem and to process the data according to the predetermined procedures [1]. When time goes by, research has devoted a big contribution into the area of communication or interaction between human and machine. Researchers emphasizes that the way human interact with the machine should be ease and naturalness by using the intrinsic features that available from human for example hands, eyes, fingers and etc [2]. Recent progression of research, the researchers are working toward the direction of human behaviour is engaged into the machine and further extend to the machine is furnished with intelligence and it is able to perform according to the need of human during the interaction [3].

In this research, we presents a digital drawing block which demonstrates an intention of spectator and multimedia creator

is converged, furthermore, to produce a new piece of creation. This piece of new creation is actually a combination effort or intention of spectator and multimedia creator through the intention converging. In addition, this piece of new creation shows spectator and multimedia creator's intention or effort being utilised and appreciated, it is also engender a unique piece of creation from different participation of spectator. Moreover, enjoyment and satisfaction of multimedia creator and spectator is increased and fulfilled, due to, amount of communication exchange between spectator and multimedia creator is adequate.

The aim of this research is to converge the intention of spectator and multimedia creator and produce a new piece of well communicated creation upon the perception of multimedia creator and spectator. Firstly, a Digital Drawing Block will be presented to the spectator with the initiate intention of the multimedia creator. After that, human computing approach is adopted, so that, the spectator can interact with the Digital Drawing Block in effortless, intuitive and naturalness with hand gesture. Adjacent,

computer vision technology is implemented and served a purpose of observer to observe and capture spectator's gesture during the interaction. Consecutively, this capture features from the spectator's gesture proposed in this paper will be speed, movement, location, angle, time and distance. These features used to construct a feature database and then sent to machine learning for recognition. Machine learning technique proposed in this paper is Bayesian Network, it performances a human behaviour recognition based on the feature data gathered from the proposed feature code. Eventually, the machine learning deduces and determines the spectator's actual behaviour consequently classifies it to the rightful human behaviour category, graphic which represented intention of spectator will be generated accordingly. Finally, Digital Drawing Block exhibits the new creation which converge the intention of multimedia creator and spectator.

The research paper is organized as follow. A concise related works is explained in section II, followed by, proposed method and outlines conceptual idea graphically in section III. Section IV is conclusion and future works to conclude research work done in this paper and future enhancement.

## II. RELATED WORK

Contemporary, Interactive multimedia has a great progression in Entertainment E-learning and Gaming, as well as, it becomes more pervasive today. Media revolution had transformed Interactive Multimedia from CD-ROM to Computer Vision Technology and embedded learning agent to enhance the interaction experience of spectator, in additional, linking the technology to real world [4]. Nevertheless, the method and amount of information exchange between multimedia creator and spectator is still limited, which result in interaction and enjoyment of spectator is restricted [5]. Moreover, [6] has pointed critical challenges in media today and future which is obtaining higher priority to be addressed and resolved. They are personalization, media should allow for incredibly efficient way of allocating the desired information based on user preferences and user profile. Followed by, adaptive multimedia content, it creates new approaches to media containers that allow a semantic-based, personalized creation and deliver concept. Lastly, content mobility and accessibility, media should adapt to a form that suits and bring ease the user at a given moment in time and independent of locations and devices.

Human Computer Interaction is widely recognized as an interdisciplinary field [7]. [8] has stated HCI is a discipline concerned with the design, evaluation and implementation of interactive computer systems for human use, including the study of major development surrounding HCI. HCI is concerned with the joint performance of tasks by human and machines, as well as, the communication's structure between human and machine, importantly, human capabilities to use machines (including the learn ability of interfaces). HCI is constantly defining new modalities of communication, and new ways of interacting with machines [9]. [10] has acknowledged that gestures are a most naturalness and user-friendly interaction to convey information for which other modalities are not efficient or suitable. It can be used as a single modality, or combined in multimodal interaction schemes.

