DRAFT
Full paper published in
Proceedings of 8th Workshop on Computational Models
of Natural Argument (CMNA 08)
Patras, Greece, 21July 2008

Representing and appraising Toulmin model arguments in trust cases

Janusz Górski and **Łukasz Cyra** and **Aleksander Jarzębowicz** and **Jakub Miler**[1]

Abstract. The paper presents a Toulmin-based argument model used in trust cases, which allows to argue various properties of IT systems. Argument patterns encountered in trust cases are discussed together with some real-life examples. A method of argument appraisal is introduced together with the corresponding aggregation mechanism. Practical applications of trust cases in industrial and research projects are outlined.

1 INTRODUCTION

Information systems are used almost everywhere, including safety-related and safety-critical applications. Design and implementation of such systems requires a special concern to identify hazardous situations which could result in accidents and mishaps caused by a system in its environment. Examples include systems deployed in healthcare, military, transportation or power plants. It is commonly required by regulations and standards that safety of such system be demonstrated by its manufacturer [1, 2]. The documented body of evidence which summarizes all manufacturer's efforts to analyze safety and mitigate hazards is called a safety case [3].

As demonstration of safety becomes an actual concern, several methodologies of safety case development have been proposed among which Claims-Arguments-Evidence [4], Goal Structuring Notation [5] and Trust-IT [6] can be mentioned and a number of supporting software tools are available [7, 8]. Recently, there is a growing interest in applying 'cases' to demonstrate other aspects of dependability like privacy or security. It has been reflected in current research aiming to extend safety cases into so called trust cases [9] or assurance cases [10].

All cases usually contain a complex, evidence-based justification that a given system meets some specified objectives (like safety, security, privacy etc.) in a given context. The body of the case contains multiple internal dependencies and therefore it is difficult to represent the case as a linear textual document. For instance, a piece of evidence may be referenced from many fragments of the case, arguments may be based on various inference rules, several additional pieces of information may be added to strengthen the argument. A free textual form yields problems in reading maintaining such documents. Therefore cases are usually organized as hierarchical argument structures. Such structures are supposed to express explicitly how conclusions are drawn from available data and the reasoning about more general conclusions is derived from more specific ones. This approach is based on the Toulmin's argument model [11].

Following Toulmin, we have proposed an argument model for trust cases which has been implemented in our Trust-IT framework [9]. Trust-IT framework consists of the language of expressing trust cases, the method of defining arguments (including argument patterns), the process of incremental trust case development in cooperation with stakeholders and the supporting software tool.

Our experience with trust cases resulted in a set of argument patterns which (like software engineering's analytical patterns) represent re-usable structures. Examples of argument patterns we found useful are presented in section 3.

[1] Department of Software Engineering, Gdansk University of Technology, Poland, email: {jango, lukasz.cyra, olek, jakubm}@eti.pg.gda.pl

A case developed for a non-trivial system is a complex structure not easy to read and understood. Assessing such an argument requires expertise and can be a laborious process. To support experts and to provide them with means to express their assessments (and the resulting uncertainty) we have developed an argument appraisal method and implemented it in a tool. The method is discussed in section 4.

2 ARGUMENT MODEL

Trust-IT uses Toulmin's argument model [11] which is also commonly accepted by other approaches to (safety, assurance, trust) cases because of its generality.

2.1 Toulmin's argument model

Toulmin's argument model is presented in Figure 1.

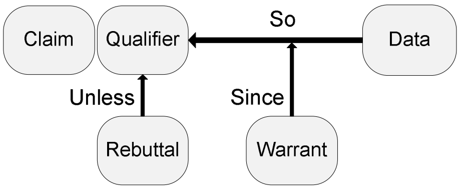


Figure 1. Toulmin's argument model

Toulmin's notation describes the scheme for the structure of a typical argument. It distinguishes:

- the *claim* being a conclusion which is to be demonstrated,
- *data* being facts we appeal to as a foundation for the claim
- the *warrant* which links data and other grounds to the claim
- the *qualifier* representing the degree of confidence that can be placed on the claim, and
- *rebuttal* representing counter-arguments that can be used.

