CSI MASTERFORMAT CODE	CLASS NAME	HIERARCHICAL LEVEL	TEST CASES			CLASSIFICATION ACCURACY [%]
			TOTAL	RIGHT	WRONG	
3000	Concrete	1	284	268	16	94.37
6000	Wood_and_plastics	1	348	327	21	93.97
16000	Electrical	1	239	229	10	95.82
	LEVEL 1		871	824	47	94.60
3050	Basic_materials_and_methods	2	105	78	27	74.29
3100	Forms_and_accessories	2	36	32	4	88.89
3200	Concrete_reinforcing	2	11	11	0	100.00
3400	Precast_concrete	2	31	28	3	90.32
3500	Cementitious_decks_and_underlayment	2	37	31	6	83.78
3600	Grout	2	11	11	0	100.00
3900	Restoration_and_cleaning	2	47	39	8	82.98
6050	Basic_materials_and_methods	2	50	45	5	90.00
6100	Rough_carpentry	2	56	52	4	92.86
6400	Architectural_woodwork	2	151	136	15	90.07
6600	Plastic_fabrications	2	67	63	4	94.03
16100	Wiring_methods	2	65	64	1	98.46
16500	Lighting	2	99	96	3	96.97
16700	Communications	2	57	51	6	89.47
16800	Sound_and_video	2	19	17	2	89.47
	LEVEL 2		842	754	88	89.55
3410	Plant_precast_structural_concrete	3	18	15	3	83.33
3450	Plant_precast_architectural_concrete	3	13	9	4	69.23
3910	Concrete_cleaning	3	10	9	1	90.00
3920	Resurfacing	3	21	15	6	71.43
3930	Rehabilitation	3	17	13	4	76.47
6060	Wood	3	14	13	1	92.86
6080	Factory_applied_coatings	3	20	20	0	100.00
6090	Fastenings	3	8	8	0	100.00
6130	Heavy_timber_construction	3	17	16	1	94.12
6150	Wood_decking	3	12	11	1	91.67
6410	Custom_cabinets	3	60	58	2	96.67
6420	Paneling	3	18	16	2	88.89
6430	Wood_stairs_and_railing	3	34	34	0	100.00
6440	Ornaments	3	38	35	3	92.11
16120	Conductors_and_cables	3	28	28	0	100.00
16130	Raceway_and_boxes	3	32	32	0	100.00
16510	Interior_luminaires	3	20	19	1	95.00
16520	Exterior_luminaires	3	24	16	8	66.67
16550	Special_purpose_lighting	3	28	16	12	57.14
16710	Communications_circuits	3	26	19	7	73.08
16720	Telephone_and_intercommunication_equipment	3	12	9	3	75.00
	LEVEL 3		470	411	59	87.45

Figure 4: Hierarchical Classification Results

Conclusions

A large percentage of the communications exchanged and documents stored in inter-organizational construction management information systems is based on textual information. Automatic classification of these documents according to a common project schema can improve information organization and access among project organizations that use these systems. Furthermore, such classification methods can be applied to other construction management tasks, including: automated access to project specifications, identification of project problem areas, and generation of knowledge for future activities and projects

In this paper, a framework for automated document classification was described and the prototype of a construction document classification system was presented. The system automates the steps involved in the proposed document classification process. Issues of class scalability and support for hierarchical classification were considered during its implementation. The system supports the generation of classification models for construction projects. After creating these models, construction documents can be easily and quickly classified according to the user-defined project components.

Two case studies were conducted to verify the feasibility of the proposed approach. The first case study tested the performance of text classification algorithms using a building project database as testbed. The results for different classification methods were presented. Support Vector Machine was the algorithm that gave the best results. Its accuracy performance was comparable to human-based document classification. The second case study analyzed the classification accuracy for hierarchical classification structures. A construction products' database, originally classified according to a hierarchical structure, was used in this analysis. The preliminary results presented were very promising, demonstrating the potential and applicability of automated document classification methods for construction management.

References

Aalami, F.B., Fischer, M. and Kunz, J.C. (1998) "AEC 4D-CAD production model: definition and automated generation." *CIFE WP 052.*

aecXML (1999). aecXML <http://www.aecxml.org> (Out 12, 2000).

Anumba, C. J. and Evbuomwan, N. F. O. (1999). "A taxonomy for communication facets in concurrent life-cycle design and construction." *Computer-Aided Civil and Infrastructure Engineering*, 14, 37-44.

