# Applications of Influence Diagrams to Information Retrieval

Luis M. de Campos
Juan F. Huete
Juan M. Fernández-Luna

Departamento de Ciencias de la Computación e Inteligencia Artificial
E.T.S.I. Informática y Telecomunicaciones, Universidad de Granada
Spain



# Outline

- Structured Information Retrieval.
- Bayesian network-based models.
- Why probabilities are not enough? Making decisions.
- Influence diagrams.
- Influence diagram-based models for structured information retrieval.
  - Topology.
  - Inference.
- Conclusions.

# Structured Information Retrieval

Classical Information Retrieval methods treat documents as they were atomic entities:

| Document Example |
|---|
| *Information Retrieval on the Web.* |
| The application to systems ... |
| 1.- Introduction and Overview. |
| Although the litterature on specific implementation ... |
| 2.- Implementation and Efficiency of the Model. |
| The first kind of related activity .. |



# Structured Information Retrieval

Classical Information Retrieval methods treat documents as they were atomic entities:

| Document Example |
|---|
| *Information Retrieval on the Web.* |
| The application to systems ... |
| 1.- Introduction and Overview. |
| Although the litterature on specific implementation ... |
| 2.- Implementation and Efficiency of the Model. |
| The first kind of related activity .. |

- Documents are indexed as a single unit.
- The IR system retrieves complete documents.
- The different document components are not taken into account.

# Structured Information Retrieval

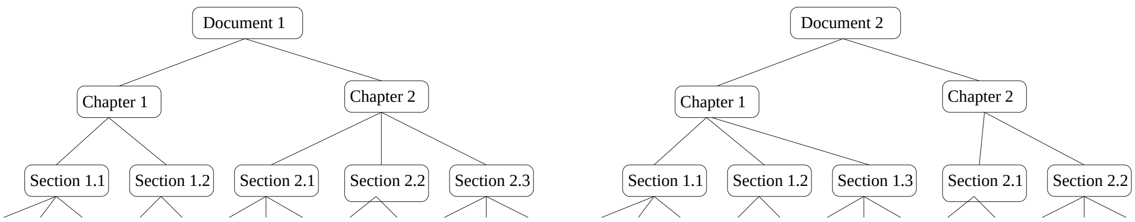
Modern methods have to deal with structured documents, whose content is organized around a well defined structure (SGML, XML, MPEG-7, books, scientific papers,...):

```
<?xml version="1.0" encoding="ISO-8859-1"?>
    <article>
        <title> Information Retrieval on the Web. </title>
        <parag> The application to systems ... </parag>
        <section>
            <title>Introduction and Overview </title>
            <parag> Although the litterature on specific implementation ...</parag>
        </section>
        <section>
            <title>Implementation and Efficiency of the Model </title>
            <parag> The first kind of related activity </parag>
        </section>
    </article>
```



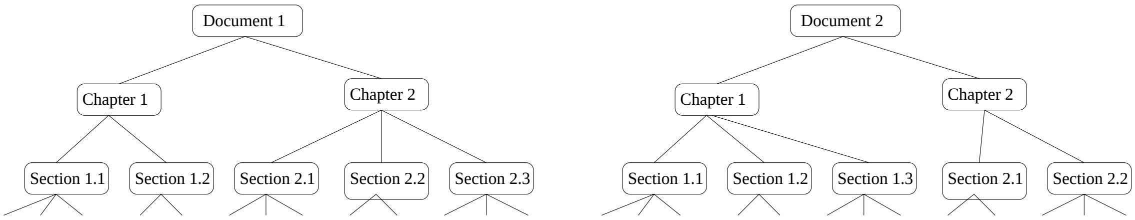
# Structured Information Retrieval (2)

The inclusion of the structure of a document affects the design and implementation of an IR system:



# Structured Information Retrieval (2)

The inclusion of the structure of a document affects the design and implementation of an IR system:

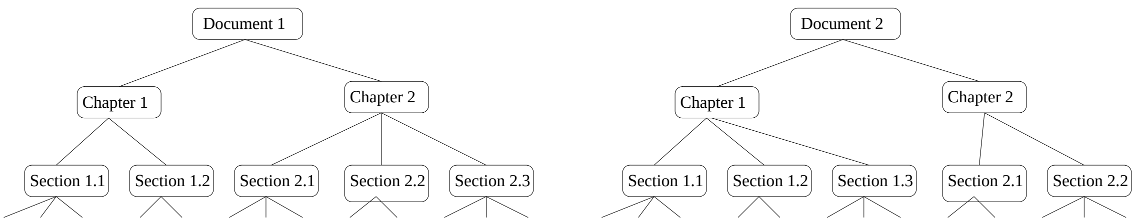



- Indexing: users can search both by content (what) and structure (where).



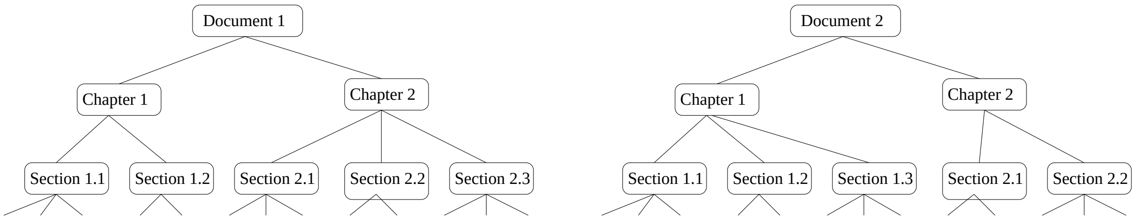
# Structured Information Retrieval (2)

The inclusion of the structure of a document affects the design and implementation of an IR system:




- Indexing: users can search both by content (what) and structure (where).
- Retrieval: use both structure and content to estimate the relevance of documents or document components.

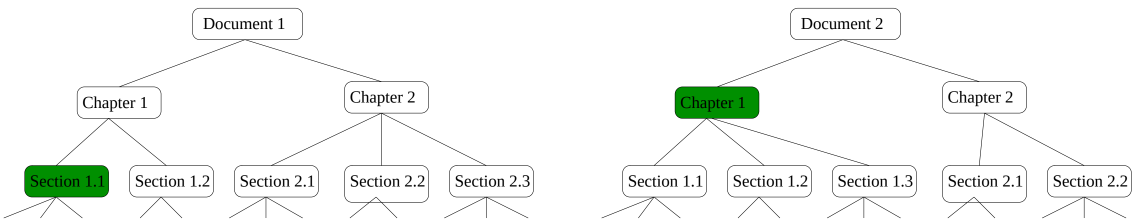
# Structured Information Retrieval (3)




Goal: To retrieve, dynamically,



# Structured Information Retrieval (3)

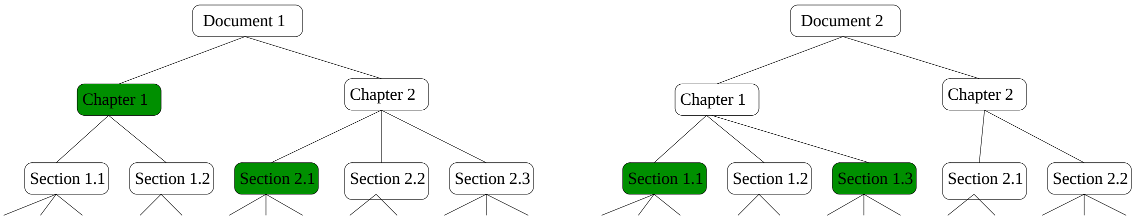



Goal: To retrieve, dynamically,

- Document components of varying complexity.

Information Retrieval and Influence Diagrams. The Future of Web Search Workshop – p.6/44

# Structured Information Retrieval (3)

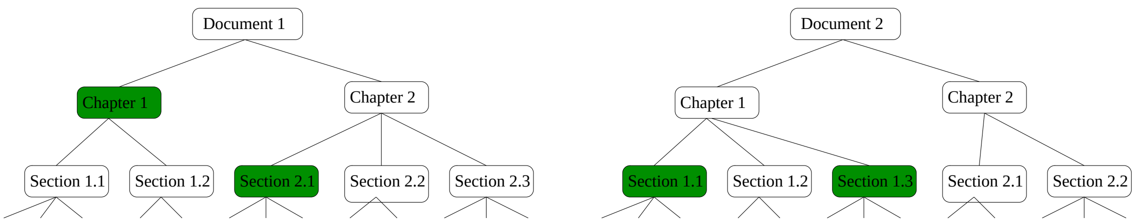



Goal: To retrieve, dynamically,

- Document components of varying complexity.
- Different parts (or *entry points*) of the same document, corresponding to different structural elements.



# Structured Information Retrieval (3)




Goal: To retrieve, dynamically,

- Document components of varying complexity.
- Different parts (or *entry points*) of the same document, corresponding to different structural elements.

As classical IR, structured IR is also an inherently uncertain process.

- The dominant formalism to deal with probabilistic uncertainty is Bayesian networks.

# Bayesian Networks

A Bayesian network (BN) has two components:

- Qualitative, directed acyclic graph $G = (V, E)$.
- Quantitative, conditional probability distributions.



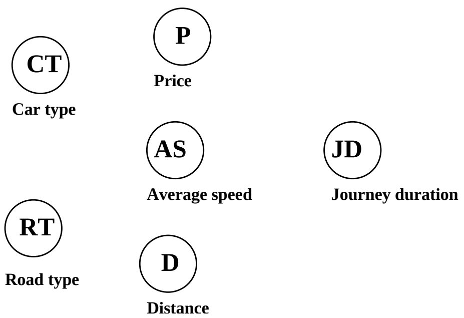
# Bayesian Networks

A Bayesian network (BN) has two components:

- Qualitative, directed acyclic graph $G = (V, E)$.
- Quantitative, conditional probability distributions.

Qualitative component:

- Nodes in $V$: random variables in the domain.



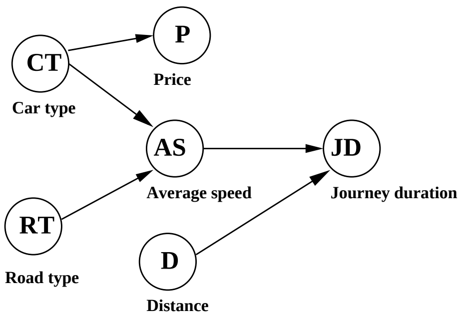
# Bayesian Networks

A Bayesian network (BN) has two components:

- Qualitative, directed acyclic graph $G = (V, E)$.
- Quantitative, conditional probability distributions.

Qualitative component:

- Nodes in $V$: random variables in the domain.
- Arcs in $E$: causality, relevance or direct dependence relationships between variables





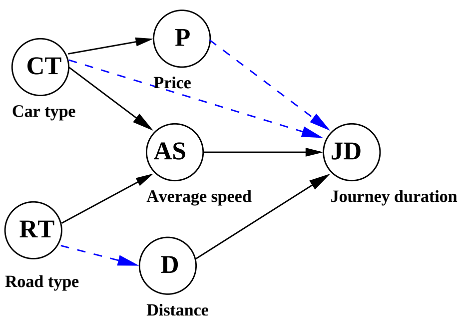

# Bayesian Networks

A Bayesian network (BN) has two components:

- Qualitative, directed acyclic graph $G = (V, E)$.
- Quantitative, conditional probability distributions.

Qualitative component:

- Nodes in $V$: random variables in the domain.
- Arcs in $E$: causality, relevance or direct dependence relationships between variables (lack of arcs means conditional independence relationships and indirect dependence relationships).