2.2 Trust-IT argument model

The Trust-IT argument model is shown in Figure 2.

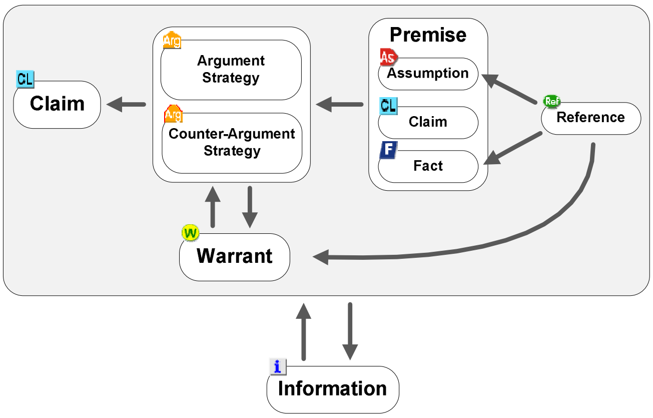


Figure 2. Trust-IT argument model

The Trust-IT argument model closely follows the Toulmin's argument model. The conclusion of the argument is stated in a *claim* (denoted CL). Following [12, 13, 14], we assume that the argument establishes a conveyance relationship between the conclusion and the ground data (the *premises*), on which it is based. The conveyance relationship is represented by the *argument strategy* (denoted Arg), which outlines the basic idea how the conclusion is drawn from the premises. Toulmin's idea of rebuttal of the claim is implemented by *counter-argument strategies* (denoted Arg).

Following Toulmin, every argument strategy has a *warrant* (denoted W), which explains the inference from the premises to the conclusion. The warrant can be self-evident or it can be further justified by its own argument. As we admit warrants which are not necessarily deductively valid, we admit *defeasible reasoning* [15] to be used.

The Toulmin's general concept of *data* is represented by premises of three different types:

- an *assumption* (denoted As), whose justification lies beyond the scope of the entire case (e.g. a trust case),
- a more specific claim which is drawn from another (more specific) argument,
- a *fact* (denoted F) which is considered clearly apprehensible (and appraisable) and does not need to be further justified.

Assumptions and facts can be explained in detail by external evidence, which is pointed to by *references* (denoted Ref). Additionally, an *information* (denoted i) can be attached to any element of the argument. It contains explanatory information, which does not constitute part of the reasoning.

The differences between the Trust-IT model and the Toulmin's model include:

- in Trust-IT, no explicit qualifier is distinguished. This information is, however, modeled using the appraisal mechanism presented in the following part of the paper,
- admittance of explanatory nodes in Trust-IT, i.e. information nodes and references, which proved to be very useful in development of real-world cases,
- different naming of the corresponding elements in both models, and
- stratification of three types of premises (data) in Trust-IT, which play the same role considering the logical structure of the argument, but differ considering the way they are treated during argument development and appraisal.

Arguments in trust cases developed using Trust-IT have a tree-like structure composed of nodes representing elements of the model. Node types are distinguished by different icons, as shown in Figure 2. The parent-child relationship of the nodes is defined according to the logical relationships of the model elements. In Figure 2, an arrow represents that a node of a given type can be a child of a node of the type pointed to by the arrow.

3 ARGUMENT PATTERNS

Argument structures in trust cases can be developed by referring to a catalogue of argument patterns. An argument pattern is a generalization of frequently encountered ways of selecting the supporting premises and composing them into an argument. Once identified, such patterns can be reused in different contexts.

