ANGELIC II: An Improved Methodology for Representing Legal Domain Knowledge

Katie Atkinson and Trevor Bench-Capon
Department of Computer Science, University of Liverpool, Liverpool
United Kingdom
{tbc,katie}@liverpool.ac.uk

ABSTRACT

The purpose of this paper is to provide a definitive, up-to-date account of a methodology has that been proven successful for representing and reasoning about legal domains. The ANGELIC (ADF for kNowledGe Encapsulation of Legal Information for Cases) methodology was originally developed to exploit then recent developments in knowledge representation techniques that lend themselves well to capturing factor-based reasoning about legal cases. The methodology is situated firmly within the tradition of research in AI and Law that aims to build systems that are knowledge rich in terms of the domain expertise that is emulated within the systems. When the methodology was first introduced, it was demonstrated on academic examples, but it was subsequently used in and evaluated on a variety of real world domains for external clients. This set of evaluation exercises yielded a variety of learning points as the methodology was applied to different legal domains with their own particular features. These learning points, and the extensions to the methodology that follow from them, urge a consolidation exercise to provide an updated version of the methodology that reflects how it has matured over time. This paper represents a milestone in the development of the methodology in that it presents the ANGELIC II Domain Model, along with a description of its constituent parts, and demonstrates its application through a case study in a key evaluation domain.

CCS CONCEPTS

• **Computing methodologies → Knowledge representation and reasoning**.

KEYWORDS

Legal Knowledge Representation, Methodology, Design

ACM Reference Format:
Katie Atkinson and Trevor Bench-Capon. 2023. ANGELIC II: An Improved Methodology for Representing Legal Domain Knowledge. In *Nineteenth International Conference on Artificial Intelligence and Law (ICAIL 2023), June 19–23, 2023, Braga, Portugal.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3594536.3595137

ICAIL 2023, June 19–23, 2023, Braga, Portugal

ACM ISBN 979-8-4007-0197-9/23/06…$15.00
https://doi.org/10.1145/3594536.3595137

1 INTRODUCTION

The ANGELIC methodology was developed more than six years ago and is most fully described in [1]. It was initially reported in [3], where it was used to model the US Trade Secrets domain [5], the wild animals domain beginning with *Pierson v Post* [9] and the automobile exception to the 4th Amendment [41]. It was then used in a series of applications for the law firm, Weightmans, including Noise Induced Hearing Loss [4], Manual Handling and Occupier Liability. It has also been used in several other projects exploring reasoning with legal cases including US Trade Secrets [16], [15], [19] and the European Convention on Human Rights (ECHR) [13], [30]. The designs produced by the methodology have been implemented using a variety of different languages including Prolog [3], [16], JAVA [13], Javascript [30], and Carneades [19] as well as commercial platforms including those of KIRA Systems [4] and Logiak [12].

Applying the methodology to this variety of domains, aimed at a variety of tasks, meant the methodology needed to be extended to meet new challenges. In particular it was found essential to relate the represented elements to their sources and purposes to support explanation and maintenance, and necessary to provide guidance to users as to how they should ascribe factors, which formed the leaves of the hierarchy in [3] and [1], and which were there taken as givens. Several extensions have proved their worth and the current version of the methodology is now considered stable. Hence we present the extended methodology in this paper.

In Section 2 we present the original methodology as described in [1]. In Section 3 we describe the additional information about sources and values. In Section 4 we describe the information used to elicit descriptions of particular cases from the users. Section 5 summarises the elements of an ANGELIC Domain Model (ADM). Section 6 applies the methodology to the well known US Trade Secrets Domain, providing a complete model of this domain. This domain is chosen for the wealth of comparison is allows: both with other approaches such as CATO [5], IBP [8] and Grabmair's Value Judgement formalism [33], and earlier uses of the ANGELIC methodology [3], [16] and [15]. We finish with some discussion in Section 7 and concluding remarks in Section 8

2 THE ORIGINAL ANGELIC

The ANGELIC methodology follows a top down analysis to encapsulate the knowledge of a specific legal domain in a well defined structure. Any available sources are used: typically these include the law itself, authoritative commentaries (e.g. Clerk and Lindsell [31] for UK Tort Law including Occupiers Liablility) and case digests (e.g. Nelligan [38] for UK Social Security Law), decisions in

leading cases, and, where available, expert knowledge. When working with law firms, it was found that they produce guidance leaflets for clients and guidance manuals and checklists for staff, and these proved particularly useful. Governments and Courts also often issue guidance. For example the European Court of Human Rights issue guidance on the ECHR[1].

The target structure derives from two sources. The basic form is taken from the abstract factor hierarchy developed in CATO [5]. The hierarchy has the outcome as root, *issues* at the upper levels and *base level factors* as the leaves. In between are *abstract factors*, which are used in CATO to assess the significance of distinctions, but are also valuable when providing explanations. *Factors* are ascribed to cases on the basis of the *facts*, and represent a stereotypical pattern of facts which provide a reason to decide an issue for one or other of the parties to the dispute. Factors are either present or absent, and so are essentially Boolean in nature, and always favour a particular side, so that they can be divided into pro-plaintiff and pro-defendant factors. In the original ANGELIC described in [3], cases were represented as bundles of factors and so questions of how factors are ascribed and whether they have magnitudes [35] did not arise. These questions will be discussed in Section 4. Note that the parent-child relation is not "is-a", but that the presence of a child node provides a reason for or against the presence of its parent.

This structure is augmented by associating each node with a set of acceptance conditions. These are based on the acceptance conditions in Abstract Dialectical Frameworks (ADF) [25]. The resulting hierarchy conforms to the definition of an ADF, and many of the works on ANGELIC refer to the structure as an ADF. It does, not. however, make use of the properties of an ADF, and is rather specialised in that all the nodes represent statements and the structure always takes the form of a tree. In the paper we will not call the structure an ADF, but an *ANGELIC Domain Model* (ADM).