# Bayesian Networks (2)

A Bayesian network (BN) has two components:

- Qualitative, directed acyclic graph $G = (V, E)$
- Quantitative, conditional probability distributions

Quantitative component:



# Bayesian Networks (2)

A Bayesian network (BN) has two components:

- Qualitative, directed acyclic graph $G = (V, E)$
- Quantitative, conditional probability distributions

Quantitative component:

- Probability distributions are specified locally for each node.
- They measure the uncertainty about the dependence relationships.

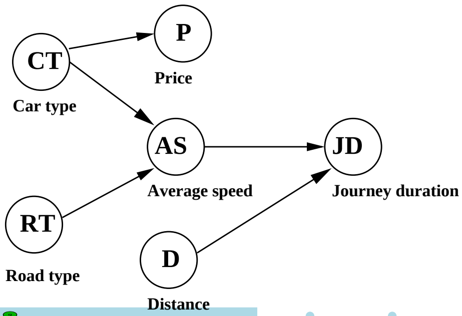
# Bayesian Networks (2)

A Bayesian network (BN) has two components:

- Qualitative, directed acyclic graph $G=(V,E)$
- Quantitative, conditional probability distributions

## Quantitative component:

- Probability distributions are specified locally for each node.
- They measure the uncertainty about the dependence relationships.
- Each variable $X$ stores a family of probability distributions $p(X|pa(X))$, one for each configuration $pa(X)$ of its parent set $Pa(X)$.




p(JD|AS,D)
p(D)
p(AS|CT,RT)
p(P|CT)
p(RT)
p(CT)

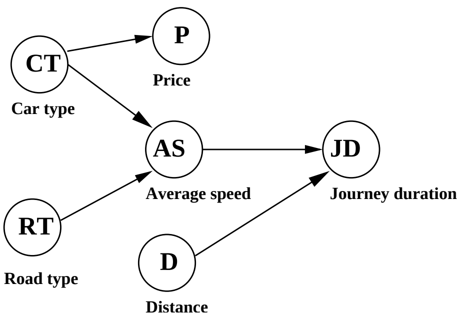


# Bayesian Networks (3)

Joint distribution: Product of the local distributions:

$$p(CT,RT,P,AS,D,JD) = p(CT)\times p(RT)\times p(P|CT)\times p(AS|CT,RT)\times p(D)\times p(JD|AS,D)$$

Inference in Bayesian networks: based on the factorization of the joint distribution, to carry out local computations $\rightarrow$ efficiency



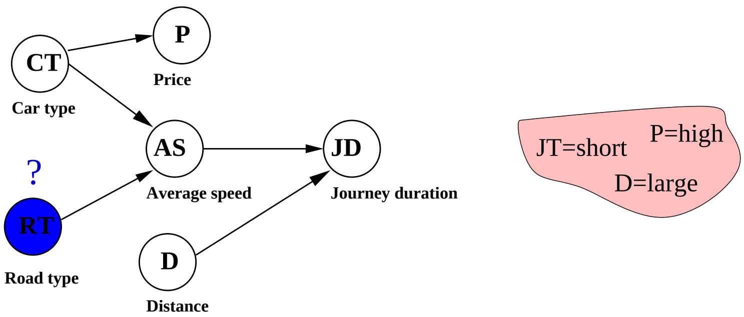
# Bayesian Networks (3)

Joint distribution: Product of the local distributions:

$$p(CT,RT,P,AS,D,JD) = p(CT) \times p(RT) \times p(P|CT) \times p(AS|CT,RT) \times p(D) \times p(JD|AS,D)$$

Inference in Bayesian networks: based on the factorization of the joint distribution, to carry out local computations → efficiency

Posterior probability of any variable or set of variables given some evidence





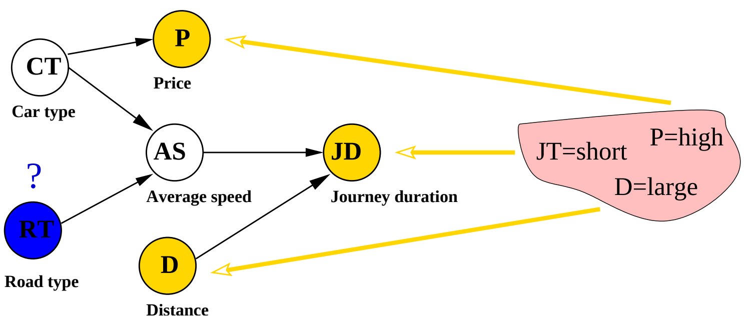

# Bayesian Networks (3)

Joint distribution: Product of the local distributions:

$$p(CT,RT,P,AS,D,JD) = p(CT) \times p(RT) \times p(P|CT) \times p(AS|CT,RT) \times p(D) \times p(JD|AS,D)$$

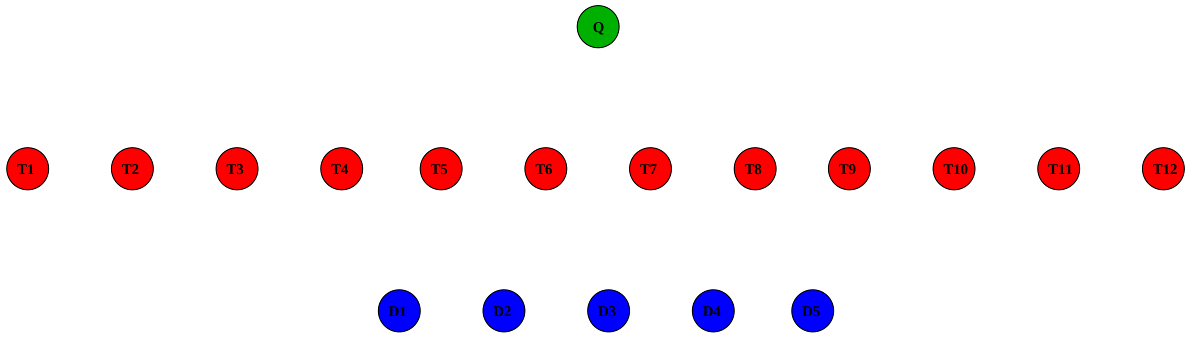
Inference in Bayesian networks: based on the factorization of the joint distribution, to carry out local computations → efficiency

Posterior probability of any variable or set of variables given some evidence




$$P(\text{Road type} \mid \text{Price=high}, \text{Distance=large}, \text{Journey duration=short})$$

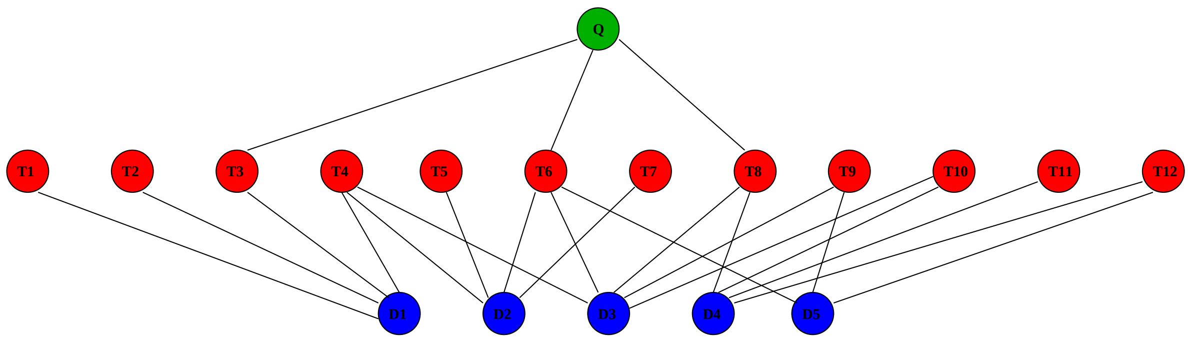
# IR models based on Bayesian nets




- Nodes: term, document and query
  - Documents $\mathcal{D} = \{D_1, \ldots, D_N\}$. Terms $\mathcal{T} = \{T_1, \ldots, T_M\}$
  - Variables $d_j \in \{d_j^-, d_j^+\}$ and $t_i \in \{t_i^-, t_i^+\}$ ({not relevant, relevant})

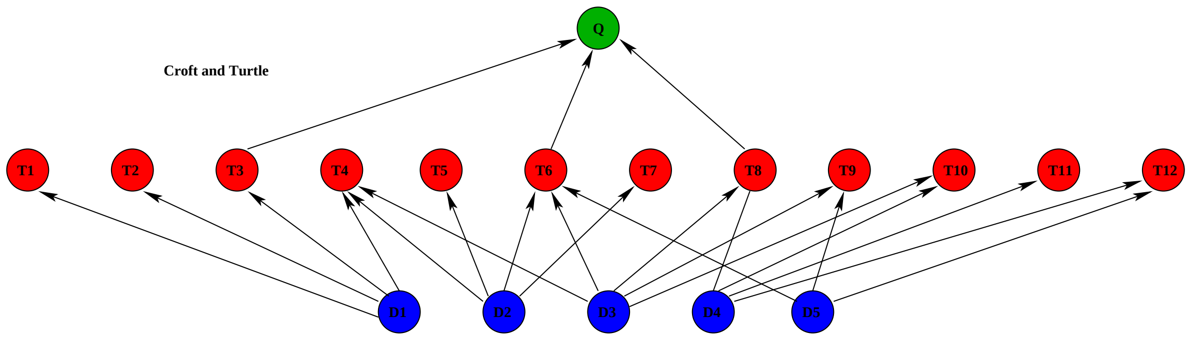


# IR models based on Bayesian nets



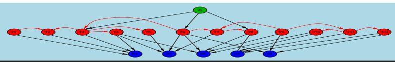

- Nodes: term, document and query
  - Documents $\mathcal{D} = \{D_1, \ldots, D_N\}$. Terms $\mathcal{T} = \{T_1, \ldots, T_M\}$
  - Variables $d_j \in \{d_j^-, d_j^+\}$ and $t_i \in \{t_i^-, t_i^+\}$ ({not relevant, relevant})
- Basic topology:
  - The terms are connected to documents where they appear.
  - The query is connected to the terms that constitute it.

# IR models based on Bayesian nets



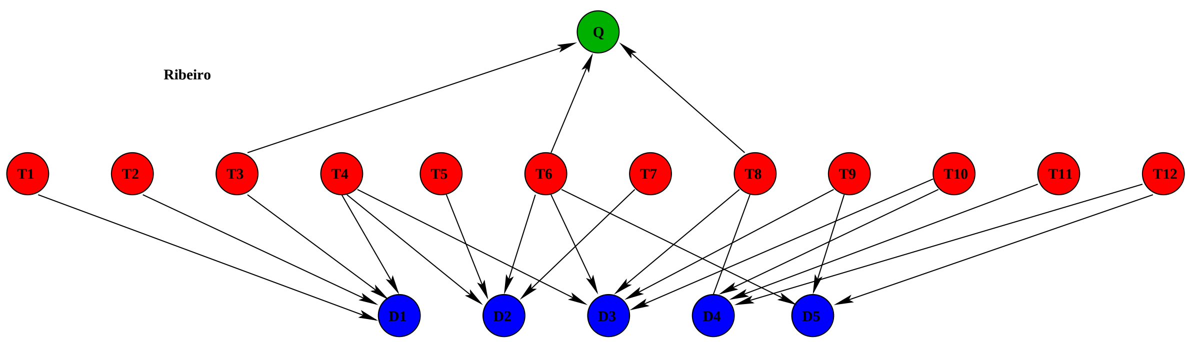

- Nodes: term, document and query
  - Documents $\mathcal{D} = \{D_1, \ldots, D_N\}$. Terms $\mathcal{T} = \{T_1, \ldots, T_M\}$
  - Variables $d_j \in \{d_j^-, d_j^+\}$ and $t_i \in \{t_i^-, t_i^+\}$ ({not relevant, relevant})
- Basic topology:
  - The terms are connected to documents where they appear.
  - The query is connected to the terms that constitute it.