While developing trust cases, we have identified a number of patterns. Some of them follow general argument schemes from the literature [16], while others are more domain-specific. Below, a number of examples of argument patterns are presented using formalized descriptions and illustrated by (less formalized for the sake of understandability) fragments of a trust case for a system for healthcare and well-being services [17]. (We do not present any patterns for counter-argument strategies, as they were seldom used in our trust cases).

Argument from risk analysis

This pattern is based on identifying and mitigating all unacceptable risks that can occur in a given context.

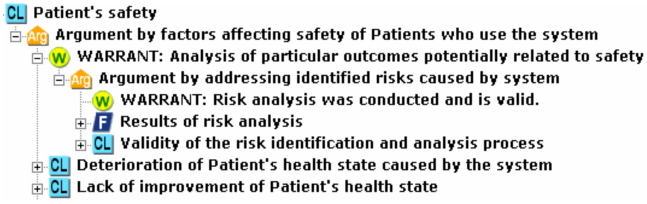
Premises:

1. System O is operated in the environment E
2. U is a user of O
3. Risk analysis of O in E identified a set R of risks potentially affecting U
4. Risk analysis was adequate
5. All risks from R are analyzed and, if classified as unacceptable, mitigated in a demonstrable way

Conclusion: ‘O in E is trustworthy for U’

Warrant: U considers O in E as trustworthy if all identified unacceptable risks potentially affecting U are mitigated. Adequate risk management provides sufficient coverage of relevant risks and adequate risk mitigation supports risk acceptance by U. Therefore, trustworthiness of O for U is plausibly established.

Example: Maintenance of patient’s safety requires mitigation of two unacceptable risks as shown in Figure 3. Such high-level risks are further decomposed with respect to their causes.



```
CL Patient's safety
⊟ Arg Argument by factors affecting safety of Patients who use the system
    ⊟ W WARRANT: Analysis of particular outcomes potentially related to safety
        ⊟ Arg Argument by addressing identified risks caused by system
            W WARRANT: Risk analysis was conducted and is valid.
            ⊞ F Results of risk analysis
            ⊞ CL Validity of the risk identification and analysis process
    ⊞ CL Deterioration of Patient's health state caused by the system
    ⊞ CL Lack of improvement of Patient's health state
```

Figure 3. Example of argument from risk analysis

Argument from compliance with ‘best practices’

This pattern frequently occurs in cases which are related to standards or recommendations. In the domain of system engineering and more specifically software engineering it is commonly accepted that compliance to recommended practices suffices to establish a desired property of the product. This pattern is based on ‘Argument from popular practice’ from [16].

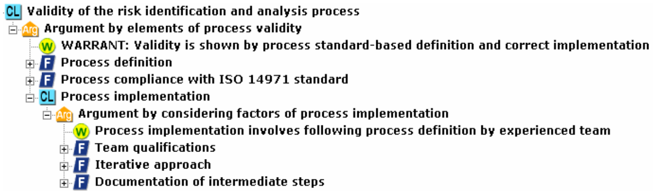
Premises:

1. R is an accepted recommendation
2. R is applicable to an object O in environment E
3. R focuses on establishing a property P
4. Object O in environment E is compliant with R

Conclusion: ‘O in E exhibits P’

Warrant: R encompasses what is recognized as a good and recommended practice to achieve P of O in E. Therefore, compliance with R justifies that P is plausibly true

Example: Arguing validity of the risk assessment process as shown in Figure 4.



```
CL Validity of the risk identification and analysis process
⊟ Arg Argument by elements of process validity
    W WARRANT: Validity is shown by process standard-based definition and correct implementation
    ⊞ F Process definition
    ⊞ F Process compliance with ISO 14971 standard
    ⊟ CL Process implementation
        ⊟ Arg Argument by considering factors of process implementation
            W Process implementation involves following process definition by experienced team
            ⊞ F Team qualifications
            ⊞ F Iterative approach
            ⊞ F Documentation of intermediate steps
```

Figure 4. Example of argument from compliance to best practices

Argument from decomposition

This pattern addresses the situation when complex properties of a system (e.g. quality, dependability) are argued about. Such complex properties need to be decomposed into simpler ones. The decomposition can be followed recursively to the level of metrics that can be measured and directly verified.