The acceptance conditions themselves take a specific form. Each has the form $head \leftarrow body$ where *head* is either ACCEPT or REJECT, and *body* is a conjunction of children of the node in question (possibly negated), or empty, to allow defaults. Variations on this form have been used in the past, including in [3]. One common variation is allowing disjunctions in the body so that conditions with the same head can be grouped together to reduce the number of conditions, but now the conjunction only form is regarded as canonical. The conditions are given in priority order, so that they are tested in turn and when one succeeds it is taken as the result and the others are ignored. The last condition is always a default: it may be either ACCEPT or REJECT, but the body is empty so that there is always a definite result for every node. The acceptance conditions have the form of a procedure in a Prolog program, but are equally readily implemented using conditionals in a procedural language. They also have a straightforward mapping to the argumentation schemes of Carneades as described in [19]

As well as this straightforward translation from design to executable code this structure has advantages which come from its effective modularisation, which aids maintenance as the law changes and new case decisions are made [2]. Finally, the ability to recover a path from the base level factors, through abstract factors and issues, to the outcome provides the basis for effective explanation, both verbal [13] and visual [19]. These points will be discussed further in Section 7.

Although not used in [3], it became established that the best way to present an ADM was in the form of a Table, with a row for each node and columns for the node id, a meaningful name, the children of the node, and the acceptance conditions. See, for example, Table 4 of [15].

3 ADDING INFORMATION

As ANGELIC begun to be used in practical applications with genuine users as stakeholders, it became apparent that the ADM needed to contain more information. In particular it is important to know the source of the information recorded in the ADM. This is essential both for verification when seeking approval for the ADM from domain experts, and for maintenance [29]. If a law is amended or a precedent overturned, it is important to know which, if any, elements of the ADM derived from that law or case.

The source is also important for explanation. While the acceptance conditions provide the equivalent of the warrant in Toulmin's scheme [44], a more satisfying explanation also provides the backing, the reason why the warrant is considered to hold. In particular, to explain why one acceptance condition is preferred to another, it is good to be able to cite the precedent which justified this ordering.

Therefore we extend the Tables recording the ADM with another column to record the source of the acceptance condition, as in [15].

Another aspect absent from the original ANGELIC but which has received considerable attention in accounts of case based reasoning since [22] is purposes or values. Although a precedent case may tell us that one factor should be preferred to another, it does not explain why that preference should hold. Berman and Hafner explained this in [22] by pointing to the legal purpose served by preferring that factor. Subsequent work such as [21] and [32] has tended to speak of values promoted and demoted by the preference. Thus to provide an additional level of explanation, and to motivate preferences not explicitly derived from precedents, the values associated with particular acceptance conditions are also useful. We therefore added another column to record the values concerned and any preference between them. Values may be promoted with varying degrees of strength. Three different degrees of both promotion and demotion were distinguished in [34], but, like [5], we use only ordinary and strong promotion and demotion. Where a factor is strongly promoted or demoted the value is shown with a "+" or a "-" suffix in the Value column

ANGELIC II thus extends the original ANGELIC by adding information as to the source of the various acceptance conditions and their ordering, and the values they promote.

4 ELICITING INFORMATION FROM THE USER

The second area where the original ANGELIC required extension was in factor ascription. In [3] the factors present in a case were taken as given, but ascription of factors demands legal knowledge, with some precedents concerned with factor ascription rather than preferences between factors [17], [18]. Typically users will require

[1]The guidance on Article 6 (Criminal Limb) can be found at https://www.echr.coe.int/documents/guide_art_6_criminal_eng.pdf. It runs to 130 pages, the last 20 of which are links to cited cases.

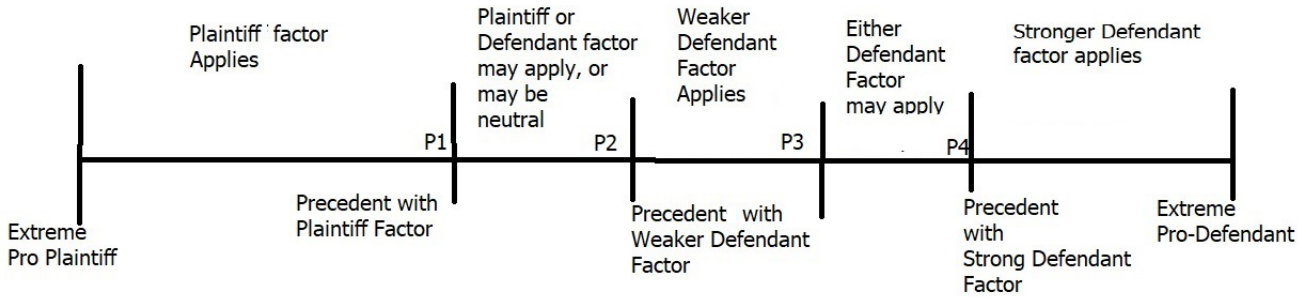


Figure 1: Dimension with three factors (one plaintiff, two defendant) and four precedents.

guidance on how to ascribe factors on the basis of the facts of the case [13].

Since CATO popularised the representation of cases in terms of factors, the move from facts to factors has received little attention. Even formalisations of precedential constraint such as [36], adopt this representation of cases. Before CATO, HYPO [6] had represented cases in terms of their facts and used *dimensions* to bridge between facts and factors. Although dimensions never entirely disappeared (e.g. [20]), factors became very much the focus. Even when the need to represent varying degrees of satisfaction was recognised following [35], these "dimensions" seemed to be treated as a special kind of factor - factors with magnitude - rather than as an intermediary between facts and factors.

For ANGELIC II we return to the relation between dimensions of factors as discussed in [42]. In HYPO cases were represented as bundles of facts. The domain was associated with a number of dimensions. A dimension was an aspect of the case which could take on a range of values, either numeric or enumerated, which favoured the plaintiff at one end and then increasingly favoured the defendant. The facts determined the point on the dimension at which a particular case fell. The point might favour the plaintiff, or the defendant, or neither. Factors were ranges on these dimensions for which the dimension provided a reason to decide for a particular side. The boundaries of these ranges might, however, be disputed, so that it would be possible to argue as to whether a factor applied or not. More than one factor favouring the same side could apply: in this way significant differences of strength could be represented.

An example dimension and its factors is shown in Figure 1. P1-4 show the positions on the dimensions of precedents P1-4. Precedent P1 established a range in which the plaintiff factor definitely applies. The weaker defendant factor applied in precedents P2 and P3. The range between P1 and P2 is currently neutral and no factor applies, but a future case may move the current bounds represented by P1 and P2. The weaker defendant factor applies between P2 and P3, but with precedent P4, where the stronger factor applied, we have a lower bound on this factor. In Figure 1 there are no precedents between P3 and P4, so either of the defendant factors could apply.

