Introducing the Email Knowledge Extraction with Social Network Analysis (EKESNA) tool for discovering an organisation's expertise network

Introducing the Email Knowledge Extraction with Social Network Analysis (EKESNA) tool, for discovering an organisation's expertise network

Ejovwoke Onojeharho[1], Thomas W. Jackson[1], Louise Cooke[1]

[1]Centre for Information Management, School of Business and Economics,
Loughborough University.
Epinal Way, Loughborough, Leicestershire, LE11 3TU, UK
E.Onojeharho@lboro.ac.uk
T.W.Jackson@lboro.ac.uk
L.Cooke@lboro.ac.uk

Abstract

Manually collating social network analysis (SNA) data can be a tedious and time consuming process, but if automated could save time and aid efficiency utilising the organisations knowledge network. In this paper the authors propose Email Knowledge Extraction with Social Network Analysis (EKESNA); a system for automating the continuous discovery and collation of organisation's social network, as well as expertise network. The research adopted a systems development methodology, which comprised four stages. The first reviewed the approaches for collecting SNA data. The second involved carrying out SNA using traditional techniques. In the third stage, the EKESNA tool was developed, piloted in-house, and a trial was run at the same organisation used in stage two. The final stage evaluated the SNA data collected using both approach. The knowledge network obtained from the traditional social network gathering method, and from EKESNA, revealed similarities during the analysis of members of the organisation that were central to the flow of Information and Knowledge. When compared with the traditional method of conducting knowledge network analysis, the EKESNA tool allows for continuous collection as new topics are discussed, and new members are introduced into the organisation. The nature of the tools continuous discovery of the organisation's knowledge network affords members up-to-date data to inform business process reengineering. The data collected is continuously

evolving meaning organisations can integrate networks around core processes, ensure integration (post-merger or reorganisation), improve the strategic decision making in top leadership networks (recruiting the right people for a particular project based on expertise and connectivity), or even identify/facilitate potential communities of practice. Analysis was limited by the sample size of both collection methods, emphasis was therefore made using centrality measures. Email knowledge extraction, and email social network systems are not new concepts, however this paper presents EKESNA which is a novel system as it combines both concepts in a way that also allows for the continuous discovery, visualisation, and analysis of networks around specified topics of interest within an organisation; linking conversations to specific expert knowledge.

1.0 The Need to Connect People to People

There are several justifications within literature for organisations increasingly focusing efforts to help employees efficiently identify expertise. Qu & Pau cited in [1, p.336] highlight that "finding out who has the right expertise to solve critical business problems is harder than ever", O'Dell & Hubert [2, p.57] say that efforts placed into identifying key experts is justified, "because of their ability to answer questions, provide historical perspective, and offer solutions". [3] Presented findings of a study on a large organisation, which showed 67% of employees were aware that information which they required was within the organisation, but 39% did not know how to access it. The author pointed out that a majority of the study group highlighted that this inability to locate expertise, resulted in duplication of work and wrong decision making, owing to inaccurate information.

Tedmori [4] highlights that the literature would suggest "people searching for information prefer to consult other people, rather than to use on-line or off-line manuals". However, the author explains that regardless of the organisation's size, identifying where knowledge lies, is a shared problem, given that its information assets are continuously growing in number and intricacy, as information being generated rapidly increases. Ehrlich cited in [5] agrees adding that while individuals within some organisations remain in the same job for long periods and develop key contact networks, for a majority of organisations, due to factors such as mergers and acquisitions, downsizing, outsourcing, or reshuffling, people do not remain in the same job for long. The issues organisations face coupled with the statistics presented from the employee survey, provide a strong argument for the need to connect people to people, along with the other potential benefits. [6] highlight some benefits which include:

- “Provision of solutions to problems
- Provision of answers to questions
- Provision of pointers to others that might know the answer; Exploiting Email: Extracting Knowledge to Support Knowledge Sharing
- Provision of pointers to other useful sources
- engagement in interaction that helps shape the dimension of the problem space
- Psychological benefits (e.g. confidence, assurance)
- social benefits or legitimation of decisions
- Improvement in the effectiveness with which a person advances their knowledge in new and often diverse social contexts
- Improvement in efficiency (e.g. reduction in time wasted pursuing other avenues)”

For this research the author identifies two tools which would aid organisations employees in finding the right people to support various tasks. Social network analysis is a tool organisations can use to access and resolve anomalies in it connections, while expert locator tools provide a means for the organisation to identify and what knowledge exist, and where that knowledge resides.