In order to converge the intention of multimedia creator and spectator, feature extraction from hand gesture of spectator is used. Hand gesture feature's code is studied in this research to discover the best fit feature's code for hand gesture. [11] proposed combined features of location, angle and velocity as hand gesture recognition. There are several methods of representing such hand's gesture trajectory feature: Cartesian velocity, and the polar and angular velocities proposed by [12], the two-dimensional edge and the time edge proposed by [13], and the raw position. Besides that, according to [11], there are many feature types such as the chain code, mesh code, momentum, and the MRF, nevertheless, all these features are based on only three basic features from a gesture trajectory which is location, orientation and velocity. [14] proposed feature's code which is able to reveal the meaning behind of the hand gesture is size, gesture timing, point of articulation and locus.

User modeling contains the system's assumptions about all aspects of the user which are considered relevant for tailoring the dialog behavior of the system to user [15]. [16] has claimed that the more user models have accommodated, only then the need of different kind of users are able to be fulfilled. In this related work, the author is comprehensible elucidate and elaborate the limitations and challenges sustain in today Interactive Multimedia. This research is focus in the method or ways to produce a more personalize interactive multimedia and to ensure sufficient communication is transmitted between spectator and multimedia creator. Moreover, concise introduction of Human Computer Interaction and which is the better way of interaction between human and machine. By the following, feature's code studies to determine the best suitable feature's code for hand gesture. As well as, further illustration of feature's code proposed and algorithm will be obtainable in Proposed Method. Lastly, different kind of user modelling will be accommodated in this paper comprehensively. Eventually, Bayesian Network will be used to evaluate the user modelling and delivery the final outcome which is the most similar and match to the spectator actual intention.

## III. PROPOSED METHOD.

### A. Human Computer Interaction

In this section, the author is demonstrate the enhancement of Human Computer Framework in fig. 2, it is referred to the current HCI Framework proposed by [17] in fig.1 A Framework of broad HCI issues and concerns by implementing an intention converging stage to combine the Technology and Human in Human Computer Interaction framework. Human Computer Interaction has rapidly growth in past decade, fig. 1 shows that broad HCI issues and concerns. According to [17], human's behaviour or intention must be accommodated during the system development, only then, the system is adaptive to the need of human. However, in the emerging world today, interaction between human and technology no longer design to adapt to any party, but is to establish a communication bridge between technology and spectator. In fig.2, communication bridge is establish in intention converging stage to discover the intention of both human and technology, eventually, produce a new intelligent creation. New intelligent creation is contains well communicated information between technology and human,

mutual understanding formed, effort of human and technology creator is acknowledged. As well as, ideally output (eg. application, interactive multimedia, system) generated which fulfilled desire of Human and Technology. Fig.3 is briefly described about the high-level process in Intention Converging Stage.

