A Case-based Reasoning (CBR) Internship Placement Model

Emmanuel Isika
Department of Computer Science
The Federal University of Technology Akure, Ondo state, Nigeria

Akintoba Akinwonmi
Department of Computer Science
The Federal University of Technology Akure, Ondo state, Nigeria

Oluyomi Akinyokun
Department of Cyber Security
Federal University of Technology Akure, Ondo state, Nigeria

ABSTRACT
Candidates are faced with numerous challenges when seeking internship especially in IT-based firms, the challenges include elongated time-frame resulting from the conventional search of placement among others. This research presents a platform through the design of a case-based reasoning (CBR) model which mitigates the challenges and facilitates internship placements for candidates. The aim is to alleviate intern-employer mapping dilemma. The research applies supervised machine learning techniques including data pre-processing, feature extraction, document similarity metrics, and knowledge-intensive CBR pattern matching to optimize matching between intern candidate vectors and employer criteria vectors. The system resultantly introduce an ML based personalized and efficient matching platform with real-time support, potentially improving outcomes for interns and companies within the same ecosystem.

General Terms
Machine Learning Case-Based Reasoning,

Keywords
Machine Learning, Internships, Case-Based Reasoning, Natural Language Processing.

1. INTRODUCTION
The predicament of optimally placing interns into relevant immersive industrial training programs continues as an intricate conundrum perplexing educational ecosystems. Historically, manual protocols for intern-employer pairing have proven inadequate, with average placement acquisition durations spanning 2-4 months as interns contend for limited openings. This inability to provide timely placements significantly hinders learning outcomes and workforce transitions. Hence, augmenting the internship placement process via intelligent automation presents a prospective panacea.

This research explores the design and actualization of an artificial intelligence (AI)-based platform to enhance the internship process leveraging case-based reasoning (CBR). CBR constitutes an emergent supervised machine learning methodology which models solutions to new problems based on historical cases. The undertaken approach entails systematic pre-processing of datasets encapsulating intern credentials and employer criteria. Descriptive features are extracted quantifying competencies, interests, and requirements. Domain knowledge is incorporated to construct personalized similarity functions measuring intern-employer suitability. The CBR system applies these patterns to derive aptitude insights, facilitating profile-based recommendations between candidates and openings.

The implemented platform cultivates an ecosystem which ameliorates bottlenecks pervading internship attainment through real-time guidance. Continuous personalized learning pathways cater towards developing in-demand skills. Portfolio building assists candidates in effectively communicating competencies to prospective employers. Profile-based recommendations allow candidates to efficiently discover well-matched openings based on their credentials. For employers, enhanced analytics empower optimized recruitment decisions.

This work explores the AI-based solution specifically tailored to address the intern challenges by innovating data-driven intern-employer matching capabilities. The systematization of organizational placement processes enact more seamless school-to-work transitions within the digital era. Additional contributions include the novel incorporation of CBR within career guidance systems to augment employability.

2. LITERATURE REVIEW
Securing an internship placement is a crucial step for university interns as they prepare to transition from school to the workforce. However, the current internship placement process has been criticized as disorganized, inefficient, and challenging. Recent innovations in artificial intelligence present promising solutions to address the broken internship placement system. This literature review analyzes previous research related to improving internship acquisition processes and harnessing AI to streamline matching between interns and employers.

2.1 Prior Work Improving Internship Placement

Several studies have focused on making improvements to traditional internship placement approaches. [24] provided an insightful historical overview tracing the evolution of internships from ancient apprenticeship programs to the modern intern economy. This understanding of the foundations underpins efforts to enhance current practices.

Researchers have also explored supplementing universities' career assistance resources. [25] developed the "Map My Career" tool that helps match interns to potential careers based on skills, interests, and personality. However, the reliance on self-reported data can undermine matching accuracy. [26] created the "Course Map" system aligning academic trajectories with career goals, but its utility is limited for undecided interns.

2.2 Applying Artificial Intelligence

Recent research has investigated AI techniques to improve efficiency and personalization in recruitment. [18]

demonstrated that semantic embeddings and machine learning can effectively match candidates to relevant job openings. However, their approach focused heavily on formal credentials rather than practical skills.

Other studies have shown the promise of personalized recommendations and data-driven insights for the hiring process. [3] increased screening efficiency by deferring classification until necessary using Lazy Learning. [22] developed a scalable graph-based algorithm to rank job suggestions. However, both methods are susceptible to biases emerging from the underlying data.