2.0 Social Network Analysis

Social network analysis (SNA) helps to identify relationships between individuals in an organisation and the level of communication within the relations. [7, p.329] explains that it “provides knowledge management practitioners with a new way to look at knowledge creation and transfer, to understand the nature of connections, to visualize them, and to analyze them for both tactical and strategic change”. Figure 1 presents a typical SNA diagram; the (Arrows) are used to represent links between individuals (Nodes). [7, 8] highlight that SNA is generally used in business organisations for the following reason:

- To identify teams and individuals playing different roles in the organisation, thought leaders; people who support the group in ways that often go unrecognised.
- Identify isolated teams or individuals (Peripherals; outsiders whose skills and expertise are often not leveraged effectively).
- Spot opportunities for connecting subgroups (Boundary Spanners; those who provide central links between two groups).
- Target opportunities where increased or improved knowledge flow will have the most impact (Bottlenecks).
- Raise awareness of the importance of informal networks.

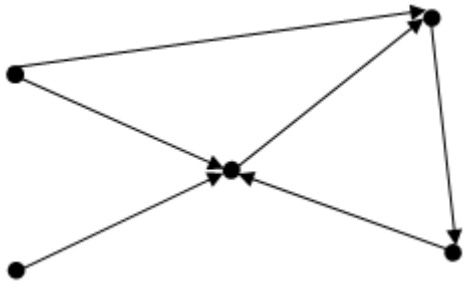

Figure 1: Information Flow Network

After consulting SNA experts, [7, p.331] reflects that "the most important fact about SNA is that it does not provide absolute values. The intent of an analysis is not to pass judgment on groups or individuals; the intent is to understand the patterns of relationships, using both quantitative and qualitative views, in order to decide how to make improvements". For the current study, the author suggest that the results could be used to perform analysis, facilitate discussion of the emerging patterns, and suggest initiatives to improve knowledge flow to support the business goals.

[7, 8] both provide steps for creating a SNA, the former provides steps for creating a general SNA study while the later focuses on steps for creating SNA study within an organisation, for the purpose of knowledge management. Table 1 presents both lists:

[8, p.144-162]	[7, p.334-335]
Identify a strategically important group	Determine the business goal for the SNA.
Assess meaningful and actionable relationships	Collect data about the knowledge and information flow patterns in an organization
Visually analyse the results	Use computer tools to create a network map and statistical measures from the data
Quantitatively analyse the results	Scan the results to look for gaps, or junctures, between individuals or groups.
Create meaningful feedback sessions	Use consultative interviews to understand the context that is behind the data and the diagnostics
Access progress and effectiveness	Present the results to the managers/sponsors and to the group that has been surveyed
	Target areas where insufficient knowledge flow has a serious impact on

	the business
	Design organizational interventions to create the environment that will enable social capital to grow

Table 1: Steps for creating a SNA. Adapted from [7, 8]

[7] Stresses that it is important to have a participation rate as close to 100% as possible, for the study group to be deemed successful. Both authors suggest that in order to carry out a SNA, researchers require software such as UCINET, Cryam NetMiner, RepTools, antology™, InFlow™, or IKNOW. They both favour UCINET for carrying out tests on the data, pairing it with NetDraw; a tool for creating graphical representations of the network. A case study conducted to look at "knowledge flow among the departments, and that among the individual employees,"[9, p.27] used a software package called NetMiner to create a SNA, within a leading organisation in Washington DC that consisted of 85 employees; the study achieved 30% participation, but still provided some key findings, such as: who the top experts were for obtaining general advice, management and leadership advice, institutional knowledge, subject matter expertise and technical or procedure knowledge.

The above seem to focus solely on the connections between individuals with an organisation, depending on the basis with which the process is conducted; the results are dependent on the content of the investigation material, which is agreed upon for that particular time and purpose. For the current study the focus was on the flow of expertise knowledge within an organisation.