Different models depending on the orientation of the arcs and the existence of additional arcs.



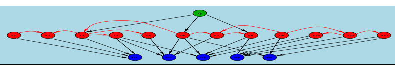


# IR models based on Bayesian nets



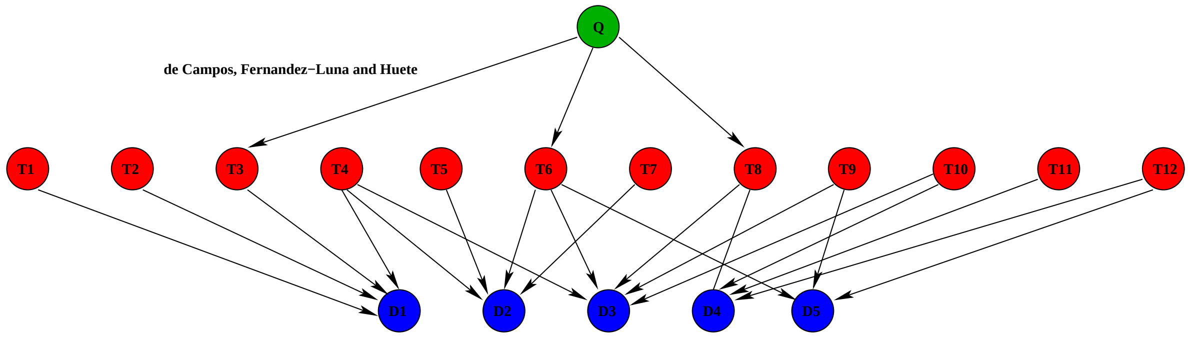

- Nodes: term, document and query
  - Documents $\mathcal{D} = \{D_1, \ldots, D_N\}$. Terms $\mathcal{T} = \{T_1, \ldots, T_M\}$
  - Variables $d_j \in \{d_j^-, d_j^+\}$ and $t_i \in \{t_i^-, t_i^+\}$ ({not relevant, relevant})
- Basic topology:
  - The terms are connected to documents where they appear.
  - The query is connected to the terms that constitute it.

Different models depending on the orientation of the arcs and the existence of additional arcs.



Information Retrieval and Influence Diagrams. The Future of Web Search Workshop – p.10/44

# IR models based on Bayesian nets




- Nodes: term, document and query
  - Documents $\mathcal{D} = \{D_1, \ldots, D_N\}$. Terms $\mathcal{T} = \{T_1, \ldots, T_M\}$
  - Variables $d_j \in \{d_j^-, d_j^+\}$ and $t_i \in \{t_i^-, t_i^+\}$ ({not relevant, relevant})
- Basic topology:
  - The terms are connected to documents where they appear.
  - The query is connected to the terms that constitute it.

Different models depending on the orientation of the arcs and the existence of additional arcs.



# Structured IR models based on BNs

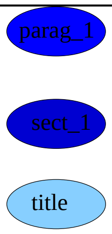
Natural extension of the models for flat documents to hierarchical ones:

- Nodes:


Terms $\mathcal{T}$: Used to express the Information needs:

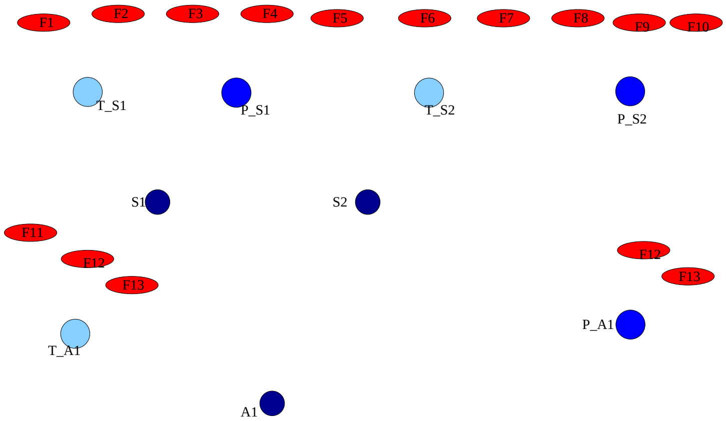
$$\mathcal{T} = \{T_1, T_2, \ldots, T_l\}$$

**states**: Non Relevant or Relevant, $T_k \in \{t_k^-, t_k^+\}$




Structural Units $\mathcal{U}$: components that may be shown to the user, *tags* (piece of data) in the XML format.
**states**: Not Relevant or Relevant, $U_i \in \{u_i^-, u_i^+\}$
*paragraph, section, article, title, ...*

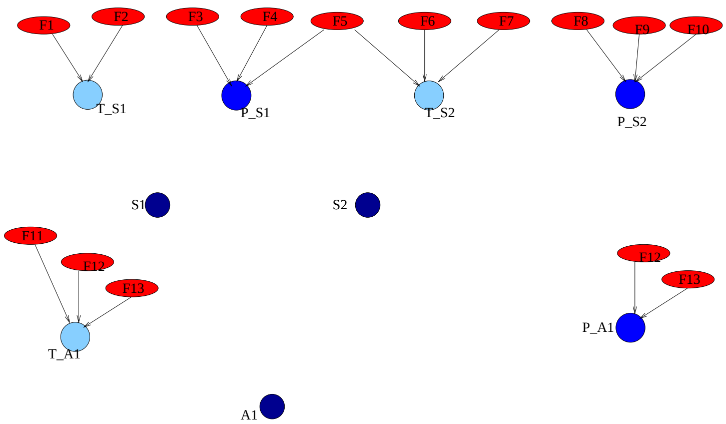
# Structured IR models based on BNs




- Arcs: The outcome of event A affects the probability of event B



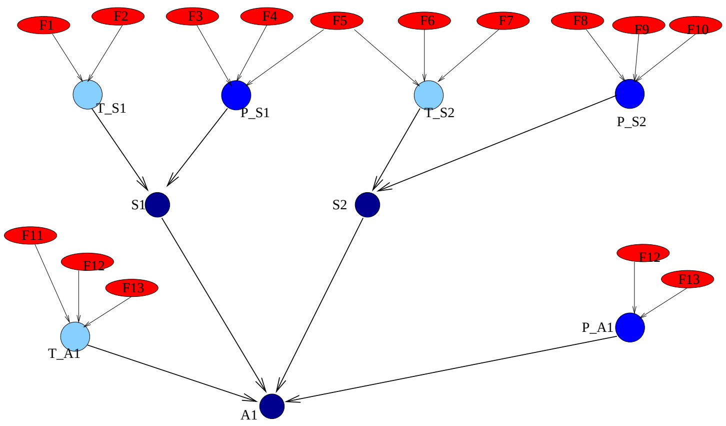
# Structured IR models based on BNs




- Arcs: The outcome of event A affects the probability of event B
  - From terms to the basic structural units where they appear.
    *paragraph, title, ...*

Information Retrieval and Influence Diagrams. The Future of Web Search Workshop – p.12/44

# Structured IR models based on BNs




- Arcs: The outcome of event A affects the probability of event B
  - From terms to the basic structural units where they appear. *paragraph, title, ...*
  - From each structural unit to the single complex unit that contains it. *article, section, subsection, ...*



# Why probabilities are not enough?

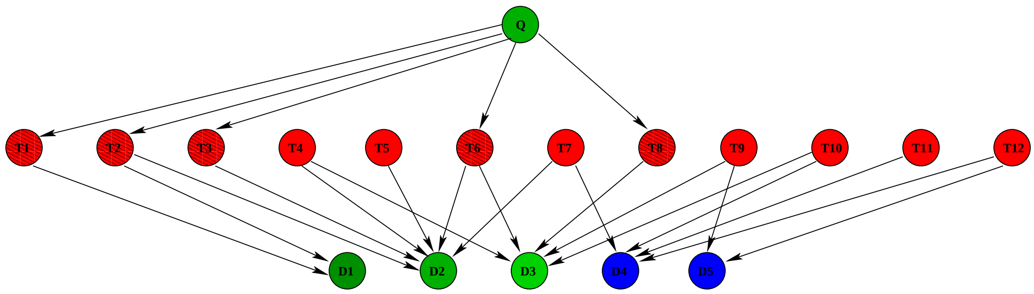
Characteristics of probabilistic (BN-based) retrieval:

- Probability of relevance ≈ degree of matching with the query: the more terms in the document appears in the query, the more relevant the document.
- The probabilities of relevance of larger documents tend to be smaller.
- The retrieval of a document is independent on the retrieval of any other document in the collection.

# Why probabilities are not enough?

Characteristics of probabilistic (BN-based) retrieval:

- Probability of relevance ≈ degree of matching with the query: the more terms in the document appears in the query, the more relevant the document.
- The probabilities of relevance of larger documents tend to be smaller.
- The retrieval of a document is independent on the retrieval of any other document in the collection.




Flat retrieval: The retrievable elements are complete documents, which are all “similar”. The previous characteristics are OK.



# Why probabilities are not enough? (2)

Structured retrieval: The retrievable elements are structural units, which are different.

- Structural units more specific and reduced (e.g. paragraphs) tend to have probabilities of relevance greater than those of more generic and wider units (e.g. chapters).

# Why probabilities are not enough? (2)

Structured retrieval: The retrievable elements are structural units, which are different.

- Structural units more specific and reduced (e.g. paragraphs) tend to have probabilities of relevance greater than those of more generic and wider units (e.g. chapters).
- The retrieval on a structural unit may affect the retrieval of other units (redundancy).



# Why probabilities are not enough? (2)

Structured retrieval: The retrievable elements are structural units, which are different.

- Structural units more specific and reduced (e.g. paragraphs) tend to have probabilities of relevance greater than those of more generic and wider units (e.g. chapters).
- The retrieval on a structural unit may affect the retrieval of other units (redundancy).

Section 1.1 relevant
Section 1.2 relevant
Section 1.3 relevant

---

Retrieve chapter 1
Do not retrieve sections

# Why probabilities are not enough? (2)

Structured retrieval: The retrievable elements are structural units, which are different.

- Structural units more specific and reduced (e.g. paragraphs) tend to have probabilities of relevance greater than those of more generic and wider units (e.g. chapters).
- The retrieval on a structural unit may affect the retrieval of other units (redundancy).

| | |
|---|---|
| Section 1.1 relevant | Section 1.1 relevant |
| Section 1.2 relevant | Section 1.2 irrelevant |
| Section 1.3 relevant | Section 1.3 irrelevant |
| Retrieve chapter 1 | Do not retrieve chapter 1 |
| Do not retrieve sections | Retrieve section 1.1 |



# Why probabilities are not enough? (2)

Structured retrieval: The retrievable elements are structural units, which are different.

- Structural units more specific and reduced (e.g. paragraphs) tend to have probabilities of relevance greater than those of more generic and wider units (e.g. chapters).
- The retrieval on a structural unit may affect the retrieval of other units (redundancy).

| | |
|---|---|
| Section 1.1 relevant | Section 1.1 relevant |
| Section 1.2 relevant | Section 1.2 irrelevant |
| Section 1.3 relevant | Section 1.3 irrelevant |
| Retrieve chapter 1 | Do not retrieve chapter 1 |
| Do not retrieve sections | Retrieve section 1.1 |

- Some types of structural units may be more interesting than others for different users, depending on their own preferences and goals.

# Why probabilities are not enough? (3)

The most probable relevant units are not necessarily the best. In addition to be probably relevant they should also be useful.



# Why probabilities are not enough? (3)

The most probable relevant units are not necessarily the best. In addition to be probably relevant they should also be useful.