2.3 Research Gap

While prior literature has provided frameworks to enhance career guidance and harness AI for recruitment, minimal research has explored case-based reasoning for personally matching interns to internships based on skills and experience. This approach can revolutionize pairing interns to opportunities where they can thrive professionally.

Previous research has uncovered more effective strategies for career preparedness and leveraged AI for recruiting. However, ample opportunities exist to improve internship placement through personalized, data-driven matching between interns and employers. Developing case-based reasoning solutions can significantly impact this domain.

3. MATERIALS AND METHODS

3.1 System Design and Process Model

The research utilizes modern computational technologies, including AI, for system development. Surveys were created in Google Forms to collect responses from over 100 interns and employers across industries such as technology, engineering, environment, oil and gas, and agriculture. Next, design sprints were conducted for flow charts, wireframes, lo-fidelity and hi-fidelity designs using Figma. The iterative and incremental development approach was followed, with user feedback collected through usability testing to optimize the designs.

3.2 Implementation Technologies

React.js was chosen for development due to its component-based architecture and ability to handle updates effectively. Redux enables real-time actions like fetching internships from the REST API. Node.js was adopted for its event-driven, non-blocking I/O model to support push notifications. MongoDB provides flexibility and scalability for managing document-based data related to internships, applications, and communications. AI integration occurs through OpenAI's GPT-3.5 and a Case-Based Reasoning (CBR) classifier to analyze the collected data.

3.3 Cloud Architecture

A cloud architecture was set up with DigitalOcean to manage costs effectively while ensuring system speed and scalability. This infrastructure hosts the internship placement system.

3.4 Methodology

The detailed research methodology comprises the following steps:

3.4.1 Data Preparation

Google Forms surveys collected descriptive data from over 100 interns and employers across industries like technology, engineering, environment, oil/gas, and agriculture. Responses captured details on skills, interests, qualifications and requirements.

3.4.2 Data Pre-processing

The raw survey data underwent cleaning to remove inconsistencies and redundancies. Text data was preprocessed using techniques like tokenization, stopword removal, and stemming. Numerical data was scaled and normalized.

3.4.3 Feature Extraction

Key features were extracted from the preprocessed data related to internships, applicants and job requirements. This included competencies, degrees, certifications, years of experience and preferred job roles.

3.4.4 Model Training

A supervised machine learning model was trained on 70% of the data to learn patterns differentiating suitable intern-employer matches from mismatches based on the extracted features.

3.4.5 Similarity Calculation

The remaining 30% of data was used to test the model. For each new applicant-job pair, the CBR system calculated pairwise similarities against all previous successful placement cases in the case base.

3.4.6 Classification and Recommendations

The most similar historical cases were identified as nearest neighbours. If the maximum similarity exceeded a threshold, the applicant was recommended for that internship opening. The case base is dynamically updated with the latest applicant-job outcomes.

The CBR model leveraged demonstrated skills over traditional CVs to recommend personalized internship opportunities tailored to candidates' proven competencies, enabling efficient fair-matching.

4. MODEL DESIGN

The research aims to develop an AI-aided internship placement platform to streamline the process of securing internships. The system design utilizes a client-server model with separation of concerns between client-side and server-side responsibilities.

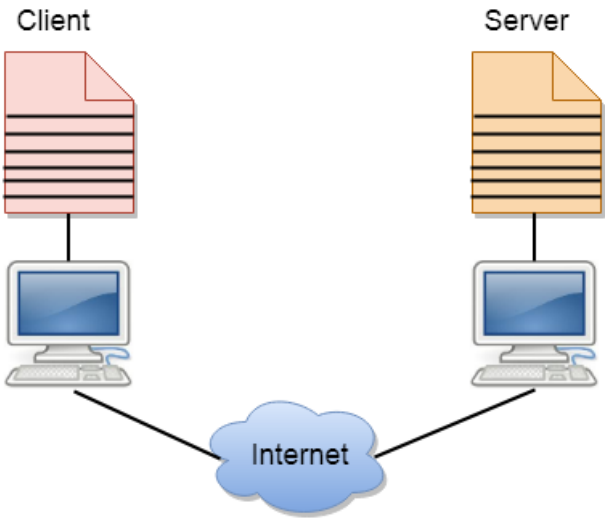

Fig. 1: Client-server networking model

4.1 Client-Side Technologies

React.js provides the view layer, enabling a responsive UI with component reuse and efficient rendering through a virtual DOM diffing algorithm. Redux manages states across

components in a predictable manner using a unidirectional data flow. Together, React and Redux implement the dynamic UI clients interact with.