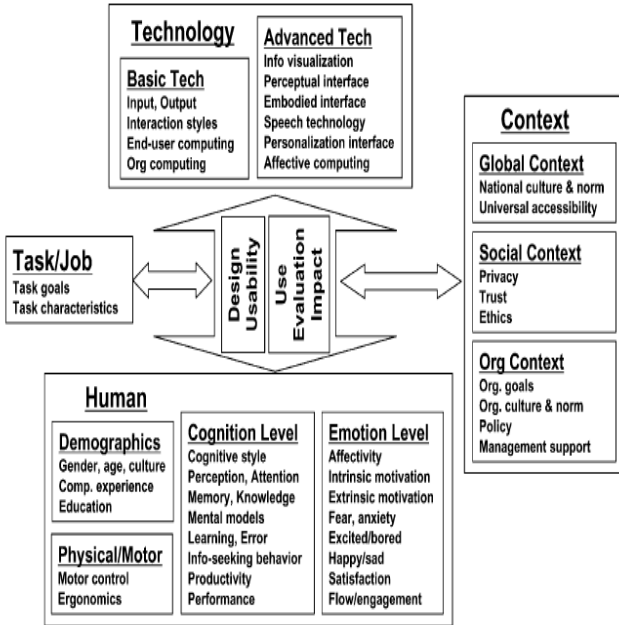


Fig. 1 A Framework of broad HCI issues and concerns

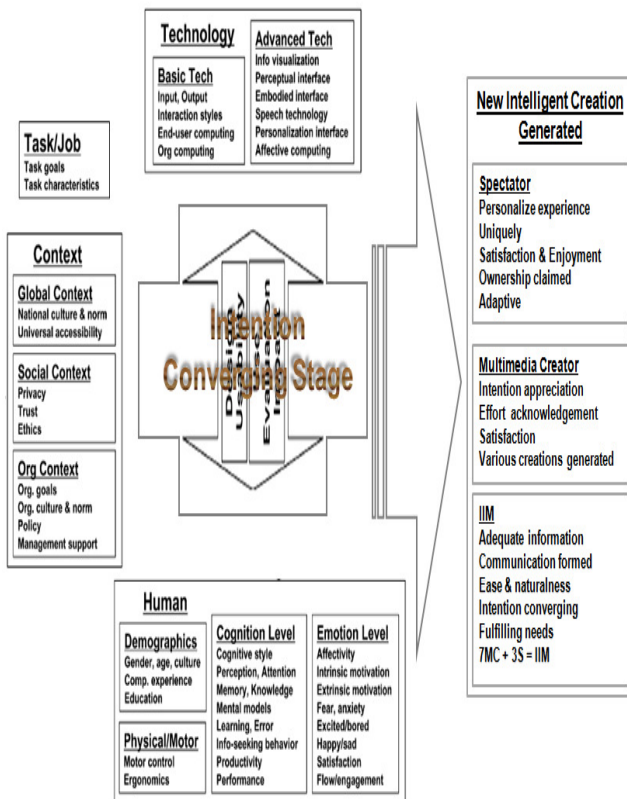


Fig. 2 A Framework of Converging Multimedia Creator's intention and Spectator's intention in HCI

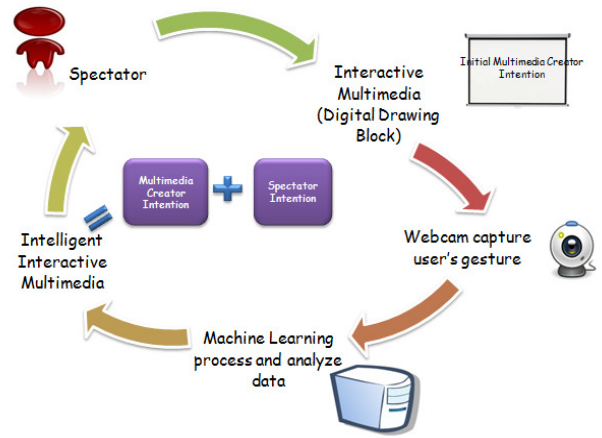


Fig.3 High –Level process of Intention Converging Stage

### B. Human Behaviour Classification

Numerous researches have been carried out in Human behaviour field in past decade. Nevertheless, there is not a standard or discipline to tell or conclude how human usually behave. Human behaviour is usually erratic and uncertain. Sometime, humans behave differently because of his or her goals have changed. In order to solve this problem, we proposed a common classification of Human Behaviour as shown in Table 1. This common classification of human behavior is proposed to give a generic perception to Interactive Multimedia that human will behave. Furthermore, this Human Behavior Category will be used to draw or generate a graphic which represented intention of spectator.

TABLE I  
HUMAN BEHAVIOUR CATEGORY

Human Behaviour Category				
HB	Active	Moderate	Inactive	No Action

### C. Feature Code Proposed

The features are gathered through the hand gesture recognition. The accuracy of result returned of human behaviour is very depending on how appropriate the feature is being defined to deduce meaning behind of the hand gesture. There are five features have proposed Movement, Speed, Location, Distance and Time. In this research, the author claimed that movement and speed will be the critical or important features to determine the rightful intention of spectator.

