CSC3031 - Research and Project Skills

Research Skills: Aspects of Academic Writing in Computer Science

John Fitzgerald

I'm grateful to the Writing Development Unit and to Dr Paul Ezhilchelvan for valuable input used in this lecture.

Aspects of Academic Writing in CS

We often feel overwhelmed when we have to prepare written work. It can be helpful to think about this almost as a development task with a design, implementation and test.

1. Writing is a Development Process
2. Structuring your Writing
3. Elements of CS ‘style’

Example Papers

Advanced Engineering Informatics 51 (2022) 101452

Contents lists available at ScienceDirect

Advanced Engineering Informatics

journal homepage: www.elsevier.com/locate/aei

Full length article

Modelling the impacts of crowds on occupants in the built environment—A static, rule-based approach to human perception and movement

Beidi Li [a], John Fitzgerald [b], Carl Schultz [a,*]

[a] Aarhus University, Denmark
[b] Newcastle University, UK

ARTICLE INFO

Keywords:
Occupant experiences
Affordances
Crowd modelling
Declarative spatial reasoning
ASP4BIM

ABSTRACT

A building occupant's experiences are not passive responses to environmental stimuli, but are the results of *multifaceted*, *prolonged* interactions between people and space. We present a framework and prototype software tool for *logically reasoning about* occupant perception and behaviour in the context of dynamic aspects of buildings in operation, based on qualitative deductive rules. In particular, we focus on the co-presence of different user groups and the resulting impact on perceptual and functional affordances of spatial layouts by utilising the concept of *spatial artefacts*. As a first proof of concept of our approach, we have implemented a prototype crowd analysis software tool in our new system ASP4BIM, developed specifically to support architectural design reasoning in the context of public-facing buildings with complex signage systems and diverse intended user groups. We evaluate our prototype on the Urban Sciences Building at Newcastle University, a large, state-of-the-art living laboratory and multipurpose academic building. Our findings are that the ASP4BIM-based prototype supports a range of novel query services for formally analysing the impacts of crowds on pedestrians that are logically derived through the use of qualitative deductive rules, that complements other powerful crowd analysis approaches such as agent-based simulation.

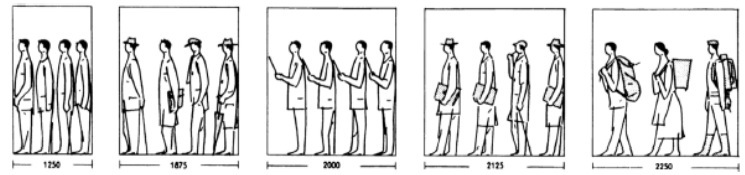


Fig. 3. Crowd spacing conditions in representative situations (units in mm). From left to right, with personal equipment, adults walking, children sprinting.
Source: Taken directly from [60].

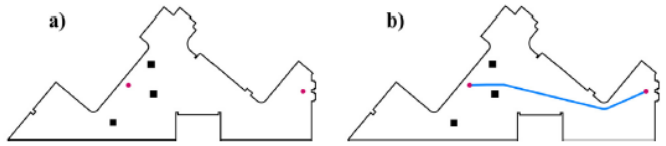


Fig. 4. (a) Environment with boundaries (black lines), obstacles (black boxes), and randomly chosen locations (ruby circles); (b) Generating the shortest path (blue lines) connecting the locations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

```
define_spatial_artefact(perp_view_space, Artefact, S, C) :-
  target(S), location(S, Loc),
  crowd(C), representation(C, R1),
  environment(S), representation(S, R2),
  Artefact = @perp_view(Loc, R1, R2).

spatial_artefact(ViewSpace, Artefact) :-
  define_spatial_artefact(ViewSpace, Artefact, _, _).

induced_by(Artefact, S) :-
  define_spatial_artefact(_, Artefact, S, _).

contingent(Artefact, C) :-
  define_spatial_artefact(_, Artefact, _, C).
```

Example Papers

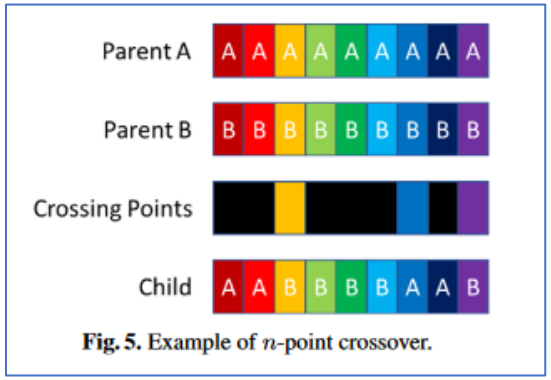


Fig. 5. Example of n-point crossover.

