Formal query systems on contexts and a representation of algebraic lattices

Qingguo Li a, Lankun Guo b,*

a College of Mathematics and Econometrics, Hunan University, Changsha, Hunan 410082, PR China
b College of Information Science and Engineering, Hunan University, Changsha, Hunan 410082, PR China

Abstract

Formal Concept Analysis (FCA) has proven to be a feasible framework for information retrieval (IR). In this paper, with the aim of exploring the mathematical properties of the FCA-based document retrieval paradigm, we propose the notions of conjunctive and disjunctive query on the formal contexts and investigate the derived query systems from the order-theoretical perspective. We show that the conjunctive query system is isomorphic to the Galois lattice of the underlying context, while the disjunctive query system can be join-densely generated from the Galois lattice of the underlying context up to isomorphism. Next, we introduce directed disjunctive query systems and study their connection with algebraic lattices. As a result, we show that the notion of directed disjunctive query provides an alternative approach to restructuring algebraic lattices.

1. Introduction

Formal Concept Analysis (FCA) has proven to be a useful method for data analysis and knowledge discovery [47,48]. In FCA, the data is introduced as a formal context, i.e., a set of objects provided with a set of attributes and a binary incidence relation between them. A context induces a Galois connection between the powersets of objects and attributes, respectively. Its fixpoints represent pairs of maximal sets from opposite sides of the connection which are called formal concepts. When ordered by set-theoretic inclusion, the concepts compose a complete lattice which is called a concept lattice or Galois lattice. In practice, FCA has been successfully employed in a wide range of applications such as knowledge discovery [3,14,25,27,31], ontology engineering [15,20,22] and software engineering [12,42].

FCA has its root in order theory especially in lattice theory [4,17] and also plays an important role in characterizing some important order structures. It is well known that the classical FCA provides an approach to restructuring complete lattices [46]. In [21], Hartung investigated the representation of bounded lattices by introducing topology into the classical FCA framework. Zhang and Shen [49,50] studied the connection between Chu spaces, information systems and Domain theory, and proposed approximable concepts as a generalization of formal concepts. They showed that approximable concepts precisely correspond to the information states of Scott’s information systems. Based on this work, Hitzler et al. [23] investigated the notion of approximable concept from the categorical viewpoint and obtained the equivalence between a particular category of formal contexts and that of algebraic lattices. Afterwards, Lei and Luo [30] introduced the notion of rough approximable concept and provided a new approach to representing algebraic lattices. These results reveal the potential of the combination of FCA and other mathematical structures in characterizing more general order structures.
In FCA-based document retrieval applications, the documents usually serve as the formal objects while the index terms serve as the formal attributes. Such a document-term relationship can be transformed into a Galois lattice in which every concept is interpreted as an entity consisting of a collection of documents concerning its extent and that of index terms regarding its intent. In this way, FCA provides a lattice-based methodology for discovering the hierarchical classification of documents from a document-term relationship [41,43]. The most attractive feature of this method is its ability of integrating browsing and direct term querying within a unique supporting structure. It also facilitates the employing of other useful features such as automatic indexing and using of thesaurus in a flexible framework [35]. In recent years, several extensions of the classical FCA have been proposed in order to improve the capacity of FCA in the real-world applications [11,44,52]. Much research has also been conducted on combining the classical FCA and other data analysis tools such as rough sets [26,45] and fuzzy sets [13,36,37].

A basic function of IR systems is to present relevant resources that match the user’s statement of the information being sought. However, the performance of the conventional keyword-based retrieval is usually limited because the user-initiated query may not perfectly match some description of the documents. In order to improve the retrieval performance, a query reformulation process is often employed. In this aspect, much effort has been made to enhance the FCA-based approaches for IR. For instance, Godin et al. [19] showed that browsing and query formulation can be combined into the FCA-based retrieval paradigm for a unique retrieval space including both terms and documents. Carpineto and Romano [7] developed a system named REFINER based on Galois lattices to conduct the refinement and enlargement of the user-initiated queries. By incorporating background knowledge and a thesaurus in the query formulation process, Carpineto and Romano [6] and Priss [34] independently developed enhanced FCA-based systems for the usual document-term relationship. In [51], by combining FCA and a notion of information anchors, Zhang et al. provided an innovative approach to information retrieval for domain-specific document collections.