Brüggemann, B. M., Holz, K. -P., and Molkenthin, F. (2000) "Semantic documentation in engineering". *Proceedings. of the ICCCBE-VIII,* Palo Alto, CA, August, 2000, 828-835.

Eastman, C.M. (1999). *Building product models: computer environments supporting design and construction.* CRC Press, Boca Raton, FL.

Fischer, M., and Kunz, J. (1995) "The circle: architecture for integrating software." *Journal of Computing in Civil Engineering*, 9(2), 122-133.

Fruchter, R. (1999) "A/E/C Teamwork: a collaborative design and learning space." *Journal of Computing in Civil Engineering*, 13(4), 261-269.

Fuhr, N., Hartmann, S., Knorz, G., Lustig, G., Schwantner, M., and Tzeras, K. (1991). "AIR/X a rule- based multistage indexing system for large subject fields." *Proceedings of RIAO- 91 3rd International Conference*, Barcelona, Spain, 606-623.

Han, J. and Kamber, M. (2001) *Data Mining: Concepts and Techniques.* Morgan Kaufmann.

IAI (1996). International Alliance for Interoperability. < http://iaiweb.lbl.gov/> (Oct 12, 2000)

Joachims, T. (1997). "A probabilistic analysis of the Rocchio algorithm with TFIDF for text categorization". *Proceedings of ICML- 97*, Nashville, TN, 143-151.

Joachims, T. (1998). "Text categorization with support vector machines: learning with many relevant features". *Proceedings of ECML- 98*, Chemnitz, Germany, 137-142.

Kosovac, B., Froese, T., and Vanier, D. (2000). "Integrating heterogeneous data representations in model-based AEC/FM systems." *Proceedings of CIT 2000*, Reykjavik, Iceland, 1, 556-566.

Lewis, D. D. (1992) "An evaluation of phrasal and clustered representations on a text categorization task". *Proceedings of SIGIR- 92*, Kobenhavn, Denmark, 37-50.

McCallum, A. K. (1996) "Bow: A toolkit for statistical language modeling, text retrieval, classification and clustering." <http://www.cs.cmu.edu/~mccallum/bow>.

Maher, M.L. and Simoff, S.J. (1998) "Ontology-based multimedia data mining for design information retrieval" *Proc. of Computing in Civil Engineering*, ASCE, pp. 212-223.

MasterFormat (1995). *MasterFormat 1995 Edition*. Construction Specifications Institute.

Mitchell (1997) *Machine Learning*. McGraw Hill. New York, NY.

O'Brien, W. J. (2000). "Implementation issues in project web-sites: a practitioner's viewpoint." *Journal of Management in Engineering,* ASCE. May 2000, 16(3), 34-39.

OSMOS (2000). Open system for inter-enterprise information management in dynamic virtual environments – OSMOS Project. <http://cic.vtt.fi/projects/osmos/index.html> (Oct 12, 2000)

OCCS (2001). Overall Construction Classification System. <http://www.occsnet.org> (Nov 15, 2001).

Rezgui, Y., Brown, Y., Cooper, G., Yip, J., Brandon, P., and Kirkham, J. (1996) "An information management model for concurrent construction engineering." *Journal of Automation in Construction*, 5(4), 343-355.

Rojas, E. M. and Songer, A. D. (1999). "Web-centric systems: a new paradigm for collaborative engineering." *Journal of Management in Engineering*, 15(1), 39-45.

Salton G. and Buckley C. (1988) "Term weighting approaches in automatic text retrieval.", *Information Processing and Management*, 2(5), 513-523.

Schapire, R. E. and Singer, Y. (2000). "BoosTexter: a boosting-based system for text categorization", *Machine Learning 39* , 2/3, 135-168.

Scherer, R. J. and Reul, S. (2000) "Retrieval of project knowledge from heterogeneous AEC documents". *Proceedings of the ICCCBE-VIII,* Palo Alto, Ca, August, 2000, 812-819.

Sebastiani, F. (1999) "Machine learning in automated text categorisation." *Technical Report IEI-B4-31-1999*, Istituto di Elaborazione dell'Informazione, CNR, Pisa, Italy.

Soibelman, L. and Caldas, C. (2000)."Project extranets for construction management: the American experience." *Proceedings of Entac-2000*, May, 2000, Salvador, Brazil.