Proposed solution: To add a decision module, in order to determine which structural units are more useful, as a function of

- the relevance probabilities,
- the utility of these units for the user,
- and the context where these units appear.

# Influence diagrams

Decision making: the choice of one among a number of possible alternatives.
Decision analysis: Framework for analyzing decision problems by structuring and breaking them down into more manageable parts, explicitly considering:

- problem structure, available information
- possible alternatives, relevant preferences
- uncertainties involved



# Influence diagrams

Decision making: the choice of one among a number of possible alternatives.
Decision analysis: Framework for analyzing decision problems by structuring and breaking them down into more manageable parts, explicitly considering:

- problem structure, available information
- possible alternatives, relevant preferences
- uncertainties involved

| State of the world $\theta$ | Value of alternatives | |
|---|---|---|
| Disease A | treat | do not treat |
| $\theta_1$=yes | $v_{yt}$ | $v_{yn}$ |
| $\theta_2$=no | $v_{nt}$ | $v_{nn}$ |

# Influence diagrams

Decision making: the choice of one among a number of possible alternatives.
Decision analysis: Framework for analyzing decision problems by structuring and breaking them down into more manageable parts, explicitly considering:

- problem structure, available information
- possible alternatives, relevant preferences
- uncertainties involved

| State of the world $\theta$ | Probability | Value of alternatives | |
|---|---|---|---|
| Disease A | $p(A)$ | treat | do not treat |
| $\theta_1$=yes | $p(A=yes)$ | $v_{yt}$ | $v_{yn}$ |
| $\theta_2$=no | $p(A=no)$ | $v_{nt}$ | $v_{nn}$ |

Decision making under risk: We know (or estimate) the probabilities of states. Decision criterion: Select the alternative which maximizes the Expected Value: $EV(a_i) = \sum_{j=1}^{n} p(\theta_j) y_{ji}$



# Influence diagrams (2)

Without using utilities (more precisely, using boolean utilities): If A=yes then treat; if A=no then do not treat.

| State | Probability | Value of alternatives | |
|---|---|---|---|
| Disease A | $p(A)$ | treat | do not treat |
| yes | $p(A=yes)$ | $v_{yt}=1$ | $v_{yn}=0$ |
| no | $p(A=no)$ | $v_{nt}=0$ | $v_{nn}=1$ |

Decision: Treat iff $p(A=yes) > p(A=no) \iff p(A=yes) > \frac{1}{2}$

# Influence diagrams (2)

Without using utilities (more precisely, using boolean utilities): If A=yes then treat; if A=no then do not treat.

| State | Probability | Value of alternatives | |
|---|---|---|---|
| Disease A | $p(A)$ | treat | do not treat |
| yes | $p(A = yes)$ | $v_{yt} = 1$ | $v_{yn} = 0$ |
| no | $p(A = no)$ | $v_{nt} = 0$ | $v_{nn} = 1$ |

Decision: Treat iff $p(A = yes) > p(A = no) \iff p(A = yes) > \frac{1}{2}$

To treat a person do not suffering A may have a very different value that not to treat a person suffering A (e.g. $v_{nt} = 0$, $v_{yn} = -1$; in this case the decision of maximum expected utility is Treat iff $p(A = yes) > \frac{1}{3}$).



# Influence diagrams (3)

Influence Diagrams: generalization of Bayesian networks to deal with decision problems.

# Influence diagrams (3)

Influence Diagrams: generalization of Bayesian networks to deal with decision problems.

- Nodes in influence diagrams:
    - 

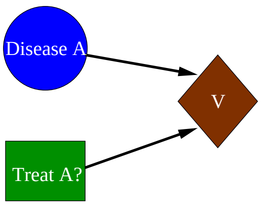
      Chance nodes: represent random variables (states of nature).
    - 

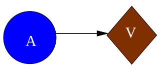
      Decision nodes: represent alternatives.
    - 

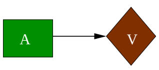
      Utility nodes: represent (numerically) consequences of the decisions.



# Influence diagrams: Examples (4)




- Arcs in influence diagrams:
    - 

      The outcome of event A affects the value of utility.
    - 

      Decision A affects the value of utility.
- Quantitative Information:

| Disease A | Value of alternatives | |
|---|---|---|
| $p(A)$ | treat | do not treat |
| $p(A = yes) = 0.3$ | 10 | 0 |
| $p(A = no) = 0.7$ | 5 | 10 |

# Influence diagrams: Examples (5)

We may have to manage several decisions:

Examples:

- A medical test could improve our knowledge about the possibility of suffering disease A
  - First decision: Perform test?
  - Second decision: Treatment for A?

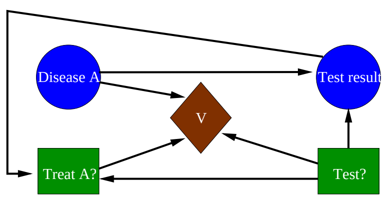


# Influence diagrams: Examples (5)

We may have to manage several decisions:

Examples:

- A medical test could improve our knowledge about the possibility of suffering disease A
  - First decision: Perform test?
  - Second decision: Treatment for A?



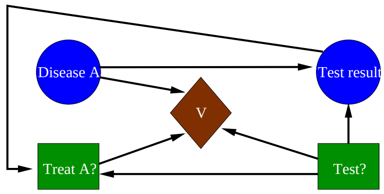

Arcs in influence diagrams:

# Influence diagrams: Examples (5)

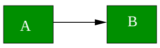
We may have to manage several decisions:

Examples:

- A medical test could improve our knowledge about the possibility of suffering disease A
  - First decision: Perform test?
  - Second decision: Treatment for A?




## Arcs in influence diagrams:



Decision A occurs before decision B. Decisions A and B are sequential.

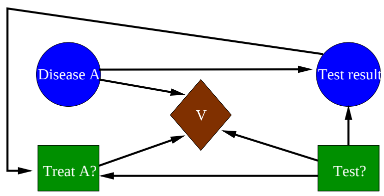


# Influence diagrams: Examples (5)

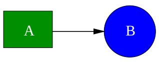
We may have to manage several decisions:

Examples:

- A medical test could improve our knowledge about the possibility of suffering disease A
  - First decision: Perform test?
  - Second decision: Treatment for A?




## Arcs in influence diagrams:



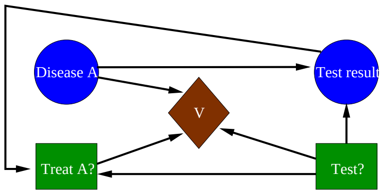
Decision A affects the probabilities of event B. Decision A is relevant for event B.

# Influence diagrams: Examples (5)

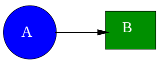
We may have to manage several decisions:

Examples:

- A medical test could improve our knowledge about the possibility of suffering disease A
  - First decision: Perform test?
  - Second decision: Treatment for A?




## Arcs in influence diagrams:




Decision B occurs after event A. The outcome of A is known when deciding about B.



# Influence diagrams: Examples (6)

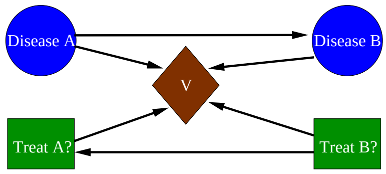
Or the patient may also suffer another more serious disease B, which is commonly caused by A

- First decision: Treatment for B?
- Second decision: Treatment for A?

# Influence diagrams: Examples (6)

Or the patient may also suffer another more serious disease B, which is commonly caused by A

- First decision: Treatment for B?
- Second decision: Treatment for A?

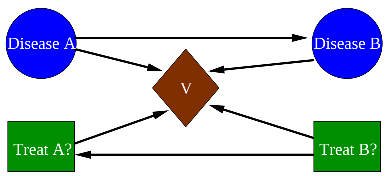





# Influence diagrams: Examples (6)

Or the patient may also suffer another more serious disease B, which is commonly caused by A

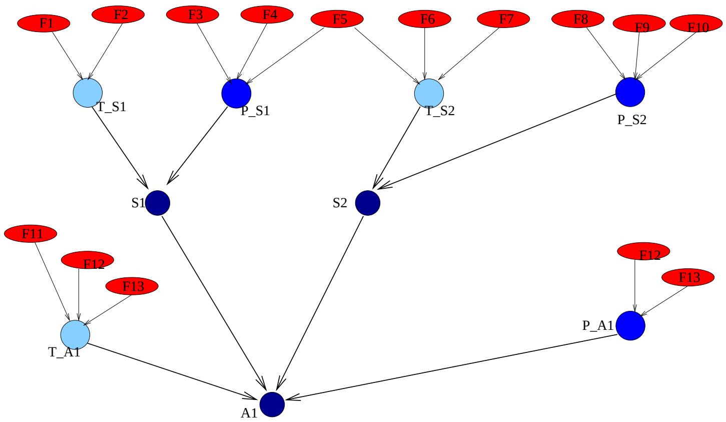
- First decision: Treatment for B?
- Second decision: Treatment for A?




Solving an ID is finding the set of decisions that maximize the expected utility, for each one of the scenarios of interest.

# Building ID for Structured IR

Problem: Modeling the original BN towards a decision model




We have to include Decision and Utility nodes



# Building ID for Structured IR (2)

- Decision Nodes, $\mathcal{R}$:
  Representing decision alternatives for the system.

# Building ID for Structured IR (2)

- Decision Nodes, $\mathcal{R}$:
  Representing decision alternatives for the system.
  - Retrieve a Section
  - Retrieve the Article
  - Retrieve the Title of a Section



# Building ID for Structured IR (2)

- Decision Nodes, $\mathcal{R}$:
  Representing decision alternatives for the system.
  - Retrieve a Section
  - Retrieve the Article
  - Retrieve the Title of a Section


For each retrievable structural unit the system has to decide between retrieval or not retrieval

$$\forall U_i \Rightarrow R_i \text{ with } R_i \in \{r_i^+, r_i^-\}$$

# Building ID for Structured IR (2)

- Decision Nodes, $\mathcal{R}$:
  Representing decision alternatives for the system.
  - Retrieve a Section
  - Retrieve the Article
  - Retrieve the Title of a Section


For each retrievable structural unit the system has to decide between retrieval or not retrieval

$$\forall U_i \Rightarrow R_i \text{ with } R_i \in \{r_i^+, r_i^-\}$$

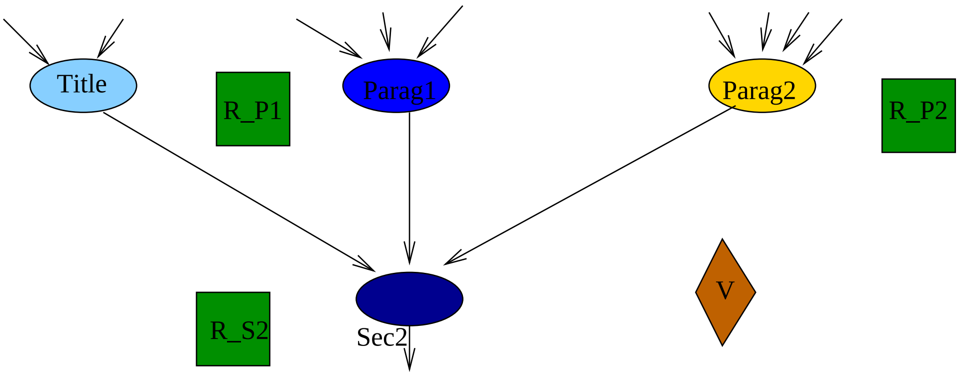
- Utility or Value Nodes, $\mathcal{V}$: Used to measure the utility (profit) of the decisions


Goal: **utility** of retrieval a unit **depends on the context**



# Building ID for Structured IR (3)

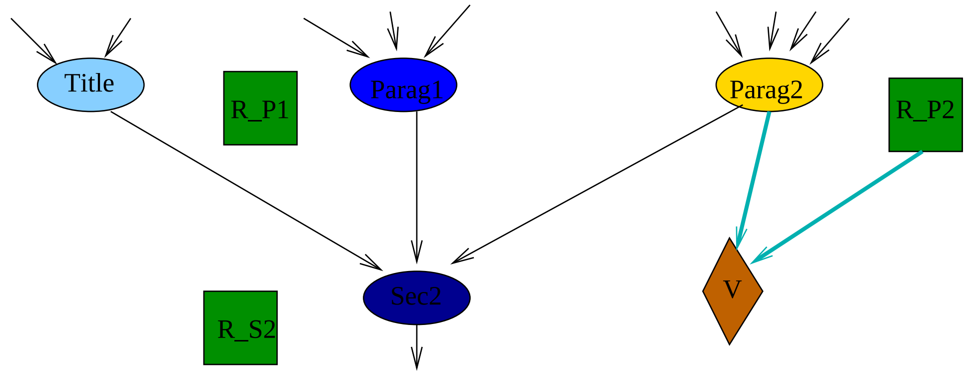
Relations with utility nodes:




The utility of retrieval *paragraph 2* will depend on ...