From the above background, we notice that due to the incorporation of various mechanisms into the FCA-based retrieval paradigm, a user-initiated query usually involves a class of elemental queries; and such classes may be arranged into a hierarchy according to their corresponding retrieved documents. In this paper, we tackle the problem of query classification in the enhanced FCA-based document retrieval paradigm from a mathematical perspective. Particularly, we propose the notions of formal conjunctive and disjunctive query which formalize two approaches of generating queries from elemental ones by Boolean operators AND and OR. Then we concentrate on the order-theoretic properties of the derived query systems.

The rest of this paper is organized as follows. In Section 2, we briefly recall some basic notions about algebraic lattices and FCA. In Section 3, we introduce the notions of formal conjunctive and disjunctive query on a formal context and investigate the order-theoretic properties of the derived query systems. In Section 4, we propose the notion of formal directed disjunctive query and study the connection between the corresponding query systems and algebraic lattices. In Section 5, we reach the conclusion and talk about future research.

2. Preliminary

2.1. Complete lattices and algebraic lattices

We recall some basic notions about algebraic lattices. Most of them are collected from [10].

Let \((L, \leq)\) be a partially ordered set (for short, poset). The element \(a \in L\) is called an upper bound (respectively, a lower bound) of \(X \subseteq L\) if \(x \leq a\) (respectively, \(a \leq x\)) for any \(x \in X\). The least upper bound (respectively, the greatest lower bound) of \(x, y \in L\), if exists, is denoted by \(x \lor y\) (respectively, \(x \land y\)) which is also called the join (respectively, the meet) of \(x\) and \(y\). If the least upper bound (respectively, the greatest lower bound) of \(X \subseteq L\) exists, it is denoted by \(\bigvee X\) (respectively, \(\bigwedge X\)). A subset \(X \subseteq L\) is said to be join-dense (respectively, meet-dense) in \(L\) if for any \(y \in L\) there exists a subset \(X' \subseteq X\) such that \(y \geq \bigvee X'\) (respectively, \(y \leq \bigwedge X'\)).

Let \((L, \leq)\) be a poset. Given a subset \(X \subseteq L\), we write \(\downarrow X = \{y \in L | (\exists x \in X) y \leq x\}\). If \(X = \downarrow x\), then \(x\) is called a down-set of \(L\). For convenience, we use \(\downarrow x\) to represent \(\downarrow \{x\}\) for a singleton \(\{x\} \subseteq L\). A subset \(D \subseteq L\) is said to be directed (respectively, filtered) if it is nonempty and its every finite subset has an upper bound (respectively, a lower bound) in \(D\). A poset \(L\) is said to be complete with respect to directed subsets (for short, directed complete) if every directed subset of \(L\) has the least upper bound. Given \(x, y \in L\), we say \(x\) way below \(y\), written as \(x \ll y\), if for any directed subset \(D \subseteq L\) for which \(\bigcup D\) exists, \(y \leq \bigcup D\) always implies the existence of some element \(d \in D\) such that \(x \leq d\). An element \(x \in L\) is said to be compact if it satisfies \(x \ll x\). The set of all compact elements of \(L\) is denoted by \(K(L)\). A poset \(L\) is said to be algebraic if it satisfies the Axiom of Compact Approximation, i.e., for any \(x \in L\), \(x \cap K(L)\) is a directed subset of \(L\) and \(x = \bigcup (\{x \cap K(L)\})\). A directed complete algebraic poset is called an algebraic domain. An algebraic domain which is also a complete lattice is called an algebraic lattice. More results about Domain theory can be found in [18].

2.2. Formal contexts and formal concepts

A (formal) context is a triplet \((G, M, I)\) where \(G\) and \(M\) are sets and \(I \subseteq G \times M\) a binary relation. Elements of \(G\) and \(M\) are usually called objects and attributes, respectively. The relation \(I\) describes the incidence relation between objects and attributes.
Given a context \((G,M,I)\), the operators \(\cdot^\cup\) on the powersets of the object and attribute sets are defined separately as follows: for any subsets \(A \subseteq G\) and \(B \subseteq M\),
\[
A^\cup = \{ m \in M \mid (\forall g \in A)(g, m) \in I \},
\]
\[
B^\cup = \{ g \in G \mid (\forall m \in B)(g, m) \in I \}.
\]

We recall some basic properties of the above operators which will be used in the sequel sections.