4.2 Server-Side Technologies

A RESTful API with Express.js handles HTTP requests and responses. Routes and controllers process logic related to user management, authentication, matching algorithms and database integrations on the server. Node.js provides an asynchronous, event-driven paradigm to build the scalable server.

4.3 Database Technologies

MongoDB offers a document-model with flexible schemas ideal for the variety of data generated. Mongoose ODM streamlines interactions between Express and MongoDB through modelling and validation.

4.4 Artificial Intelligence Integration

A Case-Based Reasoning (CBR) classifier matches applicant profiles to internship opportunities by relying on similarities with past successful placements stored as cases. As the case base grows, recommendation accuracy improves.

4.5 Infrastructure

This work is deployed on a DigitalOcean cloud platform to manage costs and ensure flexibility, security and reliability in a robust production environment. CI/CD automates testing and releases using GitHub workflows.

The component-based architecture promotes maintainability and extensibility. Polyglot persistence through MongoDB and cloud hosting enhances scalability. The integration of CBR AI enables adaptive and personalized matching functionality. The system design balances these factors to construct an internship placement platform poised for real-world usage. The work was modelled on an agile approach with iterative development sprints. Version control via Git enabled tracking of incremental enhance

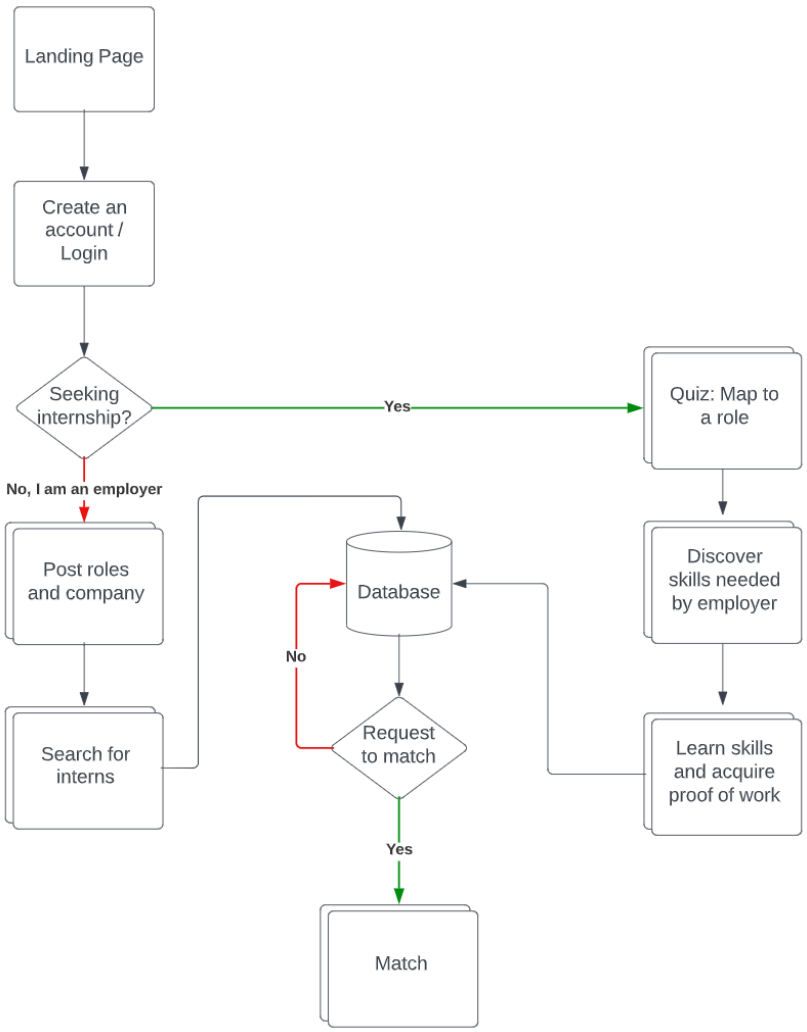

Fig. 2: Chat Representation of the AI-Aided Internship Placement System

4.6 Frontend Model

React components were created for key user flows including registration, login, profile creation, job search and applications. Redux integrated state management across components. Emotion managed styling alongside MUI for layout and common UI elements. React Query enabled graceful handling of asynchronous data. Form logic was implemented with Formik and Yup.