# Building ID for Structured IR (4)

Relations with utility nodes:



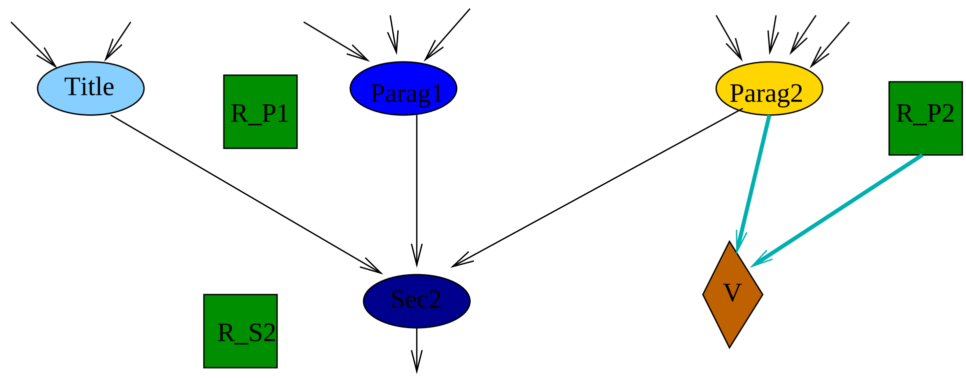

So, we can encode decision rules as

- *If the paragraph is relevant AND we retrieve it we gain 100$ .*
- *If the paragraph is relevant AND we do not retrieve it we gain 0$.*



# Building ID for Structured IR (4)

Relations with utility nodes:



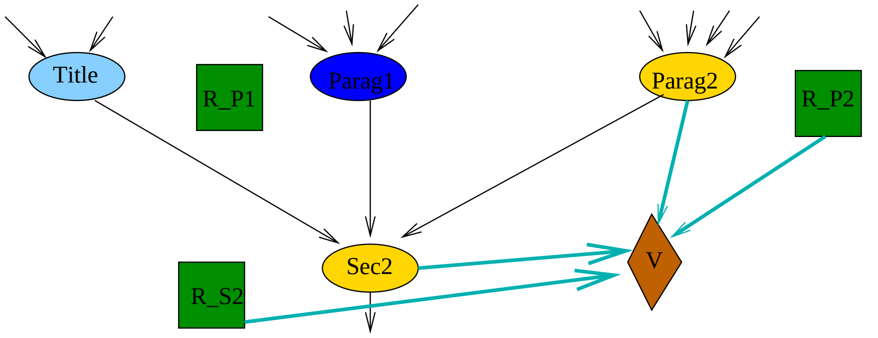

So, we can encode decision rules as

- *If the paragraph is relevant AND we retrieve it we gain 100$ .*
- *If the paragraph is relevant AND we do not retrieve it we gain 0$.*

But, what about the **context** in which the decision has been made ?

# Building ID for Structured IR (5)

Relations with utility nodes:




Assuming that node **Sec2** summarizes the context, we could encode decision rules as

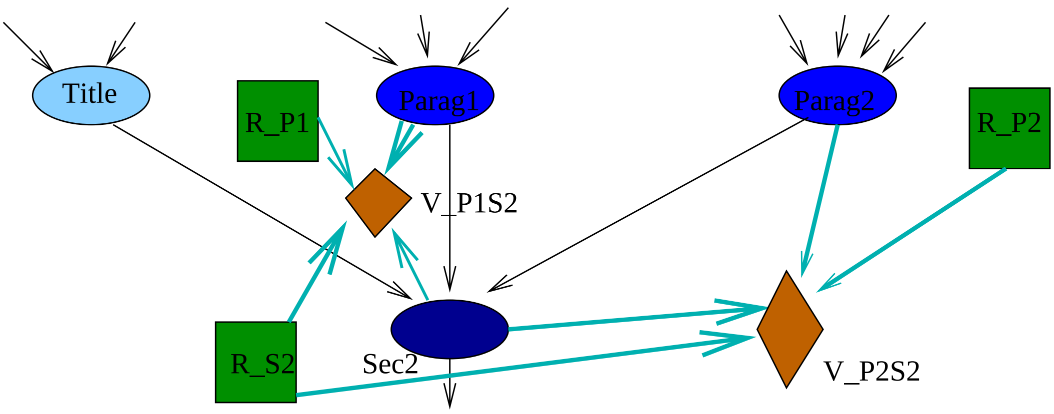
- *If the paragraph and the section are relevant AND we retrieve the section and not the paragraph we gain 100$* .
- *If the paragraph and the section are relevant AND we do not retrieve the section but retrieve the paragraph we gain 50$* .



# Building ID for Structured IR (5)

Relations with utility nodes:

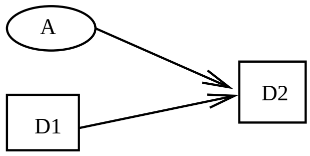
Finally,




Note that each utility node depends on a unit $U_i$ and the (unique) unit that contains it $U_j = U_{C(U_i)}$. So they will be denoted by $V_{i,j}$ .

# Analyzing Decision nodes

Relations with decision nodes:
Indicate that the value of the source node is available when the decision is made.



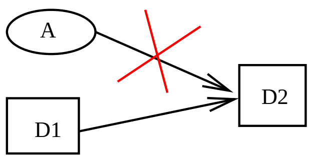

Considering that:

- we can not know a priori the relevance value of a unit



# Analyzing Decision nodes

Relations with decision nodes:
Indicate that the value of the source node is available when the decision is made.



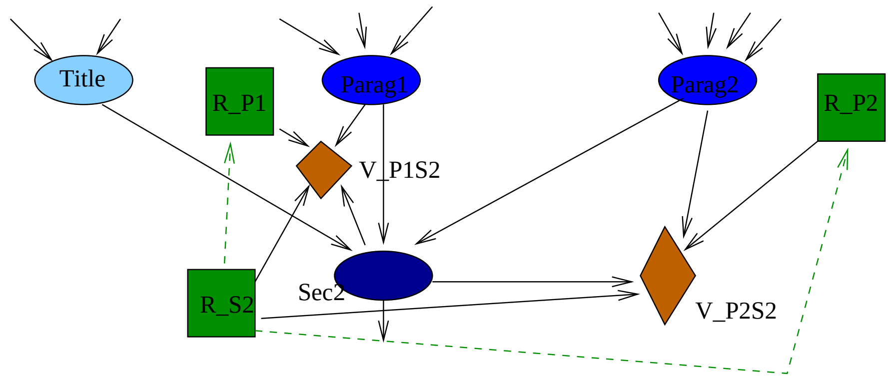

Considering that:

- we can not know a priori the relevance value of a unit
- the redundancy *it will convenient not to retrieve a unit $U_i$ if we have previously retrieved the unit $U_k$ that contains it.*

we can define a partial ordering between decisions nodes.
*General (article)* $<$ *Specific (section)*

# Analyzing Decision nodes (2)

Relations with decision nodes:

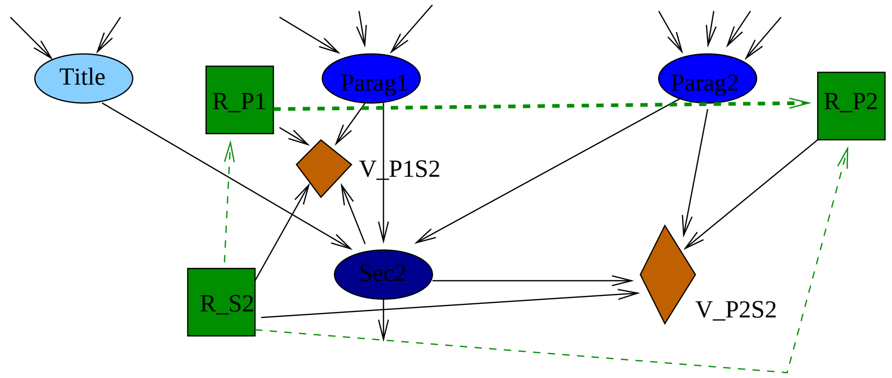



To complete the ordering between decision nodes:
we consider *left* $<$ *right* in the same level of the hierarchy (siblings units),



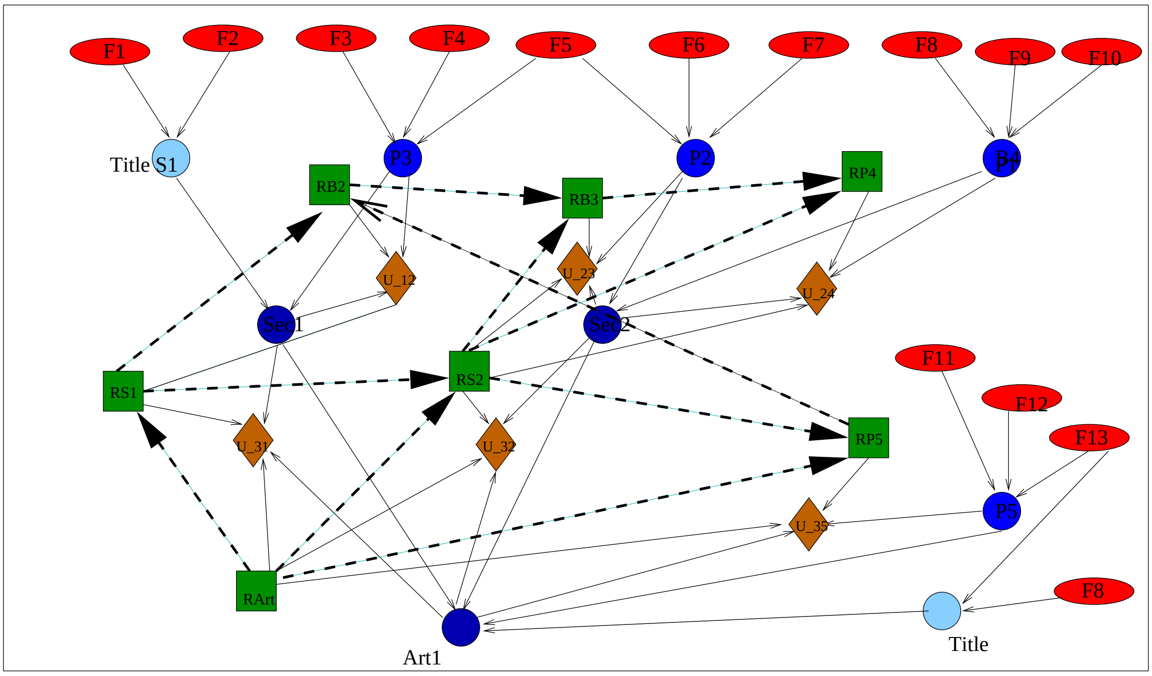
# Analyzing Decision nodes (2)

Relations with decision nodes:




and putting all together we get ...