**Proposition 2.1** \([17]\). Let \((G,M,I)\) be a context. For any \(A, A_1, A_2 \subseteq G, B, B_1, B_2 \subseteq M, \{A_j\}_{j \in J} \subseteq \mathcal{P}(G)\) and \(\{B_t\}_{t \in T} \subseteq \mathcal{P}(M)\),
\[
\begin{align*}
1) & \quad A_1 \subseteq A_2 \Rightarrow A_1^\cup \supseteq A_2^\cup; \quad B_1 \subseteq B_2 \Rightarrow B_1^\cup \supseteq B_2^\cup; \\
2) & \quad A \subseteq A^\cup; \quad B \subseteq B^\cup; \\
3) & \quad A^\cup = A^\cup, B = B^\cup; \\
4) & \quad \left( \bigcup_{j \in J} A_j^\cup \right) = \bigcup_{j \in J} A_j^\cup; \quad \left( \bigcup_{t \in T} B_t^\cup \right) = \bigcup_{t \in T} B_t^\cup.
\end{align*}
\]

For singletons \(\{g\} \subseteq G\) and \(\{m\} \subseteq M\), we abbreviate \(g^\uparrow\) as \(g^\downarrow\) and \(m^\uparrow\) as \(m^\downarrow\). As is well known, the above operators form a Galois connection between \((\mathcal{P}(G), \subseteq)\) and \((\mathcal{P}(M), \subseteq)\). For subsets \(A \subseteq G\) and \(B \subseteq M\), the pair \((A,B)\) is called a (formal) concept of \((G,M,I)\) if \(A^\cup = B\) and \(B^\cup = A\). In this case, \(A\) and \(B\) are called the (formal) extent and (formal) intent of \((A,B)\), respectively. Moreover, it is easy to check that a subset \(A \subseteq G\) is an extent of a concept if and only if \(A = A^\cup\); dually, a subset \(B \subseteq M\) is an intent of a concept if and only if \(B = B^\cup\). In the following, we use \(\mathcal{B}(G,M,I)\) to denote the set of all concepts of \((G,M,I)\).

The partial order \(\leq\) on \(\mathcal{B}(G,M,I)\) is defined as follows: for any concepts \((A_1,B_1)\) and \((A_2,B_2)\),
\[
(A_1, B_1) \leq (A_2, B_2) \iff A_1 \subseteq A_2 \quad \text{and} \quad B_1 \subseteq B_2.
\]

The equivalence between \((A_1, B_1)\) and \((A_2, B_2)\) formally captures the logical reciprocity: the larger the extension, the smaller the intension. In the following, we use \(\mathcal{B}(G)\) (respectively, \(\mathcal{B}(M)\)) to denote the set of all extents (respectively, intents) of \((G,M,I)\). It is easy to see that \((\mathcal{B}(G,M,I), \leq)\) is isomorphic to \((\mathcal{B}(G), \subseteq)\) and \((\mathcal{B}(M), \subseteq)\).

From the viewpoint of knowledge representation, every context \((G,M,I)\) can be understood as a contextual knowledge space in the following sense: (i) \(G\) represents a collection of objects in a specified domain; (ii) \(M\) represents the properties that the objects in \(G\) may possess; (iii) \(I\) specifies which objects have what properties. If we take a collection of documents as objects and their index terms as attributes, then every document-term relationship can be formally represented by a formal context. Moreover, a hierarchical classification of documents can be abstracted from a document-term relationship. The basic idea of FCA-based document retrieval is that every concept can be thought of as a query described by a collection of index terms concerning the intent with a collection of retrieved documents regarding the extent. More generally, if \(A\) represents a collection of documents, then \(A^\uparrow\) represents the common terms indexing every document in \(A\); and dually, if \(B\) represents a collection of index terms, then \(B^\downarrow\) represents the documents which possess every term in \(B\).

Some properties of formal concepts just reflect some basic characteristics of IR systems. If we understand any subset \(B \subseteq M\) as an elemental conjunctive query composed of the terms in \(B\), then the subset \(B^\downarrow\) can be interpreted as the collection of retrieved documents and \(B^\uparrow\) as the minimal query enlargement of \(B\). In this sense, Item (1) of Proposition 2.1 just reflects the fact that if more index terms are used, which means that the query is more specified, then fewer documents are retrieved. Items (2) and (3) of Proposition 2.1 demonstrate the function of FCA in query reformulation in the sense that every elemental conjunctive query can be augmented to a new one with the same collection of retrieved documents.