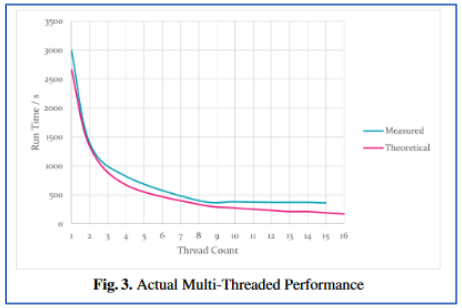


Fig. 3. Actual Multi-Threaded Performance

Genetic Algorithms for Design Space Exploration of Cyber-Physical Systems: an Implementation in INTO-CPS

Maximilian Rose and John Fitzgerald

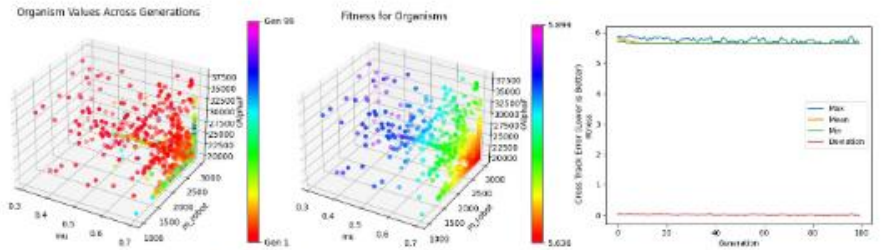
School of Computing, Newcastle University, United Kingdom
{M.Rose, John.Fitzgerald}@newcastle.ac.uk

Abstract. We describe the initial implementation and evaluation of a facility to support Design Space Exploration (DSE) over multi-paradigm models of cyber-physical systems, using Genetic Algorithms (GAs). Existing features of the INTO-CPS toolchain have been updated to allow concurrent exploration of DSE parameters and a new GA-based approach to DSE. Based on a comparison with the previously reported DSE of the multi-model of an agricultural robot, the developed solution appears to reduce the time required to perform a DSE by providing user-configurable parameters for the number of concurrent DSE simulations. It also provides a modular, user-extendable, GA-based DSE system that can be used with the INTO-CPS toolchain.

Keywords: digital twin, FMI, co-simulation, Design Space Exploration, Genetic Algorithms

Scenario	Exhaustive Best Combination cAlphaF	μ	Mass	GA Best Combination cAlphaF	μ	Mass	Cross Track Error Difference
Sin 2	20000	0.7	3000	20000	0.70	3000	-1.46×10^{-6}
Speed Ramp 1	38000	0.4	1000	37019	0.45	1038	8.40×10^{-2}
Speed Step 1	20000	0.3	1000	34725	0.45	1027	0.00
Turn Ramp 1	20000	0.3	3000	20000	0.30	3000	-9.21×10^{-7}

Fig. 12. Exhaustive Best Combination vs GA Best Combination. Results are rounded appropriately. Cross track error difference calculated as Exhaustive−GA.