3.0 Expert Locator

According to [4], traditional means of providing expert assistance was through the creation of expert databases such as knowledge directories, the process for creating these systems was deemed inefficient, as the information required needs to be entered manually by the users. This meant that these systems took a long time to set up; which meant high level costs involved with regards to financial and time resources; the systems were also difficult to maintain and update, meaning the information contained was constantly incomplete or out-dated. The author recalls how the "emergence of the World Wide Web led to the development of personal web pages where individual[s] provide information about their interest/expertise areas" (p.11). This however increased the cost of updating individual pages and also brought up a new problem of ineffective searching; because the information provided by users may not match the search terms supplied, causing it to return a large number of hits, increasing the time required to retrieve the appropriate information. [4] concludes that these shortcomings along with developments in

technology ushered in developments into automated/semi-automated expert finding systems.

Ehrlich cited in [5] and [4] describes expert locators as web based tools which connect people to people, by developing awareness of who knows what across the organisation. The former adds that these systems help to "build social capital by strengthening the ties between people who know each other, [as well as] facilitating conversations between people who do not know each other" [5, p.155]. The underpinning idea is to capture and catalogue information about an individual's expertise, and make it available for others to search, and effectively direct their queries. Several systems are highlighted by [4] including: HelpNet, ContactFinder, Yenta, InfoScout, MIT's Expert Finder, and Expertise Browser, to name a few. The author explains that the "the main feature that distinguishes expert finding systems from each other is the information source(s) that they use as the basis for expertise recognition" (p.11). [1] comment on a popular expert locator called Employee Knowledge Network (EKN), used by various large organisations including Intel, Honeywell, Boeing Canoga Park, Procter and Gamble, and CNA Insurance. The authors explain that the system utilises a Q&A approach for identifying experts with solutions to certain topics, users can then enter their questions, to which the system would return relevant stored content, and if EKN did not come back with an answer, it "then searches through user profiles to find someone with relevant expertise" (p.369).

Email systems are very popular tools of communication in organisations and provide a useful source for capturing information of expertise. Two examples of expert locators which exploit emails, found in literature are; Abuzz's BeeHive system and the Email Knowledge Extraction (EKE) system. The former uses web browsers and standard email to track user profiles and store the in-house expertise discovered, while the latter processes user emails semantically, and produces Keywords or key phrases which may hold knowledge expertise value. This research makes use of the latter because of its availability, capacity for further customisation, record topping f-measure output, and because it needs no pre-defined taxonomy.

4.0 Research Design and System Development

The purpose of conducting SNA at the organisation was to help identify the key players, and enlist them for the introduction of new knowledge management tool. The case study organisation is a National Sporting Organisation (NSO), the first Governing Body which has been in existence since 1869. According to [7] surveys, ethnographic interviews, and electronic activity mapping, can be used to capture the data required for creating SNA maps. Questionnaires were used to collect data within the case study organisation, while this "method is not practical for groups

larger than 150 to 200, the results within this range provide broad coverage and actionable results within groups up to this size" [7, p.336]. [8] propose two approaches to social network analysis; snowball and full network. The snowball approach is best suited when there is a natural starting point [10], as it works by asking the pre-determined individuals who they contact, and subsequently asking the individuals on the list obtained who they are connected to as well. The full network approach involves collecting data from the whole population at once. It was preferred for this research because it provides quickly a representation of the organisation, and it can be a conducted with ease if the process is properly managed by asking respondents to identify a limited number of specific individuals [10]. The case study organisation had a little over 350 members; of this, 245 were permanent staff, and the others listed as volunteering members. A decision was made by the authors to focus only on permanent staff, and the questionnaire constituting questions requiring information on who participants contacted for information regarding various topics (3 names for each), was administered. The initial survey lasted a month, with an additional 2 weeks added to allow the authors to chase responses for the completion of whole departments. The questionnaire response rate was calculated using the following formula;

$$Response\ Rate = \frac{\sum completed\ Responses}{\sum Selected\ Sample} * 100 \qquad \frac{174}{245} * 100 = 71.02\%$$

Collating the survey data for analysis spanned over 2 weeks as it had to be done manually. The authors quickly identified the need for a system which could collect and compile the social network data automatically, when required for future projects.

5.0 EKESNA: a System for Discovering Expertise Flow Network from Emails

This section outlines the approach developed by the authors to identify individual competencies from email messages. The EKESNA system, aims to automate the continuous discovery and collation of an organisation's contact network, as well as its knowledge / expertise network. EKESNA allows for continuous collection as new topics are discussed, and new members are introduced into the organisation. The nature of the tools continuous discovery of the organisation's knowledge network affords members up-to-date data to inform business process reengineering.