The following basic theorem shows that for any context \((G,M,I)\), \((\mathcal{B}(G,M,I), \leq)\) is a complete lattice (called the concept lattice or Galois lattice of \((G,M,I)\)); and conversely, every complete lattice can arise as the Galois lattice of an appropriate context.

**Theorem 2.1** \([17]\). Let \((G,M,I)\) be a context. Then \((\mathcal{B}(G,M,I), \leq)\) is a complete lattice in which the meet and join can be formulated as follows: for any family of concepts \(\{(A_j,B_j)\}_{j \in J}\),
\[
\bigcap_{j \in J} (A_j,B_j) = \left( \bigcap_{j \in J} A_j \bigcap \bigcup_{j \in J} B_j \right)^\downarrow,
\]

\[
\bigcup_{j \in J} (A_j,B_j) = \left( \bigcup_{j \in J} A_j \bigcup \bigcap_{j \in J} B_j \right)^\uparrow.
\]

A complete lattice \(L\) is isomorphic to \(\mathcal{B}(G,M,I)\) if and only if there exist mappings \(\gamma: G \rightarrow L\) and \(\mu: M \rightarrow L\) such that \(\gamma(G)\) is join-dense in \(L\), \(\mu(M)\) is meet-dense in \(L\) and \((g,m) \in I\) is equivalent to \(\gamma(g) \leq \mu(m)\) for all \(g \in G\) and \(m \in M\).

**Example 2.1.** Let Table 1 be a document-term relationship consisting of six documents and eight index terms, where “D” stands for the word “Document”. The crosses in the table demonstrate the incidence relation between documents and index terms. For convenience, the notations “LG”, “W3”, “FC”, “AR”, “GL”, “KD”, “DM” and “SW” are used shortly for the index terms “Logic”, “World Wide Web”, “Formal Concept”, “Approximate Reasoning”, “Galois Lattice”, “Knowledge Discovery”, and other terms. The document-term relationship can be represented by a formal context. The properties of this context can be used to define a lattice. The lattice can be used to represent the document-term relationship. The lattice can be used to represent the context. The context can be used to represent the document-term relationship. The document-term relationship can be represented by a formal context. The context can be used to represent the document-term relationship. The document-term relationship can be represented by a formal context. The context can be used to represent the document-term relationship. The document-term relationship can be represented by a formal context. The context can be used to represent the document-term relationship.

Please cite this article in press as: Q. Li, L. Guo, Formal query systems on contexts and a representation of algebraic lattices, Inform. Sci. (2013), \url{http://dx.doi.org/10.1016/j.ins.2013.03.032}
Data Mining” and “Semantic Web” separately. Fig. 1 illustrates the Galois lattice derived from this context. Each node of the lattice is composed of a collection of documents concerning the extent and that of index terms regrading the intent. For simplicity, we omit the alphabet “D” in indicating the collections of documents. For instance, “123” in the diagram stands for the set \{D1, D2, D3\}.

3. Formal query systems on contexts

In this section, we consider the problem of query classification in FCA-based documents retrieval paradigm. Limited by the scope, we only consider the formalization of query classification in the context of Boolean querying [16]. Now, we introduce a conventional notation similar to the Boolean conjunctive expression used in [11]: Given a nonempty set \(X\), we associate every subset \(B \subseteq X\) with an expression in the form \(\gamma(B) = \bigwedge B \land \bigvee B\). Intuitively, if \(B\) represents a collection of index terms, then \(\gamma(B)\) can be understood as an elemental query which is conjunctively composed of the terms in \(B\).

At a higher level, more queries may be generated from the elemental ones via Boolean operators AND, OR and NOT. This is based on the following perspectives: (i) when a user wants to search for the documents which match each of the used elemental queries, it is natural to construct a new query by combining those elemental ones via AND, and (ii) if the synonymy among terms is considered, it is reasonable to construct a disjunctive query from elemental ones via OR. The situation will be more complex when NOT is taken into account. In this paper, we only consider two basic ways of generating queries from elemental ones. The following notions can be seen as their formalization where the notations \(\land\) and \(\lor\) are used to stand for Boolean operators AND and OR, respectively.

**Definition 3.1.** Let \((G,M,I)\) be a context. For any family \(\{B_j\}_{j \in J} \subseteq \mathcal{P}(M)\), \(\bigwedge_{j \in J} \gamma(B_j)\) and \(\bigvee_{j \in J} \gamma(B_j)\) are called the (formal) conjunctive query and (formal) disjunctive query associated with \(\{B_j\}_{j \in J}\) on \((G,M,I)\), respectively.