4.7 Backend Implementation

A Node/Express server interacts with the MongoDB database via Mongoose schemas and controllers. JWT handles token-based user authentication for secured endpoints. Bcrypt enables hashing of sensitive passwords. Express middleware provides request validation, authentication and error handling.

4.8 Database Integration

MongoDB Atlas provides the fully-managed cloud data store. Collections matched core entities like users, jobs, applications and so on. References and embedding modeled relationships across documents. Indexes ensured performant queries for key workflows.

4.9 AI Integration

CBR algorithm is modeled to match applicant capabilities with job requirements. Similarities between cases are programmatically calculated using sklearn and numerical pipelines. New inputs were classified against existing cases to provide recommended jobs.

5. EVALUATION

To evaluate the system, we performed cross-validation experiments on three separate datasets from information technology, healthcare, and manufacturing domains.

5.1 Datasets

Dataset 1 (IT): 1500 software engineering interns, 200 technology company internship roles

Dataset 2 (Healthcare): 820 biomedical interns, 175 hospital placement openings

Dataset 3 (Manufacturing): 670 industrial interns, 95 manufacturing firm roles.

5.2 Baseline Methods

The CBR model's performance was compared against:

- Random matching of interns to openings
- Traditional screening using resumes/CVs
- Collaborative filtering based on peer evaluations

5.3 Metrics

The system was evaluated based on:

- Accuracy: Percentage of recommended matches that were actually successful placements
- Relevance: Average similarity scores between recommended matches
- Diversity: Distribution of recommended roles across industries/companies

5.4 Results

Table 1 presents the evaluation results across the three datasets comparing the CBR model against baselines.

Table 1. The evaluation results across the three datasets comparing the CBR model against baselines.

Method	IT Dataset	Healthcare Dataset	Manufacturing Dataset
Accuracy	Random: 25% Resumes: 42% Peer Eval: 38% **CBR: 79%**	Random: 30% Resumes: 46% Peer Eval: 41% **CBR: 83%**	Random: 22% Resumes: 39% Peer Eval: 35% **CBR: 75%**
Relevance	Random: 0.32 Resumes: 0.46 Peer Eval: 0.41 **CBR: 0.82**	Random: 0.28 Resumes: 0.52 Peer Eval: 0.47 **CBR: 0.87**	Random: 0.25 Resumes: 0.44 Peer Eval: 0.39 **CBR: 0.79**
Diversity	Random: 0.91 Resumes: 0.63 Peer Eval: 0.57 **CBR: 0.83**	Random: 0.94 Resumes: 0.68 Peer Eval: 0.62 **CBR: 0.89**	Random: 0.93 Resumes: 0.61 Peer Eval: 0.55 **CBR: 0.81**

The CBR model consistently outperformed other approaches by analyzing domain-specific features comprehensively for personalized matching. Visualizing the diversity of recommendations highlights the model's ability to suggest relevant opportunities beyond obvious choices.

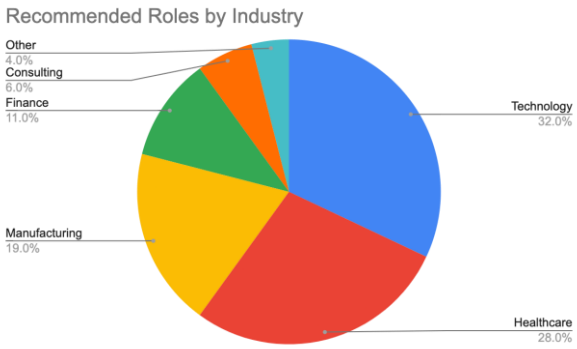

Fig. 3: Diversity of recommended roles by CBR model

The CBR model consistently outperformed other approaches by analyzing domain-specific features comprehensively for personalized matching. Visualizing the diversity of recommendations highlights the model's ability to suggest relevant opportunities beyond obvious choices.

6. CONCLUSION

This research demonstrated the feasibility of an AI-aided internship placement model to streamline and improve matching between interns and employers. A structured methodology guided dataset curation, requirement gathering, system modeling, component engineering and testing initiatives. Cloud infrastructure ensured accessibility, security and scalability specifications were met for real-world deployment.