Figure 2 represents the system architecture of the EKESNA architecture. The system comprises on the client side, a Microsoft Outlook plug-in, and on the server side, a natural language processing (NLP) engine based on an earlier EKE system detailed in [11]. EKE was chosen for this project for the following reasons: **(i) Quality of Keyword/Key phrase Extraction:** the system reported as having one

of the best F-Measure in terms of precision and recall of key phrases extracted from emails [11]. **(ii) Availability:** The authors were afforded easy access to the core component of the EKE system for further development.

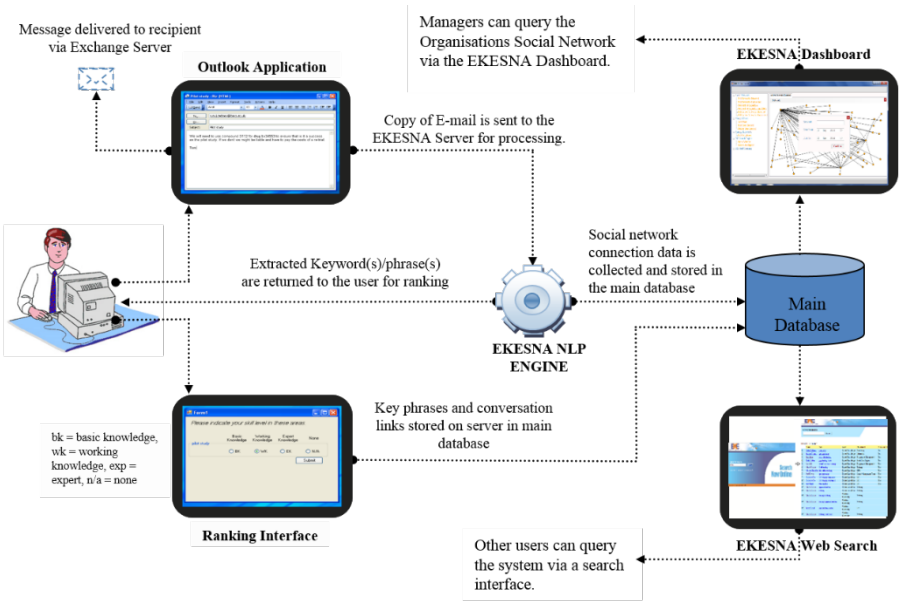

Figure 2: EKESNA Generic Architecture.

5.1 System Overview

Many expertise locator systems acquire data either by having individuals fill out profile information or by extracting information from existing sources (authored documents, blogs, published articles etc.). In deciding which data source to use for EKESNA, the authors dismissed these sources as they relied heavily on the user manually entering and updating the knowledge base. Emails seemed the obvious choice as they constitute rich expertise content and information on the expertise flow, but also because new data can be extracted constantly. As discussed earlier, social network analysis works best when all the nodes are represented; the decision to utilise email as the source supports this due to its coverage. Email use is widespread in many organisation and particularly within the case study organisation, everyone can use, and already has access meaning everyone can contribute to the knowledge network. The issue with using emails as a source stems from privacy when sensitive data is captured by the system. The selection of EKE as the base extraction system eases this problem; after considering the relevant literature around privacy issues, [12, p.113] clarified that the system "addresses this sensitive area by enabling employees to select the keyphrases they would like to share with the rest of the organisation".

Figure 2 illustrates the workflow of the system. The user creates and sends an email following a regular procedure, this triggers the EKESNA process which whilst performing the send query, a duplicate of the message containing the body and correspondents is directed to the server for processing. The following processes occur: (i) the social links between correspondents are extracted and stored on the main database. (ii) Using NLP techniques, the relevant keywords/phrases are extracted and returned to the user, who then ranks their level of expertise appropriately. Once the user completes the ranking, the keywords/phrases are stored in the main database, along with the data relating each topic to a particular conversation, ensuring capture of the expertise flow between nodes. The repository is available to be queried at a later period by all users within the organisation searching for experts on a particular topic, via a web search interface. The social network and expertise flow network is made available to authorised users via a desktop dashboard interface; from this users can create overall network data or the networks around particular topics of interest.