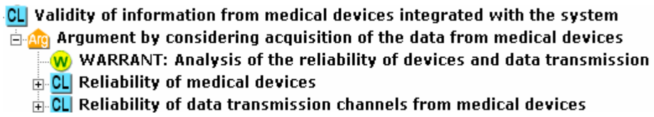
Premises:

1. Object O has property Q1
2. Object O has property Q2

Conclusion: ‘O has property Q’

Warrant: Possessing properties Q1 and Q2 by O most likely is equivalent to possessing property Q by O. Therefore, possessing property Q by O is plausibly true

Example: Arguing validity of information from medical devices as shown in Figure 5.



```
CL Validity of information from medical devices integrated with the system
⊟ Arg Argument by considering acquisition of the data from medical devices
    W WARRANT: Analysis of the reliability of devices and data transmission
    ⊞ CL Reliability of medical devices
    ⊞ CL Reliability of data transmission channels from medical devices
```

Figure 5. Example of argument from decomposition

Argument from expert opinion [14]

In many cases the available evidence requires interpretation or some additional analyses to be made. The opinion of an expert in a given domain can be used as a premise in the argument (given that the competency of this person is proved).

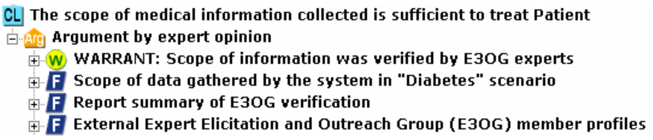
Premises:

1. Person E is an expert in domain D
2. E asserts that A is true
3. A is about D

Conclusion: A

Warrant: Assertions related to D, given by an expert in this domain are plausibly true

Example: Referring to medical experts to assess the scope of medical information collected as shown in Figure 6.



CL The scope of medical information collected is sufficient to treat Patient
- Arg Argument by expert opinion
  - W WARRANT: Scope of information was verified by E3OG experts
  - F Scope of data gathered by the system in "Diabetes" scenario
  - F Report summary of E3OG verification
  - F External Expert Elicitation and Outreach Group (E3OG) member profiles

Figure 6. Example of argument from expert opinion

Multiple argument

It is often useful to strengthen the support of the conclusion by providing several independent arguments which are based on different premises and reasoning. This pattern adapts the 'convergent type of argument' [16].

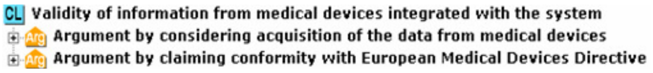
Premises:

1. Argument A1 supports property Q of object O
2. Argument A2 supports property Q of object O

Conclusion: Object O has property Q

Warrant: Warrants for A1 and A2 are mutually independent.

Example: A medical device integrated into a healthcare system can be explicitly analyzed for its reliability, or it can be claimed compliant to European Council's directive concerning medical devices (required to introduce device to the market) as shown in Figure 7.



CL Validity of information from medical devices integrated with the system
- Arg Argument by considering acquisition of the data from medical devices
- Arg Argument by claiming conformity with European Medical Devices Directive

Figure 7. Example of multiple argument

4 ARGUMENT APPRAISAL

In practice, trust cases tend to grow excessively encompassing arguments of various types and evidence of different quality. It makes the assessment of the compelling power of such structures difficult and effort consuming. Research in experimental psychology shows that human minds have difficulties in dealing with complex inference based on uncertain sources of knowledge [10], which is common in trust cases. To deal with this problem, an appraisal mechanism for argument structures was developed [18, 19], which gathers assessments of simple elements of the argument structure (i.e. assumptions, facts and warrants without arguments) and aggregates all the partial assessments into the assessment of the claims.