Figure 1 shows a very general case of a dimension giving rise to three factors. In practice most dimensions are more limited. In [6] ten of the thirteen dimensions were Boolean: one point giving rise to a factor and the other being neutral. Only one dimension, *Security Measures*, gave rise to both pro-plaintiff and pro-defendant factors,

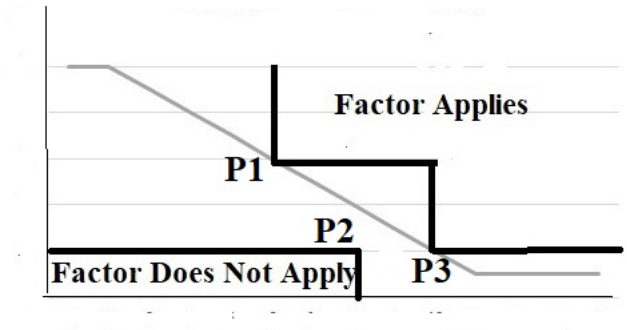


Figure 2: Factor derived from two dimensions. The grey line represents a possible demarcation of the applicability of the factor.

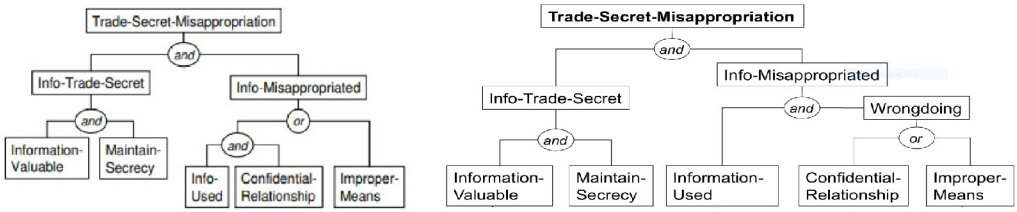
and only one, *Secrets Voluntarily Disclosed*, gave rise to two factors of different strengths for the same side ([17], Table 1). The other non-Boolean dimension was *Competitive Advantage* which depended on two facts: time saved and money saved. These elements can trade off against one another, so that precedents effectively partition a two dimensional space as shown in Figure 2. A number of additional factors were introduced in [5]: these can also be related to additional dimensions ([17], Table 2).

Factor ascription in ANGELIC II is addressed by providing a set of questions to ask the user. These are intended to elicit the facts required to locate the case on the relevant dimensions, so that knowledge of the mapping from dimension points to factors can be applied.

In [18] four methods of ascribing factors were identified. The first simply relies on the standard meaning of the words involved and so can be posed as a direct question to the user. These factors correspond to the Boolean dimensions mentioned above. Often a single question will suffice, but multiple questions are possible if the factor has a more complicated definition.

The second method uses a linear dimension. Here the question determines the point at which the case lies, either by requesting a number, or, for a dimension with a set of enumerated points, asking for the user to select from amongst the possible points.

The third way of ascribing factors is when we have two or more dimensional facts, creating a space in which there are trade offs. In



(a) IBP (b) VJAP

Figure 3: Logical Models of Issues from IBP [8] and VJAP [32]

ANGELIC II, questions are asked for each of the facts, and then a weighted sum is used to determine whether the case falls in to the space corresponding to the applicability of the factor. See [19] for an implementation of weighted sums in Carneades.

The final way of ascribing factors is to use an analogy. This applies when we have a Boolean dimension where the case does not fit the literal meaning of the case, but there are sufficient similarities to allow the factor to be ascribed analogically, or when we wish to extend a range on an enumerated dimension. This method of factor ascription is always likely to be controversial, since the judge is not obliged to accept the analogy, and very difficult to represent computationally in advance, since the determining such similarities would require an infeasibly large ontology to anticipate the many possibilities that could arise [7], [10]. Once an analogy has been made, it can be incorporated as an additional question, but before that one has to rely on the user's interpretation. For example in *Boeing*[2], the defendant had been used by Boeing as a supplier of aircraft windows made to Boeing's own design. Clearly the information had to be given to the defendant so that they could make the windows. After Boeing had terminated the contract, however, the defendant continued to manufacture the windows for sale on the spare parts market, which led Boeing to sue. Now the relevant dimension is *Secrets Voluntarily Disclosed*, with the various points corresponding to the type of people to whom it had been disclosed. Suppose we have as our points: employees, subcontractors, customers, public. The first two will not lead to the ascription of the factor *SecretsDisclosedOutsiders*, since disclosure of the information to these groups of people is essential for the conduct of business, whereas disclosure to the latter two groups will mean that the factor applies. *Suppliers* are not as yet on this list of points. The user must therefore choose between subcontractors and customers, and might well decide that the analogy with subcontractors is stronger, and so choose that, meaning that the factor does not apply. If this is subsequently endorsed by the courts, *suppliers* can be added to the options. Of course, relying on the judgement of the user in this way is not ideal, but is a reasonable pragmatic solution, especially when we are dealing with legally aware people, as is the case for applications directed towards law firms.

[2]The Boeing Company v. Sierracin Corporation, 108 Wash.2d 38, 738 P.2d 665 (1987).

When we have the set of questions which will elicit the key facts from the user we can provide acceptance conditions for the base level factors in terms of the answers to these questions.

5 ANGELIC II ELEMENTS

The complete ANGELIC II ADM thus comprises the following elements:

- **Issue Table**: This has rows for each issue, with columns for the ID, the name, the children, the acceptance conditions, the source of each acceptance condition and the value served by the acceptance condition. The children will be issues, items from the Factor Table, or items from the Base Level Factor Table. The value column will contain either the value(s) served by considering the issue, or the preference between values determining the order of the acceptance conditions. Note that the value order may differ for different issues.
- **Abstract Factor Table**: This has rows for each abstract factor, with columns for the ID, the name, the children, the acceptance conditions, the source of each acceptance condition and the value served by the node. The children will either be abstract factors, or items from the Base Level Factor Table.
- **Base Level Factor Table**: This has rows for each base level factor, with columns for the ID, the name, the associated value, the dimension(s) for the factor, and the acceptance condition in terms of the answers to questions required for the factor to be present.
- **Question List**: This will be a set of questions to ask the user. Questions may require a yes/no answer, a numeric value, or offer a set of possible answers, in which case it should be indicated where one answer or all applicable answers are required.