5.2 System Pilot

For this research, the authors followed tips documented by [13] for how they successfully implemented social software within their research organisation. The pilot enlisted 40 members of staff highlighted by the organisation as the 'super user' (early adopters and go-to persons within their department/team); of this; 38 turned up for initial briefing details of the project and attended full training. From this group a further 18 members dropped out, as they were unable to commit fully to the project. The final sample constituted a small group of 20 staff members, with at least 1 member representing each department. The authors were satisfied with this sample; it adheres to the first two success tips (i) "Start small and work with just a few groups". (ii) "Focus on groups who are enthusiastic and committed" Ward (2006, p.238). The pilot lasted 3 months, and usability feedback from participants was continuously added in newer versions. The pilot recorded over 2,200 distinct network ties between participants and other members of the organisation, and over 5,000 instances of expertise exchange between nodes within these connections.

5.3 Network Analysis

This section describes the functionalities within the tool for generating and visualising the organisation's network. Figure 3 shows the interface a user engages with. The navigation panel on the left side allows users access to different parts of the system; with regards to networks, the following functionalities are available: (i) access to visualise full networks (directed/undirected), (ii) access to generate the network by expertise (flow of a particular topic); this can also be refined with specific dates, for example to track growth of the network over time or to investigate the impact of some change actions on the network, and (ii) access to visualise the active network (directed/undirected); this provides data only on nodes that have at least one tie. The middle panel provide additional operation based

functions; in figure 3 for instance, there is option for the user to export the network as a .DL file which can be used for analysis on UCINET or NetDraw.

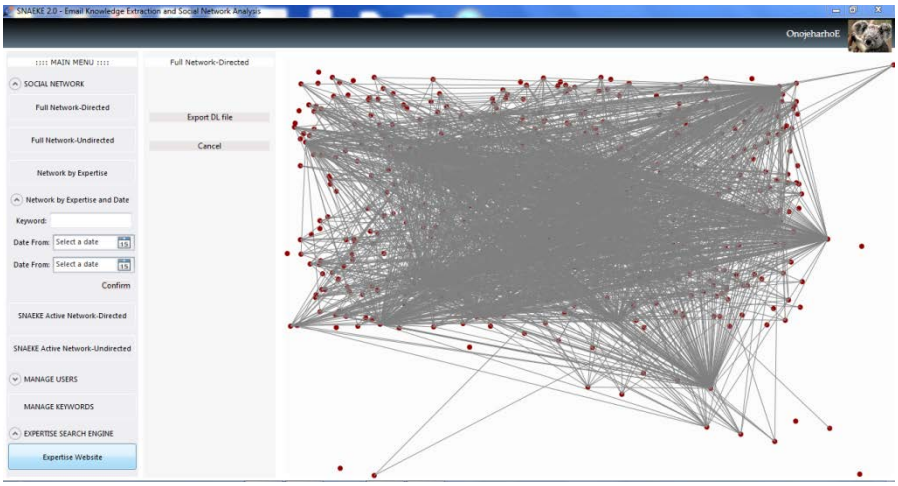

Figure 3: EKESNA dashboard interface showing a genrated full directed network

5.4 Analysing the Full Network of the organisation and the Network Flow of the topic 'PPP Project'

Analysis was limited by the sample size of both collection methods, it therefore is appropriate to focus on centrality measures when comparing both full network data sets (traditional Vs EKESNA generated). The findings identified the same members of staff as having the highest eigenvector and indegree figures; this seems justified as the most popular individual within the organisation should remain the same even within a smaller representative sample. The data collected by the EKESNA system is continuously evolving meaning organisations can better perform business process re-engineering. Full network data allows the organisation to ensure integration (post-merger or reorganisation). Network by expertise allows for improved strategic decision making in top leadership networks (recruiting the right people for a particular project based on expertise and connectivity). Figure 4 shows a network generated around the 'PPP' project at the research organisation using the EKESNA dashboard. Analysis with NetDraw identifies the key player nodes (Green = degree/ eigenvector 'the most influential person'), (Yellow = betweeness 'potential for gatekeeping, brokering, and liaising between otherwise separate parts of the network'), (Red = farness 'last to receive information'). The green node represents a major cutpoint in the network; in the scenario where that individual leaves the organisation depicted in figure 5, managers can take actions to integrate networks around core processes, or even identify/facilitate potential communities of practice, thereby promoting innovation.