4.1 Assessment scale

To support experts during the appraisal process two linguistic scales have been introduced, the *Decision scale* and *Confidence scale*. The former provides for expressing the attitude towards acceptance or rejection of the assessed element and contains four decision values: *'acceptable'*, *'tolerable'*, *'opposable'* and *'rejectable'*. The latter provides for expressing the confidence in this decision and distinguishes six levels of confidence: *'for sure'*, *'with very high confidence'*, *'with high confidence'*, *'with low confidence'*, *'with very low confidence'* and *'lack of confidence'*.

The scales can be combined which results in twenty-four values of the *Assessment scale* as shown in Figure 8. The elements of the scale, which are represented as small circles, have intuitively understandable linguistic values.

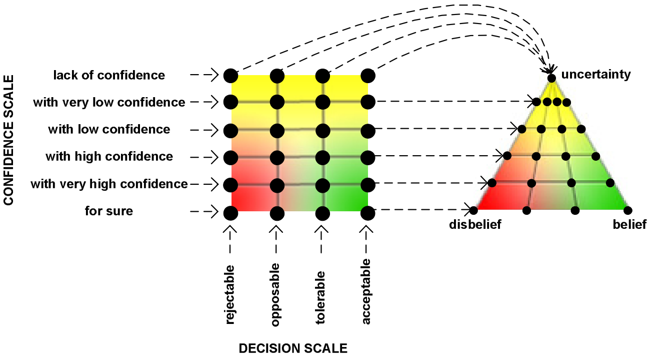


Figure 8. Assessment scale [19]

We can observe, however, that the difference between stating that something is acceptable or rejectable is significant if we have enough (or at least some) evidence supporting such an assessment. For instance, *'for sure acceptable'* or *'with very high confidence rejectable'* needs to be based on evidence. In the case of *'lack of confidence'* the situation is different. Lack of confidence refers to the situation where we do not have any evidence we can refer to and therefore it does not matter which value is chosen from the *Decision scale*. In other words, there is no reason to distinguish between '*with lack of confidence acceptable*' and '*with lack of confidence rejectable*' as both assessments express complete uncertainty about the corresponding decision. This observation has been reflected in Figure 8, where all bullets related to '*lack of confidence*' (on the left) were merged into one (on the right). The result is a triangle which closely corresponds to so-called Josang's opinion triangle [20]. In Josang's opinion triangle *'lack of confidence'* is mapped onto uncertainty. The other vertices of the triangle represent the total disbelief (equivalent to the *'for sure rejectable'* assessment) and the total belief (equivalent to the *'for sure acceptable'* assessment).

4.2 Appraisal procedure

The procedure of argument appraisal is defined as follows:

- **Step 1** – Assess basic warrants (the warrants which do not have explicit arguments) occurring in the argument. This assessment is based on the assessment of the evidence linked to the warrant but also common knowledge and logical bases for the inference.
- **Step 2** - Assess the facts and assumptions occurring in the argument. This appraisal is mostly based on the assessment of the evidence linked to the premises by the reference nodes.

The assessment of a single element (warrant, fact or assumption) proceeds as follows:

- **Step 1** - If no evidence for or against the statement is available the *'lack of confidence'* assessment is issued and the procedure terminates.
- **Step 2** - Otherwise, the ratio between the evidence supporting the acceptance and rejection of the statement is assessed , e.g. a fact can be supported by a reference which shows that some aspects of the fact are true but other are not. In this step this ratio should be estimated and an appropriate value from the *Decision scale* is chosen.
- **Step 3** - It is assessed how much evidence could be additionally provided to become sure about the decision chosen in step 2. This amount of missing evidence drives the selection from the *Confidence scale*.
- **Step 4** - Final assessment from the *Assessment scale* is obtained combining the assessments from steps 2 and 3.