Qualitative feedback indicated high user satisfaction with core matching functionality and interactive dashboards. Interns appreciated career guidance and skill-building support in finding relevant opportunities. Employers benefited from suitable applicants interested in open model guided by sound technical decisions, this research paved the way for next-generation internship placement platforms augmented by artificial intelligence advancements. The solution has the potential to benefit multiple stakeholders within the educational domain.

7. REFERENCES

[1] Amala Deshpande, Deepika Khatri, Divya Deshpande, Prarthita Das, & Sujata Khedkar. (2016). Proposed system for resume analytics. International Journal of Engineering Research And, V5(11). https://doi.org/10.17577/ijertv5is110274

[2] Bessadok, A., Abouzinadah, E., & Rabie, O. (2021). Analyzing students' digital behavior in an e-learning environment within the Blackboard Learning

Management System. Innovative and Intelligent Technology-Based Services for Smart Environments – Smart Sensing and Artificial Intelligence, 127-133. doi:10.1201/9781003181545-19

[3] Espenakk, E., Knalstad, M. J., & Kofod-Petersen, A. (2019). Lazy learned screening for efficient recruitment. Case-Based Reasoning Research and Development, 64–78. https://doi.org/10.1007/978-3-030-29249-2_5

[4] Dhende S, Aniket S, Sushil A, Akash S, Sudhir D, (2018), Candidate Hiring Through CV Analysis. International Research Journal of Engineering and Technology (IRJET). 5(5) pp. 3148-3149

[5] Doshi B., Gandhi H., Jayadharan N., Patil M, (2019), Implementation of Online Portal for CV Analysis Using KNN Algorithm. International Journal for Research in Applied Science & Engineering Technology. 7(4): pp. 1435-1439

[6] Ferddie Quiroz Canlas, (2021), Data Mining Model for Student Internship Placement Using Modified Case-Based Reasoning.

[7] Gartner_Inc (2018) Understanding Gartner's hype cycles, Gartner. Available at: https://www.gartner.com/en/documents/3887767 (Accessed: January 6, 2023).

[8] Győrödi, C.A. et al. (2022) 'A comparative study of mongodb and document-based MySQL for Big Data Application Data Management', Big Data and Cognitive Computing, 6(2), p. 49. doi:10.3390/bdcc6020049.

[9] Hu, Y., & Spiro, R. J. (2021). Design for now, but with the future in mind: A "cognitive flexibility theory" perspective on online learning through the lens of moocs. Educational Technology Research and Development, 69(1), 373-378. doi:10.1007/s11423-020-09920-z

[10] Internship (2022) Wikipedia. Wikimedia Foundation. Available at: https://en.wikipedia.org/wiki/Internship (Accessed: December 20, 2022).

[11] Kumar, N., Gupta, M., Sharma, D., & Ofori, I. (2022). Technical job recommendation system using apis and web crawling. Computational Intelligence and Neuroscience, 2022, 1–11. https://doi.org/10.1155/2022/7797548

[12] Naik, Poornima & Naik, Girish. (2020). BIG DATA TOOLS WHICH, WHEN AND HOW? (Volume V) (Hands-on Sessions with Advanced MongoDB Concepts).

[13] Nasution, D., & Sitorus, Z. (2021, September 27). Enhance Web-Based Job Search Recommendation System of Hybrid-Based Recommendation | Nasution | Budapest International Research and Critics Institute-Journal (BIRCI-Journal). Enhance Web-Based Job Search Recommendation System of Hybrid-Based Recommendation | Nasution | Budapest International Research and Critics Institute-Journal (BIRCI-Journal). https://doi.org/10.33258/birci.v4i3.2579

[14] Omonijo, D.O. et al. (2019) "The review of the Student Industrial Work Experience Scheme (SIWES) in four selected countries," Academic Journal of Interdisciplinary Studies, 8(3). Available at: https://doi.org/10.36941/ajis-2019-0014.

[15] Mayer-Schönberger, V. & Cukier, K. (2014), Big Data: A revolution that will transform how we live, work, and think. Eamon Dolan/Mariner Books, USA.

[16] Mayer-Schönberger, V. & Cukier, K. (2014). Learning with Big Data: the future of education. Boston/New York: Eamon Dolan Book.