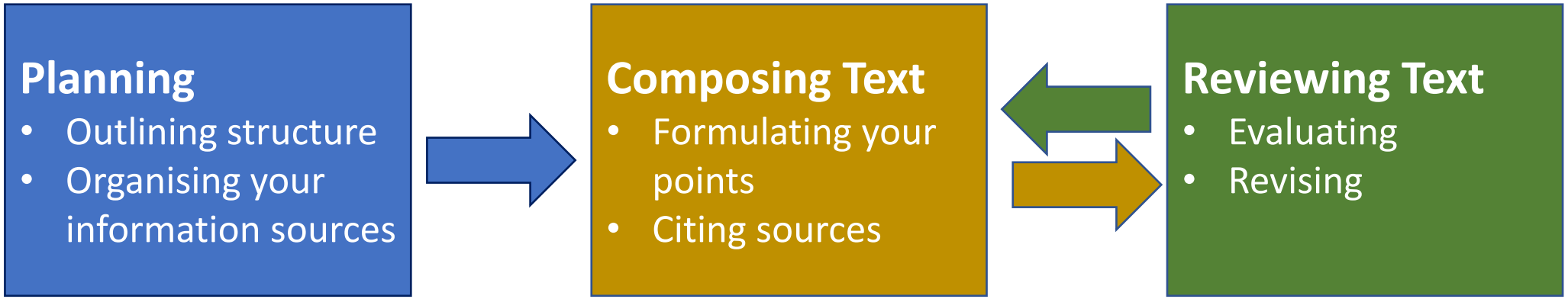


(a) Gene values across all generations. (b) Fitness against gene values. (c) Fitness trends.

Fig. 13. Sin Results.

Writing is a Development Process

- Thinking of writing like a development project
- Break the writing task into smaller units (Planning)
- Just as you might test and fix a unit of code, you can review and edit each unit of text
- You may even want to go back and re-plan



Writing is a Process: Planning

- Know your reader!
  - Review the specification and the marking scheme
  - Think about your reader, what they may know already and what they are expecting from your text
  - This is like setting the aim and objectives for your writing project
- Make up a rough plan of the text, including the sections required
  - Just write the section headings down – this helps ensure you don't forget any of them
  - Don't try to fill the sections with detail at this point – keep your thinking high-level
  - Beware over-planning – remember you can come back and re-plan

Writing is a Process: Composing

- Within each section, write down your main points – just the main points.
- Don’t worry at this stage about the quality of your prose – just focus on defining the main points.
- Some of them might be grouped into subsections – think of this as a work decomposition structure.

Writing is a Process: Reviewing and editing

Try to create distance between you and the text, so that you see it with fresh eyes. Try the following:

- Print it out
- Change the typeface
- Slow down (use a ruler or follow your finger over the text)
- Try reading it aloud (or get the computer to)
- *Note if you are using a professional proofreader, read the Writing Development Unit's Guide to ensure you avoid collusion.*

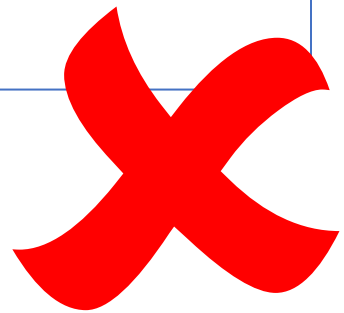
Structuring Your Writing

- Introductions and conclusions
- Structuring the Argument
- Writing in Paragraphs

Structuring: Introductions & Conclusions

- We often start a CS writing project by composing the middle sections of a text
  - Because these describe what we've done – they are the parts we know best.
- Introductions and conclusions can be tricky.
  - Only about 10% of the word count each.
  - A reader will very often read the intro and conclusions first.
  - Important to make a good impression here, so take time over them.
- Think from the perspective of the reader.

- *"Intro: tell then what you're going to say*
- *Main body: say it*
- *Conclusion: tell them that you've said it"*



Structuring: Introduction

- Think about what the reader wants to know:
  - What are you doing?
    - How did you interpret the project aim and objectives?
    - What's the motivation? Why is this work is useful, interesting or important?
  - Why are you doing this?
  - How will you do this?
    - How have you scoped the work?
    - What is the structure of the remainder of the text?

Structuring: Conclusions

- Thinking again from the reader's point of view …
- The following structure is common:
  - Where are we and how did we get here? A short reminder of the areas covered in the main body
  - Where does that leave us? A review of the aim and objectives: in each case, to what extent were they achieved. If not, why not?
  - What next? Directions for further work. This is often a feature of excellent dissertations.
- Don't introduce new material in the conclusion: it is all about adding value to the material you have provided up to this point.

Structuring: Arguments

- In your report you will often make claims that are significant.
- For each claim:
  - Support it with reasons
  - Base the reasons on evidence
  - Acknowledge and respond to other views
  - You may need to explain your principles of reasoning
- Here is an example from the example paper about computational analysis of the effect of crowds on lines of sight in buildings …

Claim

A crowd between a target and a subject negatively impacts the target's visibility and intelligibility … The severity of this impact is proportional to the occurrence, frequency and duration of each visual disruption by pedestrians in the crowd.