6 US TRADE SECRETS: A CASE STUDY

In this section we will apply ANGELIC II to US Trade Secrets Law, in order to consider the question of whether a Trade Secret has been misappropriated. This domain has been widely studied, including CATO [5], Issue Based Prediction (IBP) [8], CATE [28] and Value Judgment-based Argumentative Prediction (VJAP) [32] and [17]. Each of these works contains extensive analysis of the domain, and

Table 1: Issue Table for US Trade Secrets. Values may be strongly promoted (indicated by "+").

ID	Issue	Children	Acceptance Conditions	Source	Value
I1	TradeSecretMisappropriation	I2, I3	ACCEPT IF I2 AND I3 REJECT	ROT 757	QM, LM, CA MW RE
I2	InfoTradeSecret	I4, I5	ACCEPT IF I4 AND I5 REJECT	ROT Comment b	MW, RE
I3	InfoMisappropriated	I6, I7	ACCEPT IF I6 AND I7	ROT 757 GEN	QM, LM, CA
I4	InfoValuable	F6p, F8d F11d, F16d InfoObtainable	REJECT IF F11d REJECT IF F8d REJECT IF InfoObtainable ACCEPT IF F6p REJECT IF F16d ACCEPT	Silfen Lewis MBL Mason ROT Comment b	MW > LM+ > RE > LM
I5	MaintainSecrecy	F6p F19d F27d MeasuresOutsiders	REJECT IF F27d REJECT IF F19d ACCEPT IF F6 REJECT IF NOT MeasuresOutsiders ACCEPT	Sheets Robinson Emery ROT Comment b	RE
I6	InfoUsed	InfoMisue, OwnEfforts, F8d, F18p	ACCEPT IF InfoMisue REJECT IF OwnEfforts ACCEPT IF F18p REJECT IF F8d ACCEPT	ROT GEN(a) ROT GEN ROT Comment b(1) ROT Comment b(4)	QM > LM > MW
I7	Wrongdoing	I8, I9	ACCEPT IF I8 ACCEPT IF I9 REJECT	ROT 757(b), (d) ROT 757(a), (c)	LM QM
I8	ConfidfentialRelationship	F23d NoticeConfid, ExplicitAgreement	REJECT IF F23d ACCEPT IF NoticeConfid ACCEPT IF ExplicitAgreement REJECT	ROT GEN (b) ROT GEN (d) ROT Comment b(3)	CA
I9	ImproperMeans	F3d InfoMisuse IllegalAct OwnEfforts	REJECT IF F3d REJECT IF OwnEfforts ACCEPT IF IllegalAct ACCEPT IF InfoMisuse REJECT	Prentice Ferranti Technicon KG	LM > QM

so we will use these to provide our domain expertise. In addition we have the *Restatement of Torts*[3] (ROT), in which section 757 sets out the *liability for disclosure or use of another's Trade Secret*, and the decisions in a large number of cases used in previous works.

6.1 Issues

Section 757 of ROT begins by stating the General principle:

> **GENERAL PRINCIPLE**. One who discloses or uses another's trade secret, without a privilege to do so, is liable to the other if
> (a) he discovered the secret by improper means, or
> (b) his disclosure or use constitutes a breach of confidence reposed in him by the other in disclosing the secret to him, or
> (c) he learned the secret from a third person with notice of the facts that it was a secret and that the third person discovered it by improper means or that the third person's disclosure of it was otherwise a breach of his duty to the other, or
> (d) he learned the secret with notice of the facts that it was a secret and that its disclosure was made to him by mistake.

This general principle is followed by two comments: ROT comment (a) gives a rationale, establishing the values to be promoted by the law, and ROT comment (b) states what is meant by Trade Secret: essentially that the information must be valuable (giving "an opportunity to obtain an advantage over competitors who do not know or use it") and must have been treated as a secret ("a substantial element of secrecy must exist, so that, except by the use of improper means, there would be difficulty in acquiring the information"), Comment (b) then gives a number of things to consider when determining whether the information is a Trade Secret, which are useful for identifying factors.

Section 747 identifies the issues which provide the framework within which cases are considered. The issues identified are the

[3] https://law-journals-books.vlex.com/vid/restatement-first-of-torts-856431041

Table 2: CATO factors grouped by Issues [32]

Issue	Plaintiff Factors	Defendant Factors
InfoValuable	F6p Security Measures F8p Competitive Advantage F15p Unique Product	F11d VerticalKowledge F16d Info Reverse Engineerable F20d Info Known to Competitors F24d Info Obtainable Elsewhere
SecrecyMaintained	F4p Agreed Not To Disclose F6p Security Measures F12p Outsider Disclosures Restricted	F10d Secrets Disclosed Outsiders F19d No Security Measures F27d Disclosure In Public Forum
ImproperMeans	F2p Bribe Employee F7p Brought Tools F14p Restricted Materials Used F22p Invasive Techniques F26p Deception	F3d Employee-Sole-Developer F17d Info Independently Generated F25d Info Reverse Engineered
InfoUsed	F7p Brought Tools F14p Restricted Materials Used F18p Identical Products	F8d NoCompetitive Advantage F17d Info Independently Generated F25d Info Reverse Engineered
ConfidentialRelationship	F5p Agreed-Not-To-Disclose F13p Noncompetition Agreement F21p Knew Info Confidential	F1d Disclosure In Negotiations F4d Agreed Not Specific F23d Waiver of Confidentiality