1) *Distance*: To measure the distance from two successive points. Euclidean distance is proposed to calculate the distance as shown in 1.

$$distance(x, y) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (1)$$

2) *Time*: To find out total time consumes during the interaction and it will reflect the activeness of a spectator during his/her interaction in front of the interactive multimedia. Equation of getting time in milliseconds shown in equation 2.

$$t_{ms} = t_{end} - t_{start} \quad (2)$$

3) *Movement*: The essential step in hand motion tracking is to determine the motion of hand gesture from one frame to the next frame in sequence. This research deploys haar cascade for hand detection and hand tracking algorithm will be based on the hand's different sum pixel from one frame to next frame as shown in 3 and find out the total of movement in 5s as shown in 4, 5, and 6.

$$P(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y') \quad (3)$$

$$\text{When } P(x_1, y_1) \neq P(x_2, y_2), \quad (4)$$

$$\therefore \text{movement} = 1, \quad (5)$$

$$TM = 1 + \dots + n \quad (6)$$

4) *Speed*: Speed feature is based on the fact that each gesture is made at different speeds. For example, a simple gesture, for moving hand right to left and repeating generate almost non-varying speed but for the moving hand from right to up then to down then right will have varying speeds during the hand motion. The biggest size of speed, the fastest it is during the interaction. Speed is measure by two successive point divided by time in equation 7.

$$S_{perM} = \frac{\text{Distance}}{\text{Time}} \quad (7)$$

5) *Location*: To identify position of hand motion during the interaction. Equation 8 is used to calculate angle of the position of hand motion start from the centre point  $(X_t, Y_t)$ , the biggest size of angle in degree, the most activeness of hand motion is moving from one point to another point as shown in Fig.4.

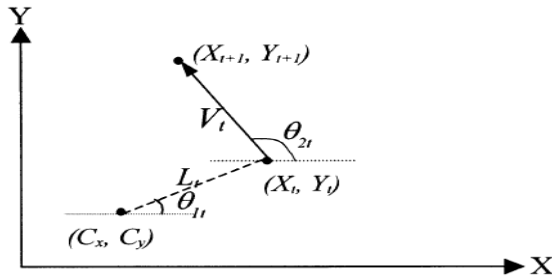


Fig. 4 Feature between two points of gesture sequence [11]

$$\theta_{2t} = \tan^{-1} \left( \frac{Y_{t+1} - Y_t}{X_{t+1} - X_t} \right) \quad (8)$$

6) *Summarization of two core features versus Human Behaviour*: In fig. 5 shows human behaviour corresponds with two core features gathered from hand detection. The highest percentage of speed and movement results active's human behaviour, while the percentage of speed and movement falls in between 0 – 50% will result no action performed from the spectator in front of the digital drawing block and so on so forth.

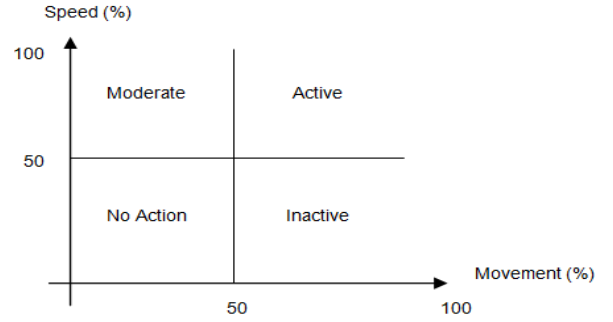


Fig. 5 Human Behaviour corresponds with Features.