Supporting argument

… In case of a perfectly aligned, homogeneous crowd moving at constant speed (see Fig. 11), the line-of-sight of a motionless subject is interrupted at equal intervals, determined by average body measurements w and s, spacing d, speed v, and angle of incidence θ between the direction vector p and the surface normal n leaving the crowd.

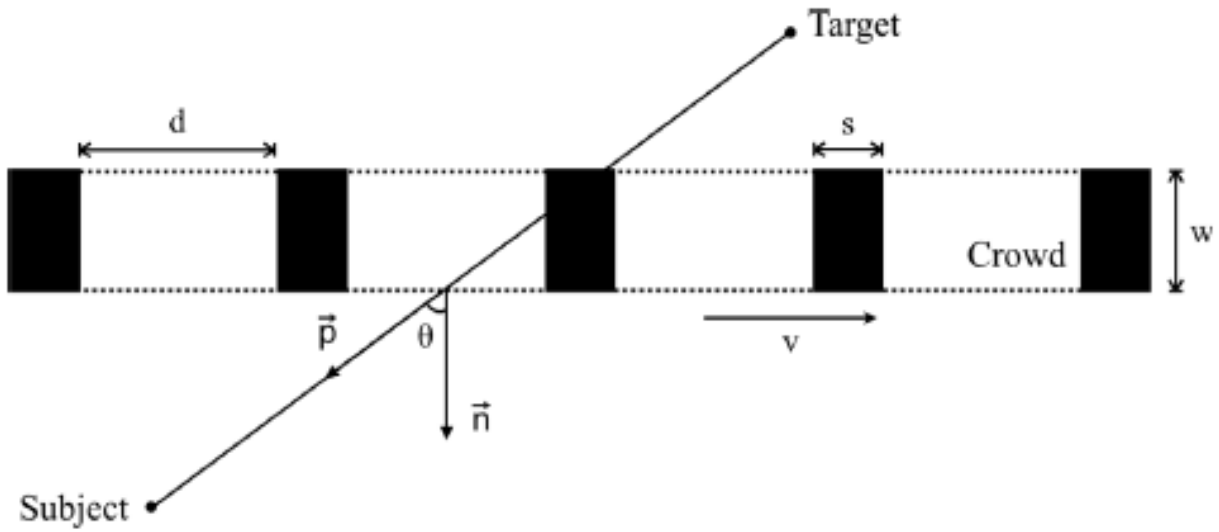


Fig. 11. Intermittent visual disruption of a regularly spaced line of people crowd.

Claim

Supporting argument

Acknowledge alternative view

Response

A crowd between a target and a subject negatively impacts the target's visibility and intelligibility … The severity of this impact is proportional to the occurrence, frequency and duration of each visual disruption by pedestrians in the crowd.

… In case of a perfectly aligned, homogeneous crowd moving at constant speed (see Fig. 11), the line-of-sight of a motionless subject is interrupted at equal intervals, determined by average body measurements w and s, spacing d, speed v, and angle of incidence θ between the direction vector p and the surface normal n leaving the crowd. Such a simplification of crowd motion is only a crude approximation of what actually happens in crowds, as people tend to position themselves behind the shoulders of other pedestrians for a better overview of the situation. Moreover, this simplified representation overlooks important cognitive processes such as adaptive route choices and complex collisional behaviour, e.g. step-and-slide, separation, cohesion, glancing [70].

Nevertheless, this simple model is sufficient to demonstrate that the severity of visibility loss is proportional to the projected crowd density, e.g. if the line of sight from subject to target is perpendicularly intercepted by a directional crowd, the visual disruption can be mitigated mentally (*continually visible*), whereas an angular visual occlusion by the crowd can lead to extended loss of sight (*occasionally visible*), even the sign's unintelligibility [57].

Structuring: Paragraphs

- Traditionally, we regard a paragraph as being based around a single "point" which is expressed in the first sentence. The rest of the paragraph either supports or develops that opening sentence.
- In practice, what constitutes a single "point" is up to the author.