Soibelman, L. and Kim, H. (2002) "Generating construction knowledge with Knowledge Discovery in Databases" *Journal of Computing in Civil Engineering*, ASCE, 16(1), 39-48.

Soibelman, L. and Peña-Mora, F. (2000). "A distributed multi-reasoning mechanism to support the conceptual phase of structural design" *J. Struct. Engineering*, ASCE, 126 (6), 733-742.

Sweet's (2002) Sweet's Product Marketplace <http://sweets.construction.com/default.jsp> (April 3, 2002)

Teicholz, P. (1999). "Vision of future practice." *Berkeley-Stanford Workshop on Defining a Research Agenda for AEC Process/Product Development in 2000 and Beyond*, Stanford, CA.

ToCEE - Towards a Concurrent Engineering Environment Project (2000). "The ToCEE client-server system for concurrent engineering." *Final Report - ESPRIT Project No. 20587.*

UniFormat (1998). *UniFormat 1998 Edition*. Construction Specifications Institute.

VEGA (1999). Virtual enterprises using groupware tools and distributed architecture – VEGA Project <http://cic.cstb.fr/ILC/ecprojec/vega/home.htm> (Oct 12, 2000)

Yang, M. C., Wood, W. H. and Cutkosky, M. R. (1998) "Data mining for thesaurus generation in informal design information retrieval". *Proc. International Computing Congress*, 189-200.

Yang, Y. and Pedersen, J.O. (1997) "A comparative study on feature selection in text categorization". *Proceedings of ICML-97,* Nashville, TN, 412-420.

Yang, Y. 1999. "An evaluation of statistical approaches to text categorization". *Information Retrieval*, 1 , 1-2, 69-90.

Yang, Y. and Liu, X. (1999) "A re-examination of text categorization methods." *Proceedings of ACM SIGIR Conf. on Research and Development in Information Retrieval (SIGIR'99)*, 42-49.

Weiss, S. A., Kasif, S., Brill, E. (1997) "Text classification in USENET newsgroups: A progress report." Department of Computer Science. The Johns Hopkins University. April 1997,

Wood, W.H. (2000) "The development of modes in textual design data". *Proceedings of the ICCCBE-VIII,* Palo Alto, CA, August, 2000, 882-889.

Zarli, A. and Rezgui, Y. (2000). "A survey of internet-oriented technologies for document-driven applications in construction open dynamic virtual environments". *Proceedings of CIT 2000 - International Conf. Construction Information Technology*, Reykjavik, Iceland, 1, 1089-1101.

Zhu, Y., Issa, R. R. (2001). "Web-based construction document processing via malleable frame" *Journal of Computing in Civil Engineering*, 15(3), 157-169.

Implementation Challenges and Research Needs of the IFC Interoperability Standard:
Experiences from HUT-600 Construction Pilot

Calvin Kam[1], Martin Fischer[2], Reijo Hänninen[3], Seppo Lehto[4], Jarmo Laitinen[5]

KEYWORDS:

Industry Foundation Classes (IFC), Interoperability, Data Exchange, Product Model

ABSTRACT:

This paper presents the findings from the use of Industry Foundation Classes for the design and construction of the Helsinki University of Technology Auditorium Hall (HUT-600) in Finland. For HUT-600, IFC implementation supported the sharing of geometric data, thermal values, and material properties among the architects, mechanical engineers, and general contractors. The implementation benefited the project with minimized data re-entry, increased accuracy of information exchange, and reduced design time during the early project phase. However, the immaturity of IFC-compliant software and middleware undermined the reliability and a more rigorous testing of the IFC's. Analyzing the HUT-600 case study, we suggest that researchers and software developers focus on partial data exchange, model server approach, and software robustness to extend the benefits and improve the reliability of the IFC's.