#### D. User Modelling Evaluated by Bayesian Network

Three generic user models are proposed for user modelling. First proposed user model is an extreme structure as shown in Table II. Second proposed user model is a mixture structure as shown in Table III. Third proposed user model is an error structure as shown in Table IV. Certainly, there will be more and more user model to be proposed and considered in development phase. These three generic user models will serve as a strategy to help Bayesian Network to determine what the most accurate Human Behaviour to be returned is. Bayesian Network will perform the task of analyzing and determining based of the strategy received and to discover the most accurate Human Behaviour which match to the spectator actual intention.

TABLE II  
USER MODEL 1

Feature	Movement	Speed	Location	Distance	Time
Human Behaviour					
Active	Fast	Fast	Biggest	Biggest	Short
Moderate	Moderate	Moderate	Moderate	Moderate	Moderate
Inactive	Slow	Slow	Smallest	Smallest	Longest
No Action	NA	NA	NA	NA	NA

TABLE III  
USER MODEL 2

Feature	Movement	Speed	Location	Distance	Time
Human Behaviour					
Active	Fast	Moderate	Biggest	Biggest	Moderate
Moderate	Slow	Fast	Biggest	Moderate	Longest
Inactive	Moderate	Slow	Smallest	Moderate	Longest
No Action	NA	Slow	NA	NA	Slow

TABLE IV  
USER MODEL 3

Feature	Movement	Speed	Location	Distance	Time
Human Behaviour					
Active	Slow	Slow	Smallest	Smallest	Longest
Moderate	Fast	Fast	Biggest	Biggest	Short
Inactive	Moderate	Moderate	Moderate	Moderate	Moderate
No Action	Slow	Fast	Biggest	Moderate	Longest

#### IV. CONCLUSIONS

In this research, we presented the momentous enhancement of current HCI framework by implementing the intention converging stage. Several methods are proposed to achieve intelligent interactive multimedia by converging the intention of spectator and multimedia creator. There is an enhancement to the framework of Human Computer Interaction, Human Behaviour's Category Classification, and Feature's code proposed to deduce meaning of the hand gesture. Bayesian Network is used to evaluate and determine the spectator's behaviour according to the User Modelling proposed. Moreover, this momentous enhancement has brought the way human interact with the machine into new milestone and it become more effective in HCI. Furthermore, this research is believed to have a significant contribution in entertainment, gaming and e-learning field. Current growth of e-learning, instructors or tutor (technology) is replaced by the multimedia (studying material) and student (user) is able to do distance learning. Student's perception always is the research focus area to deliver an adaptive multimedia which improves the student learning ability. Nevertheless, the intention or thought of instructors or tutor (technology) to convey knowledge always be neglected. Once the intention converging stage is adopted in e-learning, mutual communication is established between technology and human, only then, an effective and ideally multimedia (studying material) will be delivered. The author strongly deem that technology and human should TALK, instead of ADAPT. Along with incredible way that human interact with machine proposed in this research will become the future trends in both HCI and Interactive Multimedia. Future work of this research is experimental works should be carried out based on the several methods proposed in this paper, so that, conception of Intelligent Interactive Multimedia will become more substantial.

#### NOMENCLATURE

##### Subscripts

HCI	Human Computer Interaction
IIM	Intelligent Interactive Multimedia
MRF	Markov Random Field
TM	Total Movement
$S_{perM}$	Speed per Movement
C	Current

#### ACKNOWLEDGMENT

First and foremost, it is a pleasure to thank all subjects that has contributed to the data collection. The author would like to offer regards and thanks to all of those who had supported in the completion of conference paper especially family members and research centre in MMU.