Table 3: Abstract Factor Table for US Trade Secrets

ID	Issue	Children	Acceptance Conditions	Source	Value
AF1	InfoObtainable	F15p F20d F24d	REJECT IF F15p ACCEPT IF F20d ACCEPT IF F24d REJECT	College Arco Ferranti	LM
AF2	MeasuresOutsiders	F10d F12p	ACCEPT IF F12p REJECT IF F10d ACCEPT	Trandes Robonson	RE
AF3	InfoMisue	F7p F14p	ACCEPT IF F14p ACCEPT IF F7p REJECT	ROT GEN(a) ROT GEN(a)	QM
AF4	OwnEfforts	F17d F25d	ACCEPT IF F17d ACCEPT IF F25d REJECT	Kinnear-Weed Mason	LM
AF5	NoticeConfid	F1d, F21p, F23d	REJECT IF F23d ACCEPT IF F21p REJECT IF F1d ACCEPT	Ecologix Laser Sandlin	CA
AF6	ExplicitAgreement	F4d, F5p F13p	REJECT IF F4d ACCEPT IF F5p ACCEPT IF F13p REJECT	MBL Den=tal-ez ROT Comment b(3)	CA
AF7	IllegalAct	F2d, F23d, F26d	ACCEPT IF F23d ACCEPT IF F26d ACCEPT IF F2d REJECT	ROT GEN (a) ROT GEN (a) ROT GEN (a)	QM

same in both IBP [8] and VJAP [32], but the organisation is a little different as shown in Figure 3. We follow VJAP, which we think more faithfully reflects the text, which talks of the use of the secret before saying that liability arises from either the use of improper means or the breach of a confidential relationship.

Our five values are taken from [28]: the material worth of the information (MW); the nature of the confidentiality agreement (CA);

Q1. Was There An Agreement Between the Plaintiff and the Defendant? (Check all that apply)
(a) There was a Non Competition Agreement
(b) There was a Non Disclosure Agreement
(c) The Agreement did not specifically cover the information
(d) The Plaintiff Waived Confidentiality
(e) The Defendant Knew The information was Confidential
(f) There was No agreement

Q2. To What extent was the Information Known Outside the Plaintiff Organisation? (Check any that apply)
(a) It was a Unique Product
(b) The Information Was Known to Competitors
(c) The information Was Obtainable Elsewhere
(d) The Information Was In The Public Domain

Q3 Was the Information disclosed (Check all that apply)
(a) In negotiations with the Defendant
(b) To employees
(c) To sub-contractors
(d) To customers
(e) To the public
(f) Restrictions were placed on the disclosures
(g) The information was not disclosed

Q5 What role did the information play in the development of the product? (Check all that apply)
(a) The defendant was the sole developer
(b) The information was Independently Generated
(c) The information was Reverse Engineered
(d) An employee brought information from the plaintiff
(e) Restricted Materials were used

Q6 Is the nature of the information vertical (e.g. customer lists) or technical?
(a) Vertical
(b) Technical

Q7 Were the security measures taken? (Check one)
(a) Very strong
(b) Good
(c) Average
(d) Minimal

Q8 How similar was the defendant's product to that of that the plaintiff? (Check one)
(a) identical
(b) Very similar
(c) Some resemblance
(d) Significant differences
(e) Totally dissimilar

Q9 In terms of money, were the savings for the defendant (check one)
(a) Very large
(b) Significant
(c) Small
(d) It was more expensive

Q10 In terms of time, were the savings for the defendant (check one)
(a) Very large
(b) Significant
(c) Small
(d) It took longer

Figure 4: Questions to elicit facts from the user

the efforts taken to maintain secrecy (RE); and whether questionable (QM) or legitimate (LM) means had been used to obtain the information.

We can now produce the Issues Table, shown as Table 1. The Table includes reference to some abstract and base level factors, which will emerge in the next section.

6.2 Factors

Abstract factors play an important role in CATO, where they are used to assess the significance of proposed distinctions, but are absent from both IBP and VJAP, which move directly from issues to base level factors. They do form part of the ANGELIC II methodology, because such intermediate concepts are useful in explanation.

The best way to proceed, however, is first to identify the factors associated with each issue, and then to group these factors into abstract factors as appropriate. The factors associated with each issue in [32] are shown in Table 2[4].

We can now consider which factors should be grouped into abstract factors. The key role of abstract factors as introduced in CATO [5] is to enable distinctions between cases to be downplayed. Sometimes, where a factor present in the precedent is missing from the current case, a factor which *is* present in the current case can be used to substitute for it. For example, if F20d, *InfoKnownToCompetitors*, is missing, F24d, *InfoObtainableElsewhere* might be used to substitute for it. Similarly, where the current case contains an adverse factor not in the precedent, another additional factor can be used to cancel its effect. Suppose F10d, *SecretsDisclosedOutsiders*, was in the new case. This could be cancelled out if F12p, *OutsiderDisclosuresRestricted*, was also present.

This means that the base level factors grouped under an abstract factor must not only concern a similar aspect of the case (e.g. whether the information could be obtained elsewhere), but must also be of comparable strength, so that they are appropriate substitutes, or able to cancel to effect of an adverse factor. In [15], F16d, *InfoReverseEngineerable*, was included, along with F20d and F24d, in the abstract factor *InfoObtainable*. As shown in [19], however, this is an error, since the mere possibility of reverse engineering the information is a much weaker reason for the defendant than the actual availability of information elsewhere. Including F16d in this abstract factor would, for instance, give the wrong answer for *MBL (USA) Corp. v. Diekman*, since F16d, present in a possible precedent found for the plaintiff (e.g. *Mason v Jack Daniels*), would represent an inappropriate substitution for the stronger F20d *InfoKnownToCompetitors* present in *MBL*, which should be found for

[4]Following the analysis of [19], the original F8p *CompetitiveAdvantage*, has been replaced in Tables 1 and 4 by F8d, *NoCompetitiveAdvantage*. It seems that competitive advantage is typically assumed rather than shown, and then when discussed the question is raised by the defendant who wishes to claim that the information had no value because it yielded no competitive advantage.