4.3 Aggregation procedure

Aggregation of appraisals to acquire the appraisal of the top claim of a trust case proceeds as follows:

- **Step 1** – For each claim, whose all premises possess an appraisal, aggregate the appraisals of premises and the warrant to obtain the appraisal of the claim.
- **Step 2** - Repeat step 1 until the top claim is reached.

The way in which partial appraisals are aggregated in step 1 depends on the type of warrant occurring in an argument. The analysis of real-world trust cases showed that the great majority of warrants can be rated among four types. These are:

- C-argument - *Complementary argument* is such where the premises provide complementary support for the conclusion. In this case, falsification of one of the premises decreases, but not nullifies, the support for the conclusion. If the remaining premises are accepted, the conclusion can still be attained (possibly with less confidence). The final assessment of the conclusion is a sort of weighted mean value of the contribution of all the premises. As an example see *'Argument from decomposition'* in section 3.
- A-argument - *Alternative argument* is encountered in situations where we have two or more independent justifications of the common conclusion. In A-arguments, when assessments coming from different argument strategies agree, the confidence is reinforced, otherwise, i.e. when they contradict each other, it is decreased. Counter-arguments are treated as arguments showing the contradiction of a claim and their influence on the assessment of the conclusion is opposite to the one of an argument. This type directly corresponds to the *'Multiple argument'* pattern in section 3.
- NSC-argument - *Necessary and Sufficient Condition list argument* is such where the acceptance of all premises leads to the acceptance of the conclusion, whereas rejection of a single premise leads to the rebuttal of the conclusion. Consequently, low assessments of the premises lead to a rapid drop in assessment of the conclusion. As an example see *'Argument from risk analysis'* in section 3.
- SC-argument - *Sufficient Condition list argument* is such where acceptance of the premises leads to the acceptance of the conclusion similar to *NSC-arguments*. The difference to *NSC-argument* is that in this case rejection of a single premise leads to the rejection of the whole inference, i.e. to lack of confidence. The only reasonable conclusion in such a case is that we do not know anything new concerning the validity of this information. As an example see *'Argument from compliance with "best practices"'* in section 3.

Implementation of appraisal aggregation requires that rules are defined for all the above types of warrants occurring in arguments. The appropriate rules are described in detail in [18, 19].

5 CONCLUSIONS

This paper presented selected parts of the Trust-IT framework: the argument model, the argument patterns and the argument appraisal method. The Trust-IT argument model is based on the Toulmin's model and is used in development of argument structures called trust cases. Trust cases extend the idea of safety cases in such a way that their application is limited neither to demonstration of safety nor to computers and software. A trust case can represent an argument supporting claims about various aspects of dependability and many more, including claims like '*Greece is lovely in summer*' or '*X is guilty of crime Y'*,

Trust-IT fully supports representing Toulmin's type arguments. It explicitly represents warrants, which is crucial considering automatic argument processing and reasoning. In contrary to other notations [4, 5] which mainly concentrate on safety, Trust-IT addresses a broad scope of possible properties since the beginning.

To aid in trust case development, we have built a catalogue of argument patterns, from which particular case arguments can be derived. The argument model is associated with the appraisal mechanism, which collects partial assessments from experts and aggregates them into the assessment of the main conclusion. The trust case development is supported by a dedicated software tool called TCT (Trust Case Toolbox) [8]. TCT is a rich internet application (RIA), which supports sharing of all trust case data over the internet and provides for high level of interactivity in trust case editing. The tool also supports the appraisal of trust cases.

The presented approach was applied in a number of projects and research activities:

- A prototype system for drugs distribution and application in a hospital environment (EU 5th FP project DRIVE – 'DRugs In Virtual Enterprise') [6].

- An open platform for e-health services (EU 6th FP project PIPS - 'Personalized Information Platform for life & health Services' [17]).
- An embedded Wireless Sensor Network (WSN) based platform for health and life related services (EU 6th FP project ANGEL – 'Advanced Networked embedded platform as a Gateway to Enhance quality of Life' [21]).
- Standards Conformity Framework (SCF) [22] which supports achievement and assessment of conformity with standards.