#### REFERENCES

- [1] J.C.R.Licklider, "Man-Computer Symbiosis", IRE Transactions on Human Factors in Electronics 1960, pp.4-11.
- [2] Vladimir I.Pavlovic, Rajeev Sharma, Thomas S.Huang, "Visual Interpretation of Hand Gestures for Human-Computer Interaction", IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 19, pp. 667-695, July. 1997
- [3] Ernest Edmonds and Greg Turner, "Approaches to Interactive Art Systems", ACM 2004, pp. 113 -117
- [4] Dorée Duncan Seligman, "Interactive and Informative Art", IEEE Computer Society 2003, pp. 4-10.
- [5] Xinyuan Cai, "Principles of Human-computer Interaction in Game Design", In Proceedings of Second International Symposium on Computational Intelligence and Design IEEE 2009, pp. 92-95.
- [6] Oscar Mayora<sup>1</sup>, Petros Daras<sup>2</sup>, Marianna Panebarco<sup>3</sup>, Nick Achilleopoulos<sup>4</sup>, Peter Stollenmayer<sup>5</sup>, Doug Williams<sup>6</sup>, Nadia Magnenat-Thalmann<sup>7</sup>, Carmen Guerrero<sup>8</sup>, Michiel Pelt<sup>9</sup>, Tim McGrath<sup>10</sup>, Eugenia Fuenmayor<sup>11</sup>, David Salama<sup>12</sup>, Federico Alvarez<sup>13</sup>, Elias Kalapanidas<sup>14</sup>, Alex Shani<sup>15</sup>, Jean-Yves Le Moine<sup>16</sup>, "User Centric Media in the Future Internet: Trends and Challenges", 3rd International Conference on Digital Interactive Media in Entertainment and Arts ACM 2008, pp 441 - 446
- [7] Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., & Carey, T. (1994). Human-computer interaction. Addison-Wesley
- [8] Laurent Mignonneau, Christa Sommerer, "Designing emotional, metaphoric, natural and intuitive interfaces for interactive art, edutainment and mobile communications", In Proceedings of Computers & Graphics 29 2005, pp.837 -850
- [9] Antonio Camurri and Gualtiero Volpe, Eds., Gesture-Based Communication in Human-Computer Interaction, 5th International Gesture Workshop, GW 2003, Genova, Italy, April 15-17, 2003, Selected Revised Papers, vol.2915 of Lecture Notes in Computer Science, Springer, 2004
- [10] George Caridakis, Kostas Karpouzis, Nasos Drosopoulos and Stefanos Kollias, "ADAPTIVE GESTURE RECOGNITION IN HUMAN COMPUTER INTERACTION", IEEE 2009, pp. 270-274.
- [11] Ho-Sub Yoon, Jung Soh, Younglae J. Bae, Hyun Seung Yang, "Hand Gesture recognition using combined features of location, angle and velocity", Image Processing Div. / Computer & Software Technology Lab., ETRI 161, Kajung-Dong, Yusung-Ku, Taejon, 305-350, South Korea, Department of Computer Science / Korea Advanced Institute of Science and Technology, 373-1 Kusung-Dong, Yusung-Ku, Taejon, 305-701, South Korea, 2000, pp.1492-1501.
- [12] S. Seki, K. Takahashi, R. Oka, "Gesture recognition from motion images by spotting algorithm", In Proceedings of ACCV 1993, pp. 759-762.
- [13] L.W. Campbell, D.A. Becker, A. Azarbayejani, A.F. Bobick, "A. Pentland, Invariant features for 3-D gesture recognition", In Proceedings of the International Workshop on Automatic Face-and Gesture-Recognition 1996, pp. 157-162.
- [14] Nicola Rossini. "The Analysis of Gesture: Establishing a Set of Parameters", In Proceedings of Springer-Verlag Berlin Heidelberg 2004, pp.124-131.
- [15] Alfred Kobsa, "User Modeling and User-Adapted Interaction", In proceeding of ACM 1994, pp. 415-416.
- [16] Plinio Thomaz Aquino Junior, Lucia Vilela Leite Filgueiras, "User Modeling with Personas", In proceedings of CLIHC 2005, pp.277-282.
- [17] Ping Zhang, Na Li, "An assessment of human-computer interaction research in management information systems: topics and methods", School of Information Studies, Syracuse University, Syracuse, NY 13244, USA, 2005. pp.125-147