Table 4: Base Level Factors

ID	Factor	Value	Dimension	Required Answer
F1d	DisclosureInNegotiations	RE	Disclosure	Q3a
F2p	Bribe Employee	QM	IllegalAct	Q4c
F3d	Employee-Sole-Developer	LM	InfoUse	Q5a
F4p	Agreement-not-specific	CA	Agreement	Q1b
F5d	AgreedNotToDisclose	CA	Agreement	Q1c
F6p	AdequateSecurityMeasures	RE	SecurityMeasures	Dim: Q7(a-e) Threshold: (b)
F7p	Brought-Tools	QM	InfoUse	Q5e
F8d	NoCompetitiveAdvantage	MW	Time saved and Money saved	Weighted Sum 2 * Q9 + Q10 < 0.5
F10d	Secrets-Disclosed-Outsiders	RE	Disclosure	Dim: Q3(b-e) Threshold (b)
F11d	Vertical Knowledge	MW	Nature of Info	Q6 vertical
F12p	Outsider-Disclosures-Restricted	RE	Disclosure	Q3f
F13p	NoncompetitionAgreement	CA	Agreement	Q1a
F14p	Restricted-Materials-Used	QM	InfoUse	Q5f
F15p	Unique-Product	LM	CommonKnowledge	Q2a
F16d	Info-Reverse-Engineerable	LM	InfoUse	Q5d
F17d	InfoIndependentlyGenerated	LM	InfoUse	Q5b
F18p	Identical-Products	MW	Similarity	Dim: Q8(a-e) Threshold: (b)
F19d	InadequateSecurityMeasures	RE	Security Measures	Dim: Q7 (a-e)Threshold: (d)
F20d	Info-Known-to-Competitors	LM	CommonKnowledge	Q2b
F21p	Knew-Info-Confidential	CA	Agreement	Q1e
F22p	Invasive-Techniques	QM	IllegalAct	Q4a
F23d	Waiver-of-Confidentiality	CA	Agreement	Q1d
F24d	Info-Obtainable-Elsewhere	LM	CommonKnowledge	Q2c
F25d	Info-Reverse-Engineered	LM	InfoUse	Q5c
F26p	Deception	QM	IllegalAct	Q4b
F27d	Disclosure-In-Public-Forum	RE	CommonKnowledge	Q2d

the defendant. The abstract factors are shown in Table 3 and the base level factors in Table 4.

6.3 Values

For illustration we take our values from [28], although we could have equally well have used [32]. Five values were used in [28]:

- *Confidentiality Agreement* (CA): The purpose here is no ensure that explicit legal agreements are respected. This value encourages making any arrangements explicit through the use of legally binding agreements.
- *Reasonable Efforts* (RE): The plaintiff must have regarded the information as a secret and taken steps to protect it.
- *Legitimate Means* (LM): It is equally clear that innovation and competition should be encouraged and so there should be no punishment if the information is obtained legitimately.
- *Questionable Means* (QM): There are, however, limits on what can be done to obtain the information. Any illegal, or morally dubious, act should be discouraged.
- *Material Worth* (MW): Finally, the information needs to have material value to justify the legal dispute.

These values can be associated with base level factors, as shown in Table 4, and then passed up to their parent nodes when the base level factor is present. The preference order given by the precedents for these values is shown in the value column of Table 1. Note that when considering the issue of whether the information is valuable MW is preferred to LM, whereas when considering the issue of whether the information was used, LM is preferred to MW.

6.4 Questions

We now need a set of questions to elicit the base level factors from the user. The questions are not mapped one to one to factors, but rather grouped by topic, so that one question may relate to several factors. The questions for the Trade Secrets domain are shown in Figure 4. These questions can be used to determine whether the base level factors are present in a case or not, as specified in the required answer column of Table 4.

7 DISCUSSION

ANGELIC II combines the strengths of a factor hierarchy as developed in case based systems such as CATO [5], IBP [26] and VJAP [32] with rules as pioneered in [43]. The hierarchy provides the structure to enable the various layers of legal reasoning set out in [30] – from facts to factors to issues to outcome – to be represented separately, and for the case do be decided issue by issue so as to avoid irrelevant distinctions [17]. Writing the acceptance conditions as rules makes them explicit and precise. Moreover, it is possible to use the rules differently at different layers: those determining the outcome in terms of issues are standard logical rules providing necessary and sufficient conditions; those resolving issues in terms of factors are defeasible rules prioritised in the manner of [39];

while those moving from facts to factors can also use arithmetical expressions to handle reasoning with dimensions [16].

It is the fusion of structure and rules that makes the ADM especially suitable for representing legal domain knowledge. We now discuss some particular aspects, starting with two essential requirements on a legal knowledge representation technique: that it supports *maintenance* and that it supports *explanation* [30].

7.1 Maintenance

The need to build systems in the legal domain so that they are able to readily adapt to changes in the law has long been recognised [24]. Unlike many other domains, the law is subject to constant change. Not only is legislation frequently amended, but case law changes to meet new situations and to reflect evolving social values.

A key to maintenance in any software system (as discussed in, for example, [40]) is a modular design with coherent, loosely coupled modules. The nodes of the ADM provide an excellent example of such modularity. They provide coherent modules in that they hold all the information about one particular statement, and nothing else, and the coupling is minimal since they are called only by their parent node, limiting their interaction and providing a uniform interface between them. This means that nodes can be amended without affecting other nodes and expanded into a sub-tree if and when more detail is required. Similarly a sub-tree can be pruned without jeopardising the rest of the model if a node is considered no longer relevant.

The modular structure also permits reuse. In law, domains often have common features. For example, in several of the insurance applications developed for commercial settings claims are time-limited. This meant that the sub-tree regarding time limits developed for the first application, Noise Induced Hearing Loss [4], could be detached and reused in subsequent domains.

An ADM also facilitates discussion with, and verification by, domain experts. The hierarchy itself gives an excellent picture of how the concepts have been decomposed and the acceptance conditions are small groups of rules which assist focus and understanding.

Finally, in ANGELIC II every acceptance condition is associated with its provenance, the particular piece of legislation or case from which it was derived. This means that when legislation is changed, or a decision overruled, the affected parts of the ADM can be readily identified. This fulfils the requirements of "isomorphism" as advocated in [29].

7.2 Explanation

Explanation, both for transparent deliberation and justification of recommendations, is essential for any legal support tool: it is of no benefit to know the outcome without being able to say why that outcome is the right one (and why other options are not) [11].

The ADM supports a step by step justification of each node in terms of its children, following the traditional *how?* explanation of rule based expert systems from MYCIN [27] onwards, and adopted by rule based expert systems in law such as [43]. This is the form of explanation used in [3] and [30]. The explanation can be presented either bottom up, starting with the facts, or top down, starting withe the outcome. It can be delivered all at once, or step by step so that the users can stop when they are satisfied. The explanation can also be presented visually as an argument graph, as was shown in [19].