> "A point is a statement that contributes a significant and essential step in your whole structure, without which your conclusion will be weakened. A point like this can't stand on its own without being further unpacked with evidence, explanation, interpretation etc, which is the job of the rest of the paragraph. You should be able to get the overall gist and structure of an academic text by just reading the first line of each paragraph."
>
> *WDC note on paragraphing*

Structuring: Paragraphs

Exhaustive evaluation of all designs in a design space is infeasibly costly for all but the simplest multi-models. For even moderately complex multi-models, it is necessary to hunt for optimal designs by iteratively selecting subsets of designs, evaluating these and using the results to suggest where an optimal design might lie. Genetic Algorithms (GAs), which implement optimisation methods inspired by Darwinian principles of natural selection, may provide significant performance improvements over exhaustive methods. There are, for example, suggestions that GAs are feasible for DSE in the hardware/software codesign domain [13]. However, an implementation of GA-based DSE has not yet been integrated in the INTO-CPS toolchain to permit the evaluation of this approach.

M Rose and J Fitzgerald, Genetic Algorithms for Design Space Exploration in Cyber-Physical Systems: an Implementation in INTO-CPS, Proc. 19th Intl. Overture Workshop 2021

Structuring: Paragraphs

Topic sentence introducing the paragraph's main idea

Main body in which the opening assertion is developed or supported. Note reason and evidence

Concluding remark

Exhaustive evaluation of all designs in a design space is infeasibly costly for all but the simplest multi-models. For even moderately complex multi-models, it is necessary to hunt for optimal designs by iteratively selecting subsets of designs, evaluating these and using the results to suggest where an optimal design might lie. Genetic Algorithms (GAs), which implement optimisation methods inspired by Darwinian principles of natural selection, may provide significant performance improvements over exhaustive methods. There are, for example, suggestions that GAs are feasible for DSE in the hardware/software codesign domain [13]. However, an implementation of GA-based DSE has not yet been integrated in the INTO-CPS toolchain to permit the evaluation of this approach.

Structuring: Paragraphs

- As a test, try reading only the first sentence of each paragraph in a section – the line of the argument should still make sense.

Structuring: Paragraphs

- As a test, try reading only the first sentence of each paragraph in a section – the line of the argument should still make sense.
- Here's the Introduction page 1 of the first example paper …

Crowds are a ubiquitous feature of many built environments, and are inherently dynamic and subject to change. However, crowds are often not modelled on the same ontological level as walls and doors in state-of-the-art building models but the subject of external analyses and simulations that are not *formally* encoded and maintained. This is curious because a building is not an inanimate structure, but a dynamic body where occupants perform tasks in the presence of environmental noises such as sound [1], lighting [2], colour [3], stress [4], and crowds [5].

To give an example, when a crowd emerges, it can become a visual obstacle, a movement barrier, it can modify the qualitative character of the soundscape, and it can interfere with an occupant's personal space. Thus, the emergence and disappearance of crowds can ultimately entail a change in an occupant's experience and behaviour. A fundamental issue from the standpoint of *automated building design analysis* is that there is no uniform treatment of crowds at the level of Building Information Modelling (BIM). In contrast, the intended functions of spaces and zones in a building are often explicitly modelled (e.g. as ''space objects'' in the context of BIM models); crowds could, in principle, also be included as ''objects'' with equally rich semantics in a BIM model.

We therefore argue that the omission of crowds and other dynamic aspects as ''objects'' in the BIM model limits or even undermines the analytical outcomes of mature, powerful methods such as space syntax isovist analysis, and thus can lead to suboptimal facility management and operations, and major difficulties in assessing a building's key performance criteria such as navigability and comfort [6,7].

Furthermore, introducing *crowd objects* into a BIM model can provide a bridge between well established static analysis methods and simulation-based analysis methods. For example, such crowd objects could be hypothesised during the design stage of a building as the result of running hundreds of advanced agent-based simulations, injected back into the BIM model, and subsequently utilised in (static) isovist calculations.

In this paper, we propose a spatio-temporal reasoning-based approach to describe, predict, and explain occupant experiences and behaviour in the presence of crowds. We present a conceptual model

Structuring: Paragraphs

- As a test, try reading only the first sentence of each paragraph in a section – the line of the argument should still make sense.
- Here's the Introduction page 1 of the first example paper …
- Do the first sentences alone (in red) give a flavour of the argument here?