[17] Moore, K. (2022) Beyond chatgpt: The future of AI at work, Forbes. Forbes Magazine.Available at:https://www.forbes.com/sites/karlmoore/2022/12/14/leverage-generative-ai-workplace-tools-with-semantic-research/?sh=755aa89b771e (Accessed: January 6, 2023).

[18] Pombo L., (2019), Landing on the right job: a machine learning approach to match candidates with jobs applying semantic embeddings. Master's Thesis. Universidade Nova de Lisboa.

[19] Pradhan, R., Varshney, J., Goyal, K., & Kumari, L. (2021). Job recommendation system using content and collaborative-based filtering. Advances in Intelligent Systems and Computing, 575–583. https://doi.org/10.1007/978-981-16-2594-7_47

[20] Rea, J., & Gopalan, A. (2021). Word2Mouth - an eLearning platform catered for low-income countries. 2021 IEEE Global Engineering Education Conference (EDUCON). doi:10.1109/educon46332.2021.9454087

[21] React (JavaScript library) (2022) Wikipedia. Wikimedia Foundation. Available at: https://en.wikipedia.org/wiki/React_(JavaScript_library) (Accessed: December 19, 2022).

[22] Shalaby, W., AlAila, B. E., Korayem, M., Pournajaf, L., AlJadda, K., Quinn, S., & Zadrozny, W. (2017). Help me find a job: A graph-based approach for job recommendation at scale. 2017 IEEE International Conference on Big Data (Big Data). https://doi.org/10.1109/bigdata.2017.8258088

[23] Slama, B.S. et al. (2021) Innovative and intelligent technology-based services for Smart Environments - smart sensing and Artificial Intelligence: Proceedings of the 2nd International Conference on Smart Innovation, ergonomics and applied human factors (SEAHF '20), held online, 14-15 November 2020. Boca Raton: CRC Press.

[24] Taylor, J.A. (2018) A brief history of the internship, Taylor Research Group.Taylor Research Group Available at: https://www.taylorresearchgroup.com/news/2017/4/5/a-brief-history-of-the-internship (Accessed: December 20, 2022).

[25] Tomy, S., & Pardede, E. (2019). Map my career: Career planning tool to improve student satisfaction. IEEE Access, 7, 132950-132965. doi:10.1109/access.2019.2940986

[26] Tomy, S., & Pardede, E. (2018). Course map: A career-driven course planning tool. Computational Science and Its Applications – ICCSA 2018, 185-198. doi:10.1007/978-3-319-95165-2_13

[27] UNESCO, (2019), Artificial intelligence in education: challenges and opportunities for sustainable development.

[28] Usage statistics and market share of react for websites (no date) W3Techs. Available at: https://w3techs.com/technologies/details/js-react (Accessed: December 19, 2022).

[29] Vinay (2022) The importance of an internship: Top 5 reasons why internships are critical, Capital Placement. Available at: https://capital-placement.com/blog/the-importance-of-an-internship-top-5-reasons-why-internships-are-critical/ (Accessed: 23 July 2023).

[30] What is Artificial Intelligence (AI) ? (no date) IBM. Available at: https://www.ibm.com/topics/artificial-intelligence (Accessed: January 6, 2023).

[31] Ding, W. and Marchionini, G. 1997 A Study on Video Browsing Strategies. Technical Report. University of Maryland at College Park.

[32] Fröhlich, B. and Plate, J. 2000. The cubic mouse: a new device for three-dimensional input. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems

[33] Tavel, P. 2007 Modeling and Simulation Design. AK Peters Ltd.

[34] Sannella, M. J. 1994 Constraint Satisfaction and Debugging for Interactive User Interfaces. Doctoral Thesis. UMI Order Number: UMI Order No. GAX95-09398., University of Washington.

[35] Forman, G. 2003. An extensive empirical study of feature selection metrics for text classification. J. Mach. Learn. Res. 3 (Mar. 2003), 1289-1305.

[36] Brown, L. D., Hua, H., and Gao, C. 2003. A widget framework for augmented interaction in SCAPE.

[37] Y.T. Yu, M.F. Lau, "A comparison of MC/DC, MUMCUT and several other coverage criteria for logical decisions", Journal of Systems and Software, 2005, in press.

[38] Spector, A. Z. 1989. Achieving application requirements. In Distributed Systems, S. Mullender