The ADM can also be used to construct explanations following the Issue-Rule-Application-Conclusion (IRAC) method widely used in US law schools, as described in [14]. Here the focus is on nodes with both a pro and a con acceptance condition satisfied, and the explanation is in terms of which has priority. This customises the explanation to the particular case by drawing attention to the contested points in that case.

Finally ANGELIC II is able to augment the explanation by supplying the backing for any step (either a statutory reference or a precedent case), and by supplying a rationale in terms of the values promoted and the preferences between them.

7.3 Implementation

The ADM also readily supports implementation as a logic program (e.g. as shown in [3], [16]), an imperative program (e.g. as a Java implementation [13]), a web-based application (e.g. using Javascript [30]), or a set of argumentation schemes (e.g. using Carneades [19]). Moreover, the modularity of the ADM means that particular branches can be prototyped independently, and stubs can be used while the detail is fleshed out, facilitating top down development. This means that prototypes can be used in the development process, for testing and debugging as described in [19], and as a means of animating the model for discussion and refinement with experts and users.

These prototypes can be worked up into robust applications to provide a range of support tools of the sort described in [4], [12], [19], and [30], which can be delivered either as stand alone programs or as web applications.

All the features that we have highlighted in this section are grounded in principles for the sound design and production of decision support software. Underlying theory is essential for ensuring that the AI techniques used are performing effectively the tasks required of them. Practical considerations are equally important if we are to turn the theory into deployed solutions that legal professionals are willing and able to use.

7.4 Links with Related Work

We close our discussions by noting various links of importance between ANGELIC II and other related work. ANGELIC II embodies a number of technical insights. The structure as a whole is an Abstract Dialectical Framework [25], ensuring that the knowledge representation has a sound grounding. The structure permits logical modelling of the relevant law ([43], [26]) at the upper levels and factor based reasoning [5] at the lower levels. As a whole it forms an abstract factor hierarchy [5]. Precedents are modelled as prioritised rules as in [39], to permit multi step arguments. From a software engineering perspective the local acceptance conditions defined for nodes provide modularity [40] and the provenance of information is captured by linking to sources [29]. Values ([21], [32]) are also associated with the acceptance conditions.

8 CONCLUDING REMARKS

This paper has set out the current state of a successful methodology for capturing legal domain knowledge, building on lessons

garnered from a number of exercises conducted in both research and commercial environments, which demonstrate the broad applicability of the methodology to substantial real world legal domains. We have described the key learning points that necessitated the original methodology being adapted and the changes that were required to ensure that the methodology is sufficiently robust to be able to be used in domains that target end users operate in. End users need to have confidence in the application of new research in practice and the description of ANGELIC II we have provided in this paper is intended to serve as a solid reference point for our expanding suite of applications.

Recent pilot studies with target end users [30] have confirmed the viability of using the ANGELIC methodology in practical scenarios, with a key feature of the methodology that makes it attractive to legal professionals being the explanation features provided by tools developed using the methodology. Such features are enabled precisely thanks to the knowledge representation techniques that are exploited within the methodology. Use of these techniques also means that the models are intended to capture the legal reasoning that is conducted by the human domain experts who decide the cases, in contrast to machine learning approaches based on classification techniques. However, we do see a role for machine learning within the wider task of building AI systems for legal decision support, most notably using it for the task of automated factor ascription, of the flavour described in [23] and [37]. The focus of our future work will be on maturing the factor ascription task to enrich the automated support available for fully deploying systems built using the ANGELIC II design methodology in practice.

REFERENCES

[1] Latifa Al Abdulkarim. 2017. *Representation of case law for argumentative reasoning.* Ph.D. thesis. The University of Liverpool.

[2] Latifa Al-Abdulkarim, Katie Atkinson, and Trevor Bench-Capon. 2016. Accommodating change. *Artificial Intelligence and Law* 24, 4 (2016), 409–427.

[3] Latifa Al-Abdulkarim, Katie Atkinson, and Trevor Bench-Capon. 2016. A methodology for designing systems to reason with legal cases using ADFs. *Artificial Intelligence and Law* 24, 1 (2016), 1–49.

[4] Latifa Al-Abdulkarim, Katie Atkinson, Trevor Bench-Capon, Stuart Whittle, Rob Williams, and Catriona Wolfenden. 2019. Noise induced hearing loss: Building an application using the ANGELIC methodology. *Argument and Computation* 10, 1 (2019), 5–22.

[5] Vincent Aleven. 1997. *Teaching case-based argumentation through a model and examples.* Ph.D. thesis. University of Pittsburgh.

[6] Kevin D Ashley. 1990. *Modeling legal arguments: Reasoning with cases and hypotheticals.* MIT press.

[7] Kevin D Ashley. 2009. Ontological requirements for analogical, teleological, and hypothetical legal reasoning. In *Proceedings of the 12th International Conference on Artificial Intelligence and Law*. 1–10.

[8] Kevin D Ashley and Stefanie Brüninghaus. 2009. Automatically classifying case texts and predicting outcomes. *Artificial Intelligence and Law* 17, 2 (2009), 125–165.

[9] Katie Atkinson (Ed.). 2012. *Artificial Intelligence and Law: Special Issue on Modelling Popov v Hayashi.* Vol. 20:1.

[10] Katie Atkinson and Trevor Bench-Capon. 2019. Reasoning with legal cases: Analogy or rule application?. In *Proceedings of the 17th International Conference on Artificial Intelligence and Law*. 12–21.

[11] Katie Atkinson, Trevor Bench-Capon, and Danushka Bollegala. 2020. Explanation in AI and law: Past, present and future. *Artificial Intelligence* 289 (2020), 103387.

[12] Katie Atkinson, Trevor Bench-Capon, Tom Routen, Alejandro Sánchez, Stuart Whittle, Rob Williams, and Catriona Wolfenden. 2019. Realising ANGELIC designs using Logiak. In *Proceedings of JURIX 2019*. 151–156.