Crowds are a ubiquitous feature of many built environments, and are inherently dynamic and subject to change. However, crowds are often not modelled on the same ontological level as walls and doors in state-of-the-art building models but the subject of external analyses and simulations that are not *formally* encoded and maintained. This is curious because a building is not an inanimate structure, but a dynamic body where occupants perform tasks in the presence of environmental noises such as sound [1], lighting [2], colour [3], stress [4], and crowds [5].

To give an example, when a crowd emerges, it can become a visual obstacle, a movement barrier, it can modify the qualitative character of the soundscape, and it can interfere with an occupant's personal space. Thus, the emergence and disappearance of crowds can ultimately entail a change in an occupant's experience and behaviour. A fundamental issue from the standpoint of *automated building design analysis* is that there is no uniform treatment of crowds at the level of Building Information Modelling (BIM). In contrast, the intended functions of spaces and zones in a building are often explicitly modelled (e.g. as ''space objects'' in the context of BIM models); crowds could, in principle, also be included as ''objects'' with equally rich semantics in a BIM model.

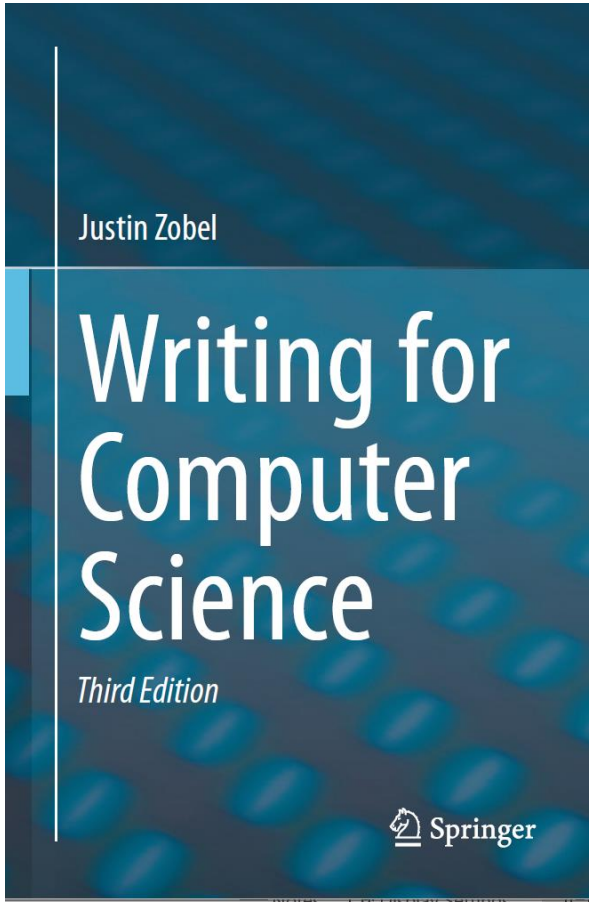
We therefore argue that the omission of crowds and other dynamic aspects as ''objects'' in the BIM model limits or even undermines the analytical outcomes of mature, powerful methods such as space syntax isovist analysis, and thus can lead to suboptimal facility management and operations, and major difficulties in assessing a building's key performance criteria such as navigability and comfort [6,7].

Furthermore, introducing *crowd objects* into a BIM model can provide a bridge between well established static analysis methods and simulation-based analysis methods. For example, such crowd objects could be hypothesised during the design stage of a building as the result of running hundreds of advanced agent-based simulations, injected back into the BIM model, and subsequently utilised in (static) isovist calculations.