# ID for Structured IR






# Specifying Probabilities and Utilities

- Probabilities (Chance node $U_i$):

  exponential number of prob. values, $2^{|Pa(U_i)|}$

  Solution: Canonical models

$$\forall U_i \, p(u_i^+ | pa(U_i)) = \sum_{F \in R(pa(U_i))} w(F, U_i)$$

$$p(art^+ | p1^+, s1^+, s2^-) = w(p1, art) + p(s1, art)$$

$$p(art^+ | p1^-, s1^+, s2^+) = w(s1, art) + p(s2, art)$$

- Utilities (e.g. $A \longleftarrow S1$ ): $V_{a,s}(A, S1, R_a, R_{s1})$ needs 16 utility values

| | | | |
|---|---|---|---|
| $v(++++) = 0$ | $v(+++-) = 5$ | $v(++-+) = 0$ | $v(++--) = -5$ |
| $v(+-++) = 0$ | $v(+-+-) = 10$ | $v(+--+) = -15$ | $v(+---) = -15$ |
| $v(-+++) = -15$ | $v(-++-) = -15$ | $v(-+-+) = 15$ | $v(-+--) = 0$ |
| $v(--++) = -15$ | $v(--+-) = -15$ | $v(---+) = -15$ | $v(----) = 15$ |

# Solving ID: Inference

- Input: Query, $Q \subseteq \mathcal{T}$:
- Output: Subset of $\mathcal{U}$
- How?: Solving the ID
  - For each document, select the strategy (set of decisions) with the highest expected utility.

| | Art | P_A1 | Sec1 | Sec2 | P_S1.1 | P_S2.1 | P_S2.2 | V |
|---|---|---|---|---|---|---|---|---|
| Str1 | - | - | + | - | - | + | - | 200$ |
| Str2 | - | - | + | - | - | - | + | 150$ |
| Str3 | + | - | - | - | - | - | - | 50$ |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| $Str_m$ | + | + | - | + | - | - | - | -250$ |



# Solving ID: Inference

| **Doc1** | Art1 | P_A1 | Sec1 | Sec2 | P_S1.1 | P_S2.1 | P_S2.2 | V |
|---|---|---|---|---|---|---|---|---|
| **Str** | - | - | [+] | - | - | [+] | - | 200$ |
| **Doc2** | Art2 | Sec1 | Sec2 | Sec3 | P_S1.1 | P_S2.2 | P_S2.3 | |
| **Str** | - | - | - | [+] | - | - | - | 180$ |
| **Doc3** | Art3 | Sec1 | PS1.1 | PS1.2 | PS1.3 | | | |
| **Str** | - | - | - | - | - | | | 250$ |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |

System output:
It is necessary to merge all the optimal strategies (one for each document), into a unique output
Solution: To rank the decisions in decreasing order of the utility of retrieval each particular unit $U_i$: $EU(r_i^+|q)$.

```
Art1\Sec2\Parag1 < Art2\Sec3 < Art1\Sec1 < ...
```

# Solving ID: Inference

Problem: High computational cost due to

- large number of variables (terms and structural units),
- large number of decision nodes in a document
  - the exponential number of strategies $O(2^{|U_D|})$ to explore
- the topology of the underlying BN with
  - multiple pathways connecting nodes,
  - the great number of parents of each unit.

Solution: Two step approach

Probability inference + Decision Making



# Reducing the set of valid strategies

Coherent criteria that help to prune the search

- If the query is not related with particular unit we shall decide “not to retrieve” the unit and none of the units included in it.

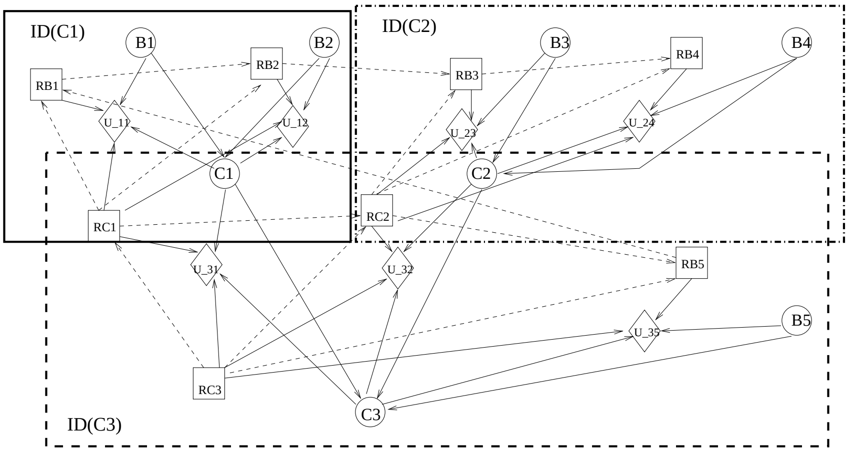
  We do not have to explore the strategies in non relevant documents
- considering specificity, if we decide “to retrieve” a unit, none of the units included in it will be also recommended.

  We reduce the strategies in relevant documents

Nevertheless the number of compatible strategies is exponential.

Final Strategy: **Bottom-up** approach making decisions with the information that can be computed “locally”.

# Splitting into local Influence Diagrams

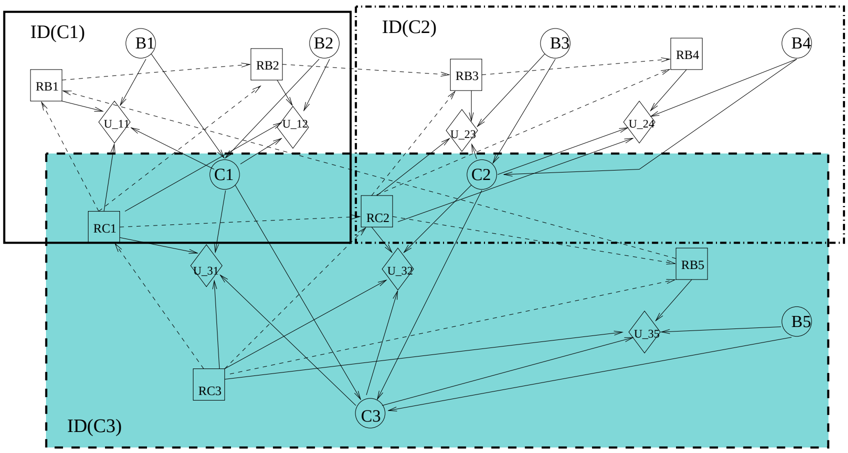



Each local ID relates a structural unit with their inmediate sub-components (*parents*).

$ID_{U_i}$ relates $U_i$ with $Pa(U_i)$



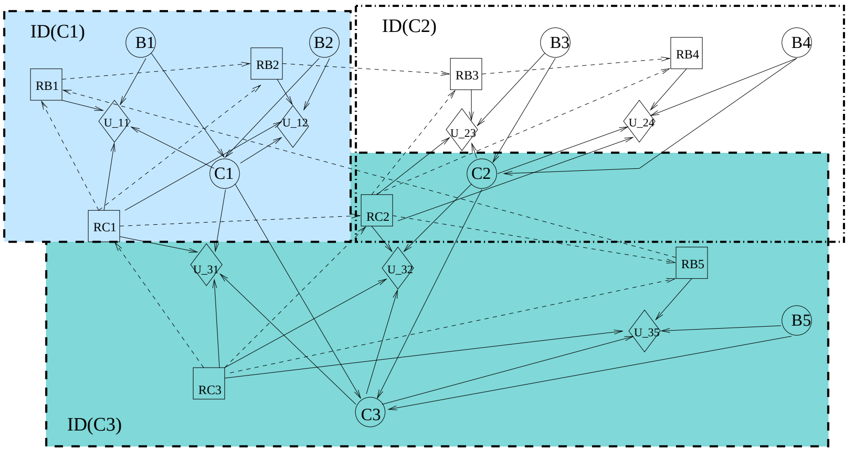
# Splitting into local Influence Diagrams




ID(Article): Candidate Strategies:

| | | | |
|---|---|---|---|
| Art | Parag1 | Sec1 | Sec2 |
| Art | Parag1 | Sec1 | Sec2 |
| Art | Parag1 | Sec1 | Sec2 |

# Splitting into local Influence Diagrams




ID(Section1): Candidate Strategies:

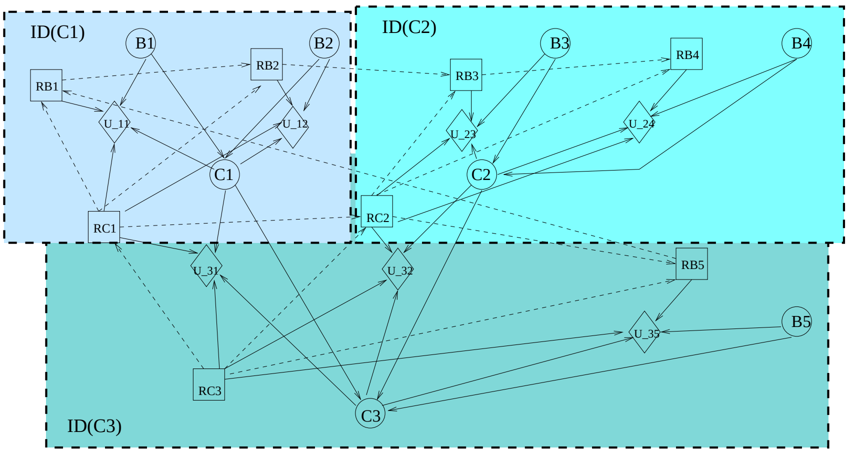
Sec1 Parag1 Parag2

Sec1 Parag1 Parag2

Sec1 Parag1 Parag2



# Splitting into local Influence Diagrams



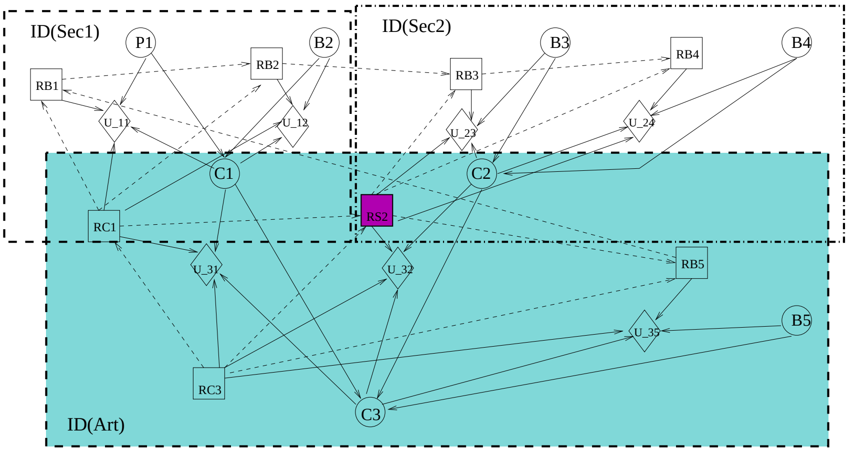

ID(Section2): Candidate Strategies:

Sec2 Parag1 Parag2

Sec2 Parag1 Parag2

Sec2 Parag1 Parag2

# Splitting into local Influence Diagrams



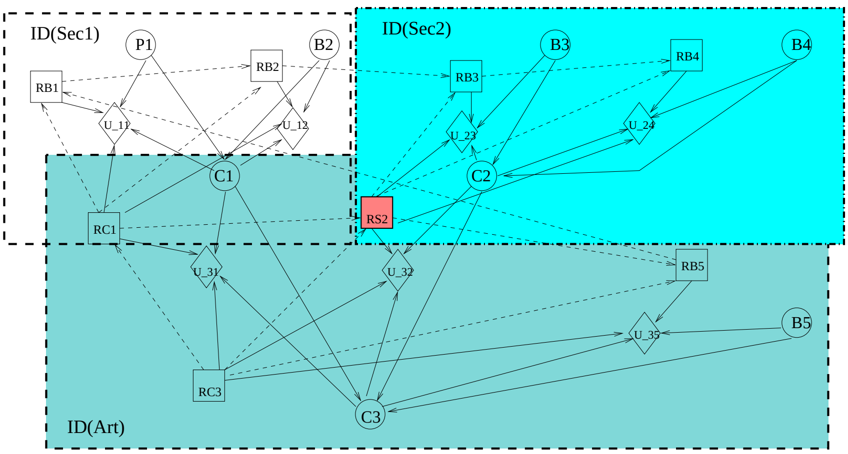

TWO local decisions for each complex structural unit.