[13] Katie Atkinson, Joe Collenette, Trevor Bench-Capon, and Kanstantsin Dzehtsiarou. 2021. Practical tools from formal models: the ECHR as a case study. In *Proceedings of the 18th International Conference on Artificial Intelligence and Law*. 170–174.

[14] Trevor Bench-Capon. 2020. Explaining legal decisions using IRAC. In *Proceedings of CMNA 2020: CEUR Workshop Proceedings*, Vol. 2669. 74–83.

[15] Trevor Bench-Capon. 2021. Using issues to explain legal decisions. *arXiv preprint arXiv:2106.14688, Presented at XAILA 2021* (2021).

[16] Trevor Bench-Capon and Katie Atkinson. 2018. Lessons from Implementing Factors with Magnitude.. In *Proceedings of JURIX 2018*. 11–20.

[17] Trevor Bench-Capon and Katie Atkinson. 2021. Precedential constraint: The role of issues. In *Proceedings of the 18th International Conference on Artificial Intelligence and Law*. 12–21.

[18] Trevor Bench-Capon and Katie Atkinson. 2022. Argument Schemes for Factor Ascription. In *Proceedings of COMMA 2022*. IOS Press, 68–79.

[19] Trevor Bench-Capon and Thomas F. Gordon. 2022. Implementing a Theory of a Legal Domain. In *Proceedings of JURIX 2023*. 13–22.

[20] Trevor Bench-Capon and Edwina L Rissland. 2001. Back to the future: dimensions revisited. In *Proceedings of JURIX 2001*. 41–52.

[21] Trevor Bench-Capon and Giovanni Sartor. 2003. A model of legal reasoning with cases incorporating theories and values. *Artificial Intelligence* 150, 1-2 (2003), 97–143.

[22] Donald H Berman and Carole L Hafner. 1993. Representing Teleological Structure in Case-based Legal Reasoning: The Missing Link. In *Proceedings of the 4th International Conference on Artificial Intelligence and Law*. 50–59.

[23] L Karl Branting, Craig Pfeifer, Bradford Brown, Lisa Ferro, John Aberdeen, Brandy Weiss, Mark Pfaff, and Bill Liao. 2021. Scalable and explainable legal prediction. *Artificial Intelligence and Law* 29, 2 (2021), 213–238.

[24] Paul Bratley, Jacques Frémont, Ejan Mackaay, and Daniel Poulin. 1991. Coping With Change. In *Proceedings of the 3rd International Conference on Artificial Intelligence and Law*. 69–76.

[25] Gerhard Brewka and Stefan Woltran. 2010. Abstract Dialectical Frameworks. In *12th International Conference on the Principles of Knowledge Representation and Reasoning*. 102–111.

[26] Stephanie Brüninghaus and Kevin Ashley. 2003. Predicting outcomes of case based legal arguments. In *Proceedings of the 9th International Conference on Artificial Intelligence and Law*. ACM, 233–242.

[27] Bruce G Buchanan and Edward H Shortliffe. 1984. *Rule based expert systems: the Mycin experiments of the Stanford heuristic programming project.* Addison-Wesley Longman Publishing Co., Inc.

[28] Alison Chorley and Trevor Bench-Capon. 2005. An empirical investigation of reasoning with legal cases through theory construction and application. *Artificial Intelligence and Law* 13, 3 (2005), 323–371.

[29] Frans Coenen and Trevor Bench-Capon. 1991. Exploiting isomorphism: development of a KBS to support British Coal insurance claims. In *Proceedings of the 3rd International Conference on Artificial Intelligence and Law*. 62–68.

[30] Joe Collenette, Katie Atkinson, and Trevor Bench-Capon. 2023. Explainable AI Tools for Legal Reasoning about Cases: A Study on The European Court of Human Rights. *Artificial Intelligence* (2023), 103861.

[31] Andrew Tettenborn (ed). 2022. *Clark and Lindsell on Torts, 23rd Edition.* Sweet and Maxwell.

[32] Matthias Grabmair. 2016. *Modeling Purposive Legal Argumentation and Case Outcome Prediction using Argument Schemes in the Value Judgment Formalism.* Ph. D. Dissertation. University of Pittsburgh.

[33] Matthias Grabmair. 2017. Predicting trade secret case outcomes using argument schemes and learned quantitative value effect tradeoffs. In *Proceedings of the 16th International Conference on Artificial Intelligence and Law*. 89–98.

[34] Matthias Grabmair and Kevin D Ashley. 2011. Facilitating case comparison using value judgments and intermediate legal concepts. In *Proceedings of the 13th International conference on Artificial Intelligence and Law*. 161–170.

[35] John F Horty. 2019. Reasoning with dimensions and magnitudes. *Artificial Intelligence and Law* 27, 3 (2019), 309–345.

[36] John F Horty and Trevor Bench-Capon. 2012. A factor-based definition of precedential constraint. *Artificial Intelligence and Law* 20, 2 (2012), 181–214.

[37] Jack Mumford, Katie Atkinson, and Trevor Bench-Capon. 2022. Reasoning with Legal Cases: A Hybrid ADF-ML Approach. In *Proceedings of JURIX 2022*. 93–102.

[38] Desmond Neligan and Charles A Emanuel. 1989. *Social Security Case Law: Digest of Commissioners' Decisions.* HM Stationery Office.

[39] Henry Prakken and Giovanni Sartor. 1998. Modelling Reasoning with Precedents in a Formal Dialogue Game. *Artificial Intelligence and Law* 6, 2-4 (1998), 231–287.

[40] Roger S Pressman. 2005. *Software engineering: a practitioner's approach.* Palgrave Macmillan.

[41] Edwina L Rissland. 1989. Dimension-based analysis of hypotheticals from supreme court oral argument. In *Proceedings of the 2nd International Conference on Artificial Intelligence and Law*. 111–120.

[42] Edwina L Rissland and Kevin D Ashley. 2002. A note on dimensions and factors. *Artificial Intelligence and Law* 10, 1 (2002), 65–77.

[43] Marek J. Sergot, Fariba Sadri, Robert A. Kowalski, Frank Kriwaczek, Peter Hammond, and H Terese Cory. 1986. The British Nationality Act as a logic program. *Commun. ACM* 29, 5 (1986), 370–386.

[44] Stephen E Toulmin. 1958. *The uses of argument.* Cambridge University Press.