In this paper, we propose a spatio-temporal reasoning-based approach to describe, predict, and explain occupant experiences and behaviour in the presence of crowds. We present a conceptual model

Aspects of Computer Science ‘Style’

- Like it or not, style can influence the perception of your work.
- Some elements of detailed style (like use of personal pronouns) are controversial, but consistency is the most important thing. Decide on a convention and stick to it.
- As examples, consider some aspects of CS style:
  - Economy
  - Tone: objective and accurate
  - Voice: active but avoiding the personal



Style: Economy

The volume of information has been rapidly increasing in the past few decades. While computer technology has played a significant role in encouraging the information growth, the latter has also had a great impact on the evolution of computer technology in processing data throughout the years. Historically, many different kinds of databases have been developed to handle information, including the early hierarchical and network models, the relational model, as well as the latest object-oriented and deductive databases. However, no matter how much these databases have improved, they still have their deficiencies. Much information is in textual format. This unstructured style of data, in contrast to the old structured record format data, cannot be managed properly by the traditional database models. Furthermore, since so much information is available, storage and indexing are not the only problems. We need to ensure that relevant information can be obtained upon querying the database.

Example from Zobel, *Writing for Computer Science.* Springer 2014

Style: Economy

The volume of information has been rapidly increasing in the past few decades. While computer technology has played a significant role in encouraging the information growth, the latter has also had a great impact on the evolution of computer technology in processing data throughout the years. Historically, many different kinds of databases have been developed to handle information, including the early hierarchical and network models, the relational model, as well as the latest object-oriented and deductive databases. However, no matter how much these databases have improved, they still have their deficiencies. Much information is in textual format. This unstructured style of data, in contrast to the old structured record format data, cannot be managed properly by the traditional database models. Furthermore, since so much information is available, storage and indexing are not the only problems. We need to ensure that relevant information can be obtained upon querying the database.

Example from Zobel, *Writing for Computer Science.* Springer 2014

Style: Economy

- Economy only comes through extensive revision. It takes time to make text shorter. Rushed and last-minute texts are often long.
- If you get a critique of your writing that suggests misunderstanding, don’t complain about the reader, but ask yourself why the reader misunderstood.

Style: Tone

- The goal here is a level of objectivity and accuracy.
  - Keep the organisation of your text straightforward and logical.
  - Try to use short words unless technically required.
  - CS writing tends to favour short paragraphs and short sentences.
  - Avoid slang and buzzwords.
  - Break the rules when you need to (but only when you need to).
- Bear in mind that not all readers have English as their first language
- … and they don't share the same cultural references
  - "This design change was a bit of an own goal"

Style: Voice

- Try to use an active voice:
  - ✗ Tree structures can be utilized for dynamic storage of terms.
  - ✓ Terms can be stored in dynamic tree structures.
  - ✗ Local packet transmission was performed to test error rates.
  - ✓ Error rates were tested by local packet transmission.
- To avoid twisting text around too much, you can use the special authorial "we", e.g.,

  The following theorem can now be proved.

  might be better as …

  We can now prove the following theorem.

Examples from Zobel

Getting Started on your Project Proposal

- Review the contents of Lecture 11 (Week 2)
- There's a clear top-level structure:
  - Motivation
  - Aim & Objectives
  - Background
  - Diagrammatic Work Plan
  - Brief Explanation of Work Plan
  - References
- If you're finding it challenging to get started, try the time-bounded freewriting exercise mentioned in Lecture 11.

The areas where you'll have an opportunity to do a little academic writing are in red.

Getting Started on your Dissertation

- We will present the elements of the dissertation in CSC3032, closer to when you need it.
- However, if you want to prepare:
  - Extend your background reading, keeping a catalogue of papers as an “annotated bibliography”. For each paper, keep notes on Validity, Synthesis and Relevance.
  - Start a dissertation document with placeholders for:
    - Introduction
    - Background
    - Sections indicating what was done and how (one per objective?)
    - Evaluation
    - Conclusions (including future work).

Reminder – Skills Sessions …

Listed in our Weekly announcement on Teams. Book via the Academic Skills Kit workshops pages (https://www.ncl.ac.uk/academic-skills-kit/events/)

- Wed 16th Feb, 1-2pm: Strategic note-taking
- Thu 17th Feb, 12-1pm: Finding academic information for your literature review
- Thu 17th Feb, 4-5pm: Finding other stuff for your literature review
- Tue 1st Mar, 3-4pm: How to Google well
- Tue 8th Mar, 12-12:45pm: EndNote overview and drop-in
- Wed 9th Mar, 4-4:30pm: Referencing help and tips

CSC3031 - Research and Project Skills

Research Skills: Aspects of Academic Writing in Computer Science

John Fitzgerald