1.- related with the unique structural unit containing $U_i$

*Opt. Strategy* **ID(Art)***:* Art ; Parag1 ; Sec1 ; Sec2



# Splitting into local Influence Diagrams




TWO local decisions for each complex structural unit.

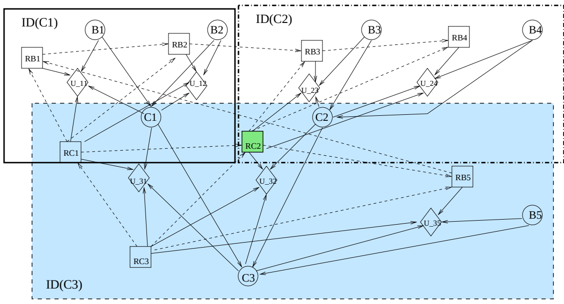
2.- related with the units contained by $U_i$

*Opt. Strategy* **ID(Sec2)***:* Sec2 ; Parag1 ; Parag2

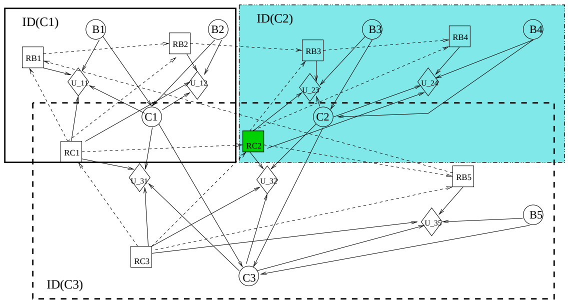
# Relating Decisions $ID_{U_i}$ and $ID_{C(U_i)}$

| | $ID_{U_i} = r_i^+$ | $ID_{U_i} = r_i^-$ |
|---|---|---|
| $ID_{C(U_i)} = r_i^+$ | ? | |
| $ID_{C(U_i)} = r_i^-$ | | |

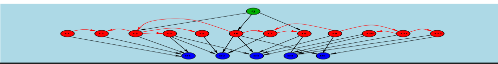
There is no doubt, retrieve the unit $U_i$.




Art; Parag1 ;Sec1; Sec2




Sec2;Parag 1; Parag2



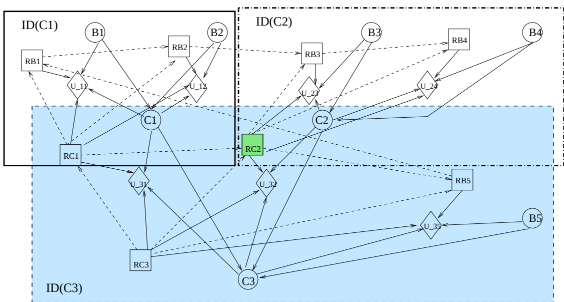


# Relating Decisions $ID_{U_i}$ and $ID_{C(U_i)}$

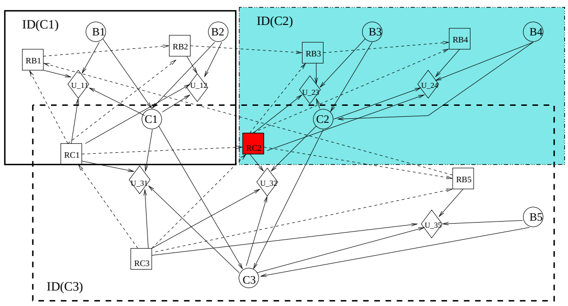
| | $ID_{U_i} = r_i^+$ | $ID_{U_i} = r_i^-$ |
|---|---|---|
| $ID_{C(U_i)} = r_i^+$ | $r_i^+$ | ? |
| $ID_{C(U_i)} = r_i^-$ | | |

$r_i^+$: $U_i$ is more relevant than its siblings.

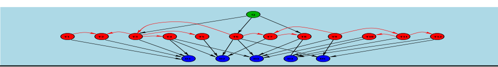
$r_i^-$: is better to retrieve some of its parents.




Art; Parag1 ;Sec1; Sec2




Sec2;Parag 1; Parag2

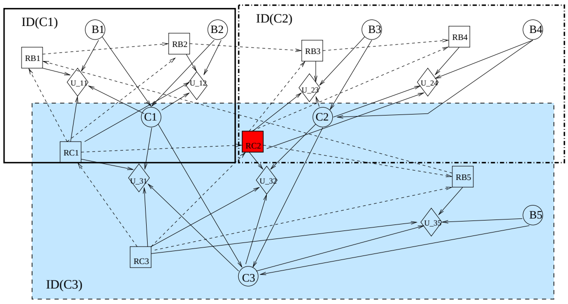


# Relating Decisions $ID_{U_i}$ and $ID_{C(U_i)}$

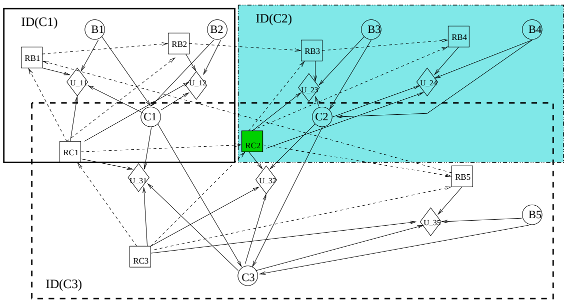
| | $ID_{U_i} = r_i^+$ | $ID_{U_i} = r_i^-$ |
|---|---|---|
| $ID_{C(U_i)} = r_i^+$ | $r_i^+$ | $r_i^-$ |
| $ID_{C(U_i)} = r_i^-$ | ? | |

$r_i^-$: $U_i$ is less relevant than its siblings.

$r_i^+$: $U_i$ is more relevant than its parents.




Art; Parag1 ;Sec1; Sec2



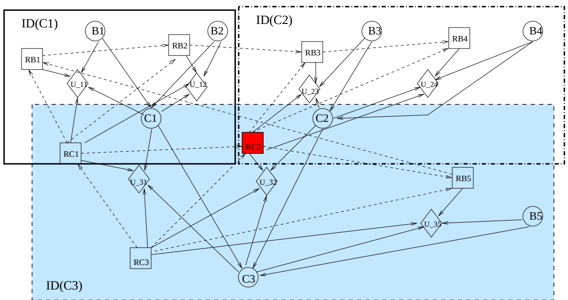

Sec2;Parag 1; Parag2



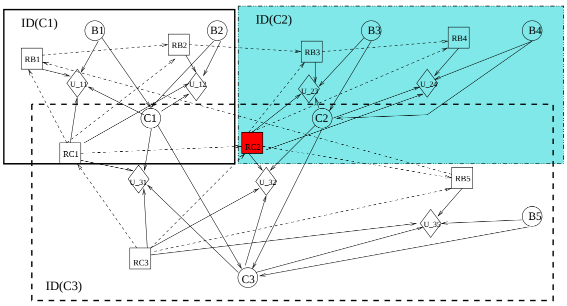
# Relating Decisions $ID_{U_i}$ and $ID_{C(U_i)}$

| | $ID_{U_i} = r_i^+$ | $ID_{U_i} = r_i^-$ |
|---|---|---|
| $ID_{C(U_i)} = r_i^+$ | $r_i^+$ | $r_i^-$ |
| $ID_{C(U_i)} = r_i^-$ | $r_i^+$ | ? |

There is no doubt: Not to retrieve $U_i$.




Art; Parag1 ;Sec1; Sec2




Sec2;Parag 1; Parag2

# Relating Decisions $ID_{U_i}$ and $ID_{C(U_i)}$

| | $ID_{U_i} = r_i^+$ | $ID_{U_i} = r_i^-$ |
|---|---|---|
| $ID_{C(U_i)} = r_i^+$ | $r_i^+$ | $r_i^-$ |
| $ID_{C(U_i)} = r_i^-$ | $r_i^+$ | $r_i^-$ |

So, decision about $U_i$ will only depend on the strategy of maximum expected utility in $ID_{U_i}$.

Ej. The decision about an article will be made considering the decisions about the units directly included in it (its paragraphs and sections).



# How the decisions have been made

Let `RetrievalList` be a list retrieval units

```
MakeDecision(U_j, RetrievalList);
{
  if (U_j is relevant to Q )
      {
       Compute Decision for U_j
       if (Decision  U_j ==Retrieval)
           RetrievaList.insert(U_j,EU(U_j) )
       else for each U_k parent of U_j
          MakeDecision(U_k, RetrievalList);
       }
}
```

Where a unit $U_j$ is relevant to a query $Q$ iff

$$P(U_j|Q) > P(U_j)$$

# Computing Decisions for $U_i$

The decision of retrieval or not $U_i$ is made considering if $EU(r_i^+) > EU(r_i^-)$ in the $ID(U_i)$ .

The computation of $EU(r_i^+)$ can be done with linear cost (in size and time) with the number of parents of $U_i$.