**Example 3.1**

1. Consider the document-term relationship given in Example 2.1. Suppose a user conducts two individual queries: one is represented by \(LG \land GL\) and the other by \(GL \land KD\). If the user wants to search for the documents which are exactly indexed by “Logic”, “Galois Lattice” and “Knowledge Discovery”, then this query can be represented by \(LG \land GL \land KD\) which is conjunctively composed of \(LG \land GL\) and \(GL \land KD\).
(2) Suppose a user searches for the documents which are indexed by "World Wide Web" and "Formal Concept", then this query can be represented by W3 ∨ FC. It is easy to see that D6 is the only document which matches this query. However, it is well known that Semantic Web is the current mainstream of the World Wide Web, and Galois lattices are the core structure of formal concept analysis. Therefore, documents indexed by both "Semantic Web" and "Galois Lattice" are also relevant to the user's query even though they do not exactly match the original query. Therefore, it is reasonable to consider the disjunctive query (W3 ∨ FC) ∨ (SM ∨ GL).

In the sequel, we use C(G, M, I) and D(G, M, I) to denote the set of all conjunctive and disjunctive queries on (G, M, I), respectively. Given a query \( \bigwedge_{j \in J} \gamma(B_j) \) (or \( \bigvee_{j \in J} \gamma(B_j) \)) on the context (G, M, I), if every \( B_j \) is an intent of (G, M, I), then it is said to be intential. Intuitively, \( \bigcap_{j \in J} B_j \) (respectively, \( \bigcup_{j \in J} B_j \)) can be interpreted as the collection of documents retrieved by the query \( \bigwedge_{j \in J} \gamma(B_j) \) (respectively, \( \bigvee_{j \in J} \gamma(B_j) \)). For simplicity, given a conjunctive query \( Q = \bigwedge_{j \in J} \gamma(B_j) \), we use the notation \( \eta(Q) \) to represent the set \( \bigcap_{j \in J} B_j \). For a disjunctive query \( Q = \bigvee_{j \in J} \gamma(B_j) \), we use the notation \( \epsilon(Q) \) to denote the set \( \bigcup_{j \in J} B_j \).

Given conjunctive (respectively, disjunctive) queries \( Q_1 \) and \( Q_2 \) on (G, M, I), if \( \eta(Q_1) \subseteq \eta(Q_2) \) (respectively, \( \epsilon(Q_1) \subseteq \epsilon(Q_2) \)), it can be interpreted as: the documents retrieved by \( Q_1 \) are fewer than those by \( Q_2 \). This suggests that the queries can be arranged into a hierarchy according to their retrieved documents. On the other hand, we notice that several lattice-based approaches for refining queries containing Boolean operators have been proposed [33, 38]. A marked disadvantage of such systems is that there exists redundancy in the query formulation because different queries may lead to the same output and the number of possible refinements may grow too large even for a very limited number of terms [40]. This motivates us to consider the problem of query classification in the FCA-based document retrieval paradigm.

**Definition 3.2.** Let (G, M, I) be a context. A relation \( \succ \) on \( C(G, M, I) \) is defined as follows: for any conjunctive queries \( Q_1 \) and \( Q_2 \) on (G, M, I),

\[ Q_1 \succ Q_2 \iff \eta(Q_1) = \eta(Q_2). \]

A relation \( \sim \) on \( D(G, M, I) \) is defined as follows: for any disjunctive queries \( Q_1 \) and \( Q_2 \) on (G, M, I),

\[ Q_1 \sim Q_2 \iff \epsilon(Q_1) = \epsilon(Q_2). \]

It is easy to check that \( \succ \) and \( \sim \) are equivalence relations on \( C(G, M, I) \) and \( D(G, M, I) \), respectively. We use \( C_\succ(G, M, I) \) and \( D_\sim(G, M, I) \) to denote their corresponding sets of equivalent classes.

**Example 3.2**

1. Consider the context given in Example 2.1 and the conjunctive queries (LG ∧ GL) ∨ (DM ∧ SW) and LG ∧ SW. It is easy to check that they have the same collection of retrieved documents [D3]. Therefore, they are equivalent with respect to \( \succ \) in the sense of Definition 3.2.