More details about the Trust-IT framework and the related research projects can be found at [23].

ACKNOWLEDGEMENTS

This work was partially supported by the EC 6th FR projects: ANGEL (IST project 2005-IST-5-033506-STP) and PIPS (Contract No. 507019 IST2.3.1.11 e-Health).

Contributions by M. Olszewski, M. Witkowicz and M. Nawrot in the development of the TCT tool are to be acknowledged.

REFERENCES

[1] Ministry of Defence Directorate of Standardisation. Defence Standard 00-55 Issue 2: *The procurement of safety critical software in defense systems.*

[2] Railways (Safety Case) Regulations 2000 Health and Safety Executive.

[3] R. Maguire, *Safety cases and safety reports*, Ashgate Publishing Ltd., UK, 2006.

[4] P. Bishop, R. Bloomfield, *A Methodology for Safety Case Development*, Safety-Critical Systems Symp, Birmingham, 1998.

[5] T. Kelly, *Arguing Safety – A Systematic Approach to Managing Safety Cases*, PhD Thesis, University of York, 1998.

[6] J. Górski, A. Jarzębowicz, R. Leszczyna, J. Miler, M. Olszewski, 'Trust Case: justifying trust in an IT solution', *Reliability Engineering and System Safety*, Vol. 89 (2005), pp. 33-47.

[7] Adelard Safety Case Editor (ASCE), Adelard, UK, 2006.

[8] Trust Case Toolbox Editor (http://kio.eti.pg.gda.pl/trust_case/download/TCTEditor_Users_Manual.pdf)

[9] J. Górski, *Trust-IT – a framework for trust cases*, Workshop on Assurance Cases for Security - The Metrics Challenge, DSN 2007 The 37th Annual IEEE/IFIP Intern. Conf. on Dependable Systems and Networks, June 25 - 28, Edinburgh, UK, 2007.

[10] L. Strigini, *Formalism and Judgement in Assurance Cases*, DSN 2004 Workshop on Assurance Cases: Best Practices, Possible Obstacles, and Future Opportunities, Florence, Italy 2004.

[11] S. Toulmin, *The Uses of Argument*, Cambridge University Press, 1958, Updated Edition, 2003.

[12] J. Katzav, C. Reed, 'On Argumentation Schemes and the Natural Classification of Arguments', *Argumentation* 18 (2), 2004, pp. 239-259.

[13] H. Prakken, *AI & Law, Logic and Argument Schemes*, International Congress of Comparative Cultures and Legal Systems, Mexico City, 2004.

[14] D. Walton, 'Justification of Argument Schemes', *Australasian Journal of Logic*, 3, 2005, pp. 1-13.

[15] Stanford Encyclopedia of Philosophy (http://plato.stanford.edu/entries/reasoning-defeasible/)

[16] D. Walton, *Fundamentals of Critical Argumentation*, Cambridge University Press, 2005

[17] PIPS project website - http://www.pips.eu.org

[18] Ł. Cyra, J. Górski, *Supporting expert assessment of argument structures in trust cases*, 9th International Probabilistic Safety Assessment and Management Conference, Hong Kong, 2008.

[19] Ł. Cyra, J. Górski, *Expert assessment of arguments: a method and its experimental evaluation*, Safecomp 2008, Newcastle, 2008 (accepted).

[20] A. Josang, T. Grandison, *Conditional Inference in Subjective Logic*, proc. 6th International Conference on Information Fusion, Cairns, Australia 2003.

[21] ANGEL project website - http://www.ist-angel-project.eu

[22] Ł. Cyra, J. Górski, *Supporting compliance with safety standards by trust case templates*, proc. ESREL 2007, Stavanger, Norway, 2007.

[23] Information Assurance Group homepage - http://iag.pg.gda.pl/iag/