Thus $EU(r_i^+) =$

$$\sum_{U_j \in Pa(U_i)} \max \left\{ \begin{array}{l} \sum_{\substack{u_j \in \{u_j^-, u_j^+\}, \\ u_i \in \{u_i^-, u_i^+\}}} V_{i,j}(u_i, u_j, r_i^+, r_j^+) p(u_j|q) p(u_i|q), \\ \sum_{\substack{u_j \in \{u_j^-, u_j^+\}, \\ u_i \in \{u_i^-, u_i^+\}}} V_{i,j}(u_i, u_j, r_i^+, r_j^-) p(u_j|q) p(u_i|q) \end{array} \right\}$$

being

$$p(u_i^+|Q) = \sum_{F \in Pa(U_i)} w(F, U_i) p(F|Q)$$



# Summarizing

Three step process:

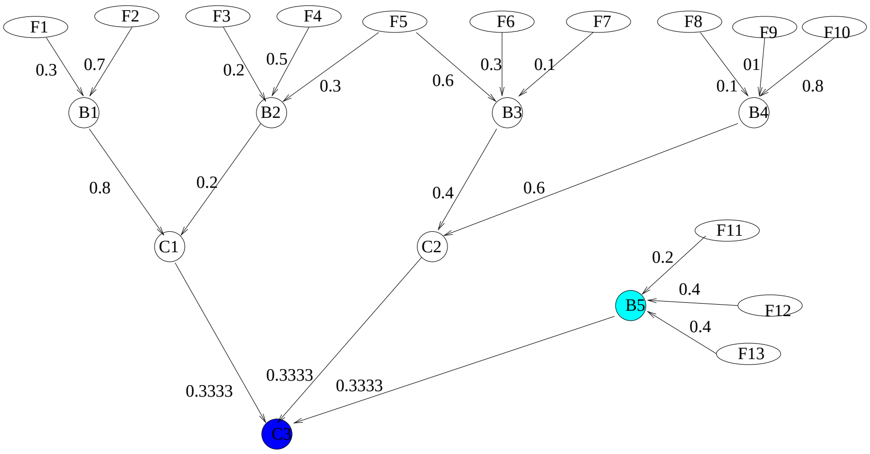
- Probabilistic Inference
  Posterior probabilities of chance nodes in the BN, $P(U_i|Q)$.
  Computed in a Top-Down manner
- Decision process
  Solve local Influence Diagrams in order to obtain the strategy.

  ```
  for each Document, Di, relevant to Q
          MakeDecision( Di, RetrievalList);
  ```

  Computed in a bottom-up manner
- Sort RetrievalList in decreasing order of EU

# Example

Prior Prob.: $p(f^{+}) = 0.5$

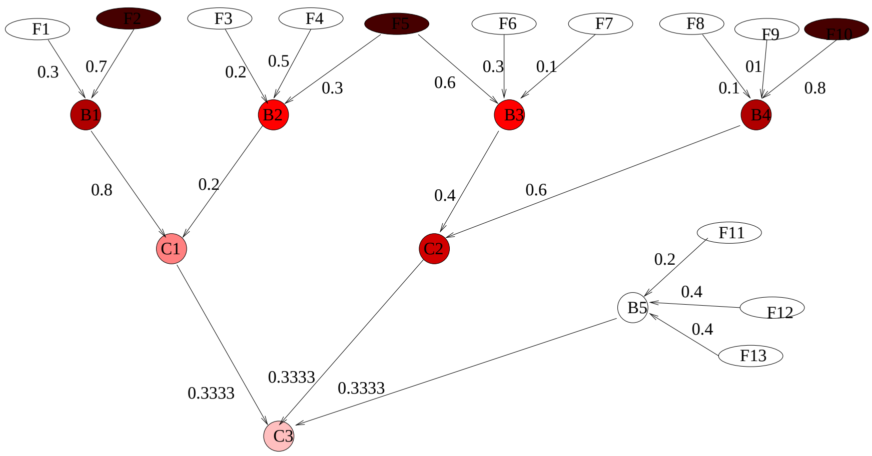



$V_{i,j} = \{U_c, U_p, R_c, R_p\}$

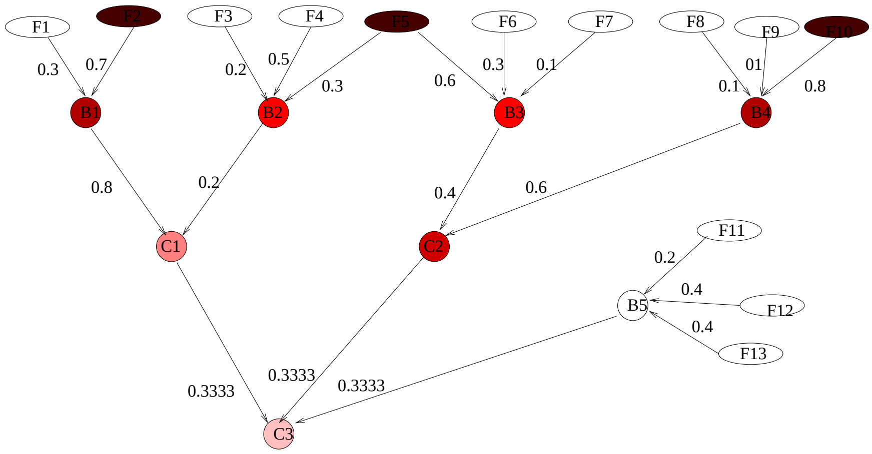
| | | | |
|---|---|---|---|
| $v(++++) = 0$ | $v(+++-) = 5$ | $v(++-+) = 0$ | $v(++--) = -5$ |
| $v(+-++) = 0$ | $v(+-+-) = 0$ | $v(+--+) = -15$ | $v(+--) = -15$ |
| $v(-+++) = -15$ | $v(-++-) = -15$ | $v(-+-+) = 15$ | $v(-+--) = 0$ |
| $v(--++) = -15$ | $v(--+-) = -15$ | $v(---+) = -15$ | $v(----) = 15$ |



# Example (2): BNs vs IDs



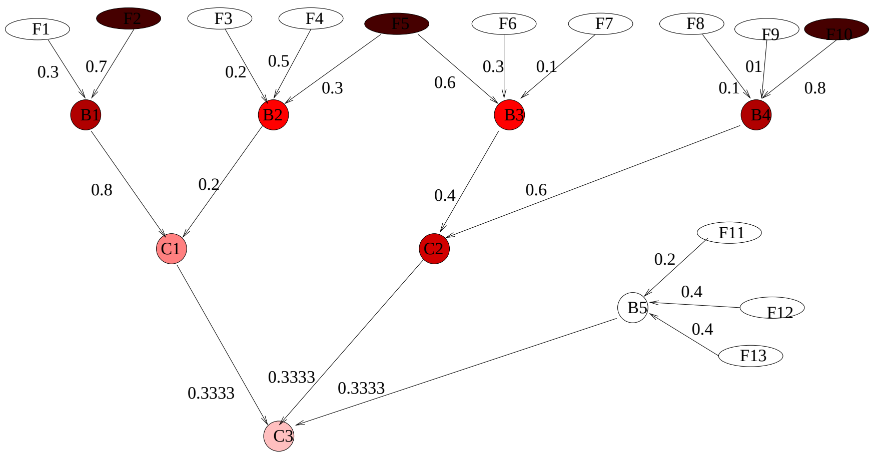
# Example (2): BNs vs IDs




| BNs | B4 | C2 | B1 | B3 | C1 | C3 | B2 | B5 |
|---|---|---|---|---|---|---|---|---|
| $Q_1$ | 0.90 | 0.86 | 0.85 | 0.80 | 0.750 | 0.703 | 0.65 | 0.50 |



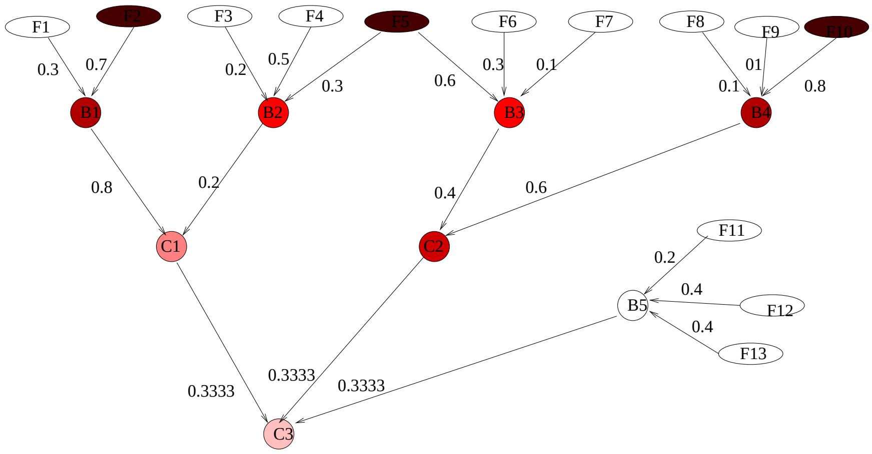
# Example (2): BNs vs IDs




| BNs | B4 | C2 | B1 | B3 | C1 | C3 | B2 | B5 |
|---|---|---|---|---|---|---|---|---|
| $Q_1$ | 0.90 | 0.86 | 0.85 | 0.80 | 0.750 | 0.703 | 0.65 | 0.50 |

| Global ID | Optimal Strategy |
|---|---|
| $Q_1$ | $r^{+}_{c2}(1.12)$ $r^{+}_{c1}(-1.35)$ |

# Example (2): BNs vs IDs




| BNs | B4 | C2 | B1 | B3 | C1 | C3 | B2 | B5 |
|---|---|---|---|---|---|---|---|---|
| $Q_1$ | 0.90 | 0.86 | 0.85 | 0.80 | 0.750 | 0.703 | 0.65 | 0.50 |

| Global ID | Optimal Strategy | |
|---|---|---|
| $Q_1$ | $r^+_{c2}(1.12)$ | $r^+_{c1}(-1.35)$ |

| Local IDs | $ID_{C1}$ | $ID_{C2}$ | $ID_{C3}$ | System Output | |
|---|---|---|---|---|---|
| $Q_1$ | $r^-_{c1}, r^-_{b1}, r^-_{b2}$ | $r^+_{c2}, r^-_{b3}, r^-_{b4}$ | $r^-_{c3}, r^+_{c1}, r^+_{c2}, r^-_{b5}$ | $r^+_{c2}(3.11)$ | $r^+_{c1}(-1.87)$ |



# Example (3)

Queries:

$Q_1 = \{f_2^+, f_5^+, f_{10}^+\}$ $Q_2 = \{f_2^+, f_6^+, f_{10}^+\}$ $Q_3 = \{f_2^+, f_5^-, f_{10}^+\}$

| $Q$ | Optimal Strategy | | C3 | C1 | C2 | B1 | B2 | B3 | B4 | B5 |
|---|---|---|---|---|---|---|---|---|---|---|
| $Q_1$ | $r^+_{c2}(1.12)$ | $r^+_{c1}(-1.35)$ | 0.703 | 0.750 | 0.86 | 0.85 | 0.65 | 0.80 | 0.90 | 0.50 |
| $Q_2$ | $r^+_{b1}(0.94)$ | $r^+_{c2}(0.25)$ | 0.658 | 0.675 | 0.80 | 0.85 | 0.50 | 0.65 | 0.90 | 0.50 |
| $Q_3$ | $r^+_{b4}(2.82)$ | $r^+_{b1}(1.89)$ | 0.593 | 0.600 | 0.62 | 0.85 | 0.35 | 0.20 | 0.90 | 0.50 |

| $Q$ | $ID_{C1}$ | $ID_{C2}$ | $ID_{C3}$ | System Output | |
|---|---|---|---|---|---|
| $Q_1$ | $r^-_{c1}, r^-_{b1}, r^-_{b2}$ | $r^+_{c2}, r^-_{b3}, r^-_{b4}$ | $r^-_{c3}, r^+_{c1}, r^+_{c2}, r^-_{b5}$ | $r^+_{c2}(3.11)$ | $r^+_{c1}(-1.87)$ |
| $Q_2$ | $r^-_{c1}, r^+_{b1}, r^-_{b2}$ | $r^+_{c2}, r^-_{b3}, r^-_{b4}$ | $r^-_{c3}, r^+_{c1}, r^+_{c2}, r^-_{b5}$ | $r^+_{b1}(1.89)$ | $r^+_{c2}(0.2)$ |
| $Q_3$ | $r^-_{c1}, r^+_{b1}, r^-_{b2}$ | $r^-_{c2}, r^-_{b3}, r^+_{b4}$ | $r^-_{c3}, r^+_{c1}, r^+_{c2}, r^-_{b5}$ | $r^+_{b4}(3.63)$ | $r^+_{b1}(2.85)$ |

# Conclusions

- Decision-based model for structured information retrieval
- Automatically detect the “best entry points”:
  
  Considering the context
  - probabilities of relevance
  - utility for the user (user preferences)

Future Research

- Estimation of the weights in chance nodes
- Elicitation of the utility values
  - Explicitly in the query
  - By means of an user profile
- Adapt the model to use link information in Web framework
- Adapt the model for Recommeding in hierarchical domais