---

[1] Graduate Research Student, Civil & Environmental Engineering, Stanford University, Stanford, CA. 94305-4020, U.S.A. <ckam@stanford.edu>
[2] Associate Professor, Civil & Environmental Engineering and (by courtesy) Computer Science, Director of the Center for Integrated Facility Engineering, Stanford University, Stanford, CA. 94305-4020, U.S.A. <fischer@stanford.edu>
[3] General Manager, Olof Granlund Oy, Helsinki, Finland <reijo.hanninen@granlund.fi>
[4] Senate Properties, Helsinki, Finland <seppo.lehto@senaatti.com>
[5] Ph.D, Customer Solutions, Solibri Inc., Helsinki, Finland <jarmo.laitinen@solibri.com>

Background

Interoperability is the ability of a system to use the parts or equipment of another system (Merriam-Webster, 2002). In each building design and construction project, there are numerous information or data parts across many different software applications and professional specialists. However, the challenge to exchange such data parts (e.g., 3D geometries, thermal values, construction schedule, building assembly, etc.) in conventional practice undermines the efficiency of project collaboration while risking the accuracy of data exchanges.

Since 1995, the International Alliance of Interoperability (IAI) has defined and promoted the use of the Industry Foundation Classes (IFC) standard. It defines interoperability in the building industry as "an environment in which computer programs can share and exchange data automatically (without translation or human intervention), regardless of the type of software or of where the data may be residing" (IAI 1995). The alliance aims at utilizing the IFC's, an interoperable standard, to define a single object-oriented data model to allow different disciplines to accurately share technical information with IFC-compliant tools. The IAI strives to provide a universal standard for sharing cross-disciplinary data among the fields of building design, construction, and operation. This allows project participants to work across different application packages and to build upon existing data, while eliminating the inefficiencies and inconsistencies associated with conventional practices of data re-entry.

Prior to HUT-600's adoption of IFC's in 2000-01, there were limited test cases of this evolving promise of interoperability: In 1999, the Building Lifecycle Interoperable Software (BLIS) project[6] tested a series of IFC "use cases" and model "views" on small set of retrospective test cases. In the SPADEX project (SPADEX 2001), the participants tested interdisciplinary data sharing within a small scale simulation environment.

The HUT-600 Construction Pilot

As the property owner of the Helsinki University of Technology, Senate Properties[7] in Finland assembled a team of designers, consultants, contractors, and researchers for its new Auditorium-600 (HUT-600) construction pilot project. The National Technology Agency (TEKES) in Finland had been promoting applied and industrial research and development. In HUT-600, TEKES sponsored the testing of state-of-the-art technologies and data standards. In particular, our international research partnership documented and analyzed the use of product models and the IFC interoperability standard.

---

[6] URL: http://www.blis-project.org/

[7] http://www.senaatti.com/index.asp?siteID=2

The virtual representation of building design and construction through an object-oriented 3D model is a crucial foundation for interoperability. The HUT-600 project adhered to the product modeling approach, in which the professional team constructed and maintained 3D models with explicit knowledge of building components, spatial definitions, material composition, and other parametric properties. Only under this product modeling approach could the team leverage the object intelligence within the 3D model for data interoperability. Since the architects, engineers, contractors, and the researchers in HUT-600 were committed to utilizing product models for simulations and analyses, we tested the extent to which the exchange of IFC-based project data could take place within an array of disciplines, applications, and simulations (Figure 1).

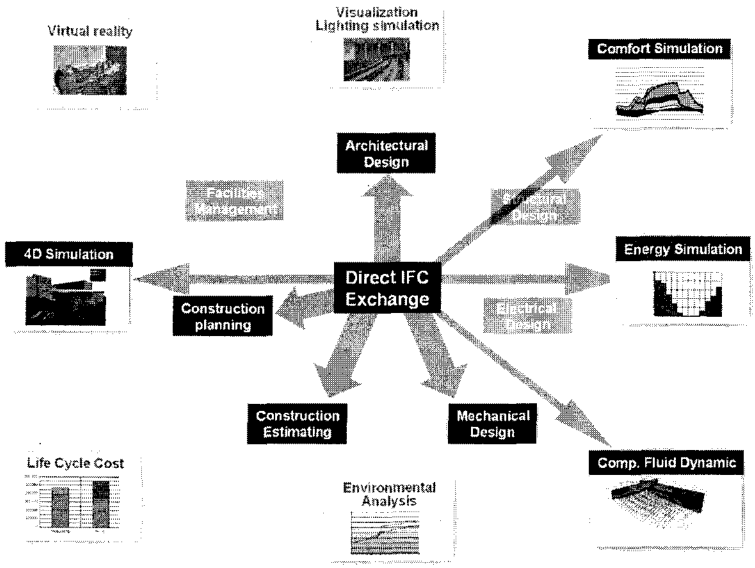

Figure 1. Scope of testing IFC-based project data exchange in HUT-600 project.