2. Consider the disjunctive queries (LG ∧ GL) ∨ (GL ∧ SW) and LG ∨ SW. It is easy to calculate that they have the same collection of retrieved documents [D1, D2, D3, D4]. This means that (LG ∧ GL) ∨ (GL ∧ SW) and LG ∨ SW are equivalent with respect to \( \sim \) in the sense of Definition 3.2. In addition, the query LG ∨ SW may be regarded as a refinement of (LG ∧ GL) ∨ (GL ∧ SW) since they have the same retrieved documents.

Next, we study the order-theoretic properties of \( C_\succ(G, M, I) \) and \( D_\sim(G, M, I) \) where the hierarchy on query classes is determined by the set-theoretic inclusion between their corresponding collections of retrieved documents. Formally, we introduce the following partial orders.

**Definition 3.3.** Let (G, M, I) be a context. A partial order \( \preceq \) on \( C_\succ(G, M, I) \) is defined as follows: for any conjunctive queries \( Q_1 \) and \( Q_2 \) on (G, M, I),

\[ [Q_1]_{\succ} \preceq [Q_2]_{\succ} \iff \eta(Q_1) \subseteq \eta(Q_2). \]

A partial order \( \preceq \) on \( D_\sim(G, M, I) \) is defined as follows: for any disjunctive queries \( Q_1 \) and \( Q_2 \) on (G, M, I),

\[ [Q_1]_{\sim} \preceq [Q_2]_{\sim} \iff \epsilon(Q_1) \subseteq \epsilon(Q_2). \]

We call \( (C_\succ(G, M, I), \preceq) \) and \( (D_\sim(G, M, I), \preceq) \) the conjunctive query system and disjunctive query system on (G, M, I), respectively.

**Example 3.3**

1. Consider the context in Example 2.1 and the conjunctive queries LG ∧ GL and (GL ∧ DM) ∧ SW. We can check that their collections of retrieved documents are [D1, D2, D3] and [D3], respectively. Therefore, the query class represented by LG ∧ GL is higher than that by (GL ∧ DM) ∧ SW in the conjunctive query system.
Now we explore some basic properties of the query systems from the viewpoint of query reformulation. For notational convenience, given a set $X$ and $\mathcal{A} \subseteq \mathcal{P}(X)$, we use the notations $\bigcup \mathcal{A}$ and $\bigcap \mathcal{A}$ to denote $\bigcup \{A | A \in \mathcal{A}\}$ and $\bigcap \{A | A \in \mathcal{A}\}$, respectively.

**Proposition 3.1.** Let $(G, M, I)$ be a context. Then for any $B \subseteq M$ and any conjunctive query $\bigwedge_{j \in J} \gamma(B_j)$ on $(G, M, I)$,

1. $[\gamma(B)]_{\infty} = [\gamma(B^0)]_{\infty}$;
2. $[\bigwedge_{j \in J} \gamma(B_j)]_{\infty}$ for any $\mathcal{C} \subseteq \mathcal{P}(M)$ such that $B = \bigcup \mathcal{C}$;
3. $[\bigwedge_{j \in J} \gamma(B_j)]_{\infty}$.

**Proof.**

1. By Definition 3.2, it is sufficient to check $B^0 = B^I$ which immediately follows from Proposition 2.1.
2. Suppose $B \subseteq M$ and $\mathcal{C} \subseteq \mathcal{P}(M)$ such that $B = \bigcup \mathcal{C}$. By Proposition 2.1, we have $\bigcap_{k \in K} C_k = (\bigcup \mathcal{C})'$ which implies that $B^0 = \bigcap_{k \in K} C_k$. This means $\gamma(B)_{\infty} = [\bigwedge_{j \in J} \gamma(B_j)]_{\infty}$.
3. It immediately follows from $\bigwedge_{j \in J} B_j = (\bigcup_{j \in J} B_j)'$ which has been presented in Proposition 2.1.

Intuitively,

1. means that the documents retrieved by an elemental query are exactly those retrieved by an elemental intential one. Since $B^0$ is an intent for any $B \subseteq M$, $B^0$ can be considered as an intential query enlargement of $B$.
2. means that every elemental conjunctive query can be decomposed into smaller ones while the common documents retrieved by them are exactly those retrieved by the original one.
3. indicates that every conjunctive query can be transformed into an elemental one which has the same retrieved documents.