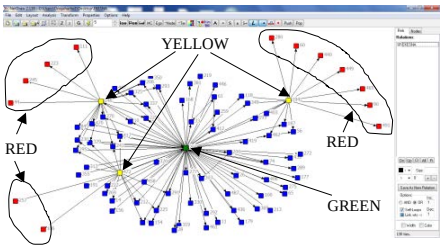

Figure 4: 'PPP' network generated from EKESNA, visualised / analysed on NetDraw

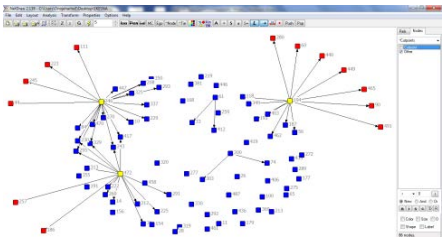

Figure 5: 'PPP' network shows the ability of a single node's potential to distort the topic flow

5.5 Expertise Search

The EKESNA web search engine returns a list of relevant people, ranked by their level of expertise and responsibility regarding queried topics. The engine applies stemming techniques to the search string to allow for a greater recall. Consider the case of someone looking for information about the PPP review process, entering "PPP" into the search window would bring up the display shown in figure 6. It is also interesting to see the expert at the top of the list is the same as the green central node observed in figure 4&5 when analysing the PPP network.

Figure 6: EKESNA expertise web search interface, showing 'PPP' experts at the organisation

6.0 Summary and Future Research

Email knowledge extraction, and email social network systems are not new concepts, however this paper presents EKESNA which is a novel system as it combines both concepts in a way that also allows for the continuous discovery, visualisation, and analysis of networks around specified topics of interest within an organisation; linking conversations to specific expert knowledge. Although analysis was limited due to the sample size of both collection methods, the

centrality measures proved similar. The EKESNA system currently relies on third party software such as NetDraw and UCINET to visualise and analyse the SNA dataset; further development should consider integrating the relevant analysis criteria, to provide actionable reports regarding the knowledge network.

7.0 References

1 Qu L & Pau S, Tools for tapping expertise in large organizations. In: Rao M, Knowledge management tools and techniques: practitioners and experts evaluate KM solution. Elsevier Butterworth-Heinemann, 2005

2 O'Dell C & Hubert C, The new edge in knowledge: how knowledge management is changing the way we do business. John Wiley and Sons, 2011

3 Gilmour D (2003). How to fix knowledge management. Harvard Business Review, 81

4 Tedmori S, Exploiting email: Extracting knowledge to support knowledge sharing. PhD thesis, Loughborough University, 2008

5 Ackerman M S Pipek V & Wulf V, Sharing expertise: beyond knowledge management. London: MIT Press, 2003

6 Cross R Parker A & Rice R E (2001). Information Seeking in Social Context: Structural Influences and Receipt of Information Benefits, IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews. 31, 4, 438-448.

7 Anklam H, Social network analysis in the KM toolkit. In: Rao M, Knowledge management tools and techniques: practitioners and experts evaluate KM solution. Elsevier Butterworth-Heinemann, 2005

8 Cross R L & Parker A, The hidden power of social networks: understanding how work really gets done in organizations. Harvard Business School Press, 2004

9 Chan K & Liebowitz J (2006). The synergy of social network analysis and knowledge mapping: a case study. 7, 19-35.

10 Hanneman R A & Riddle M, Introduction to social network methods. Riverside, CA: University of California, 2005

11 Tedmori S, Jackson TW & Bouchlaghem D (2006). Locating knowledge sources through keyphrase extraction. Knowledge and Process Management, 13, 100-107.

12 Tedmori S, Lichtenstein S, & Jackson TW, Expertise Location Using Keyphrases in Electronic Mail: Socio-ethical Challenges, proceedings of the Australian Conference for Knowledge Management & Intelligent Decision Support (ACKMIDS 2007), Melbourne, Australia, 2007.

13 Ward R (2007). Blogs and wikis: a personal journey, Business Information Review. 23, 235-240

14 Borgatti S P, NetDraw Software for Network Visualization. Analytic Technologies: Lexington, KY, 2002