IFC Implementation in HUT-600

In the HUT-600 project, IFC-based data exchange took place during the early project phases and mostly among the architects, mechanical engineers, construction managers, and the 4D research collaborators. Using IFC release 1.5.1, the project team experimented with the sharing of architectural models, thermal simulation data, mechanical component geometries, building composition, and material data (Figure 2).

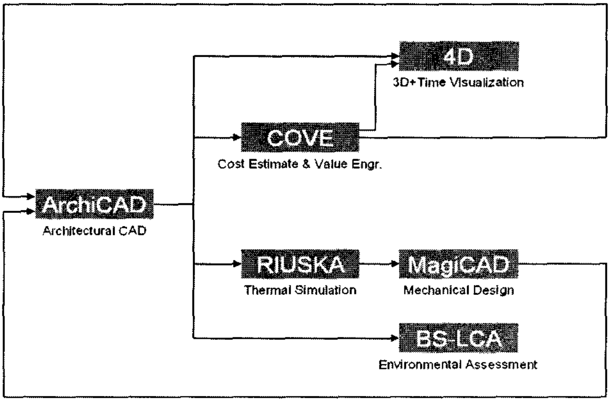

Figure 2. Flow of IFC files across software applications and disciplines on HUT-600 project.

The following subsections explain the processes of IFC exchange within the six IFC applications in HUT-600:

1. ***Architectural Design***

   The architects of HUT-600 adopted a product modeling approach to building design and documentation. They created a 3D model and assigned objects with accurate properties during the conceptual planning phase, and followed this product modeling approach through the construction documentation phase. With IFC-compliant design software (ArchiCAD from Graphisoft[8]), the architectural designers generated IFC files that contained three-dimensional building geometry, spatial zoning information, and material composition of the geometric model.

   The IFC files the architects exported were read by the mechanical engineer's software to conduct thermal simulations, the general contractor's software to generate cost estimates and schedule, the building systems consultant's software to assess environmental impacts, and the research collaborator's 4D modeling tool[9].

   Iteratively, the architects could import again the IFC files to which the mechanical engineers had incorporated their design additions and changes.

[8] URL: http://www.graphisoft.com
[9] URL: http://www.commonpointinc.com

2. ***Thermal Simulation***
During the schematic design phase, the building system consultants conducted thermal simulation to estimate the heat gain and heat loss of the building in response to the climate, architectural configurations, and the anticipated operation by the occupants.

In the construction pilot, the IFC-based product model enabled the thermal simulation tool RIUSKA[10], developed by Olof Granlund Oy, to import the 3D building geometry and its spatial data for thermal simulation. In turn, RIUSKA exported thermal data, such as cooling and heating design temperatures, via IFC, for mechanical design. RIUSKA reads and writes IFC files through a bridging middleware tool—BSPro[11].

3. ***Mechanical Design***
With the target design values set from the thermal simulations and the spatial configurations from the architectural designs, the mechanical engineers were then ready to design and optimize the cooling and heating systems. HUT-600 engineers employed Progman Oy's MagiCAD[12] to conduct 3D modeling and optimization of the Heating, Ventilating, and Air-Conditioning (HVAC) system. This object-oriented mechanical design software was capable of importing the IFC file from the thermal simulation tool. It directly imported the cooling and heating design temperatures, supply and exhaust air flow rates, and the total heat gain.

After the engineers optimized the location and sizing of the HVAC system, they exported another IFC file that contained the geometric representation of HVAC components. The architects and the research collaborators were able to import this IFC file and incorporate the ductwork, air-handling systems, and other mechanical devices into the 3D architectural model as well as the 4D model.

4. ***Environmental Impact Assessment***
HUT-600's building system consultants conducted an environmental impact assessment to assess the environmental influences of the building materials and energy for this facility. With Olof Granlund Oy's BSLCA software, the consultants quantified the amount of pollution emission, global warming, acidification, etc. in support for material and system selection. Since BSLCA is IFC-compliant through BSPro, the consultants could directly import quantitative values (such as height, length, area, etc.) and descriptive information (such as materials, composition, etc.) from the architect's IFC exports.

---

[10] URL: http://www.eren.doe.gov/buildings/tools_directory/software/riuska.htm
[11] URL: http://www.bspro.net
[12] URL: http://www.progman.fi/english/e_index.htm