An immediate consequence of Items (1) and (3) of Proposition 3.1 is that every conjunctive query is equivalent to an elemental intential query with respect to $\infty$. Therefore, every equivalent class in $\mathcal{E}(G, M, I)$ just corresponds to an extent (hence, a concept) of $(G, M, I)$. In the following, we specify the meet and join in $\mathcal{E}(G, M, I)$ and show the isomorphism between the conjunctive query system and the Galois lattice of the underlying context.

**Theorem 3.1.** Let $(G, M, I)$ be a context. Then $(\mathcal{E}(G, M, I), \subseteq)$ is a complete lattice. In particular, for any family of conjunctive queries $\{\bigwedge_{j \in J_k} \gamma(B_j)\}_{k \in K}$ on $(G, M, I)$,

$$\bigcap_{k \in K} \bigwedge_{j \in J_k} \gamma(B_j) = \left[ \bigwedge_{k \in K} \bigwedge_{j \in J_k} \gamma(B_j) \right]_{\infty},$$

$$\bigcup_{k \in K} \bigwedge_{j \in J_k} \gamma(B_j) = \left[ \gamma \left( \bigcap_{k \in K} \left( \bigcup_{j \in J_k} B_j \right) \right) \right]_{\infty}.$$

Moreover, $(\mathcal{E}(G, M, I), \subseteq)$ is isomorphic to $(\mathfrak{B}(G), \subseteq)$.

**Proof.** For Eq. (3), suppose $\{\bigwedge_{j \in J_k} \gamma(B_j)\}_{k \in K}$ is a family of conjunctive queries. Because $\bigcap_{k \in K} \bigwedge_{j \in J_k} B_j \subseteq \bigcap_{j \in J_k} B_j$ for any $k \in K$, we have $\left[ \bigwedge_{k \in K} \bigwedge_{j \in J_k} \gamma(B_j) \right]_{\infty}$ as a lower bound of $\left\{ \left[ \bigwedge_{j \in J_k} \gamma(B_j) \right]_{\infty} \right\}_{k \in K}$ in $(\mathcal{E}(G, M, I), \subseteq)$. Suppose $\left[ \bigwedge_{j \in J_k} \gamma(B_j) \right]_{\infty}$ is another lower bound of $\left\{ \left[ \bigwedge_{j \in J_k} \gamma(B_j) \right]_{\infty} \right\}_{k \in K}$ in $(\mathcal{E}(G, M, I), \subseteq)$, then it follows that $\bigcap_{k \in K} B_j \subseteq \bigcap_{j \in J_k} B_j$ for any $k \in K$. This means that $\bigcap_{k \in K} B_j = \bigcap_{k \in K} \bigwedge_{j \in J_k} B_j$ and thus $\bigwedge_{k \in K} \bigwedge_{j \in J_k} \gamma(B_j)_{\infty} \geq \left[ \bigwedge_{k \in K} \bigwedge_{j \in J_k} \gamma(B_j) \right]_{\infty}$. This means that $\left[ \bigwedge_{k \in K} \bigwedge_{j \in J_k} \gamma(B_j) \right]_{\infty}$ is the greatest lower bound of $\left\{ \left[ \bigwedge_{j \in J_k} \gamma(B_j) \right]_{\infty} \right\}_{k \in K}$ in $(\mathcal{E}(G, M, I), \subseteq)$.

For Eq. (4), it is easy to see that $\left[ \gamma \left( \bigcap_{k \in K} \left( \bigcup_{j \in J_k} B_j \right) \right) \right]_{\infty}$ is an upper bound of $\left\{ \left[ \bigwedge_{j \in J_k} \gamma(B_j) \right]_{\infty} \right\}_{k \in K}$ in $(\mathcal{E}(G, M, I), \subseteq)$ because $\bigcap_{k \in K} \left( \bigcup_{j \in J_k} B_j \right)'' = \left( \bigcap_{k \in K} \left( \bigcup_{j \in J_k} B_j \right) \right)' = \left( \bigcap_{k \in K} \left( \bigcup_{j \in J_k} B_j \right) \right)'$ and $\bigwedge_{k \in K} \left( \bigcup_{j \in J_k} B_j \right)'' \subseteq \left( \bigcup_{k \in K} \left( \bigcup_{j \in J_k} B_j \right) \right)''$ for any $k \in K$ by Proposition
2.1. Suppose \( \bigwedge_{p \in P} \gamma(B_p) \rangle = (M, I) \) is another upper bound of \( \left\{ \bigwedge_{j \in J} \gamma(B_j) \right\}_{j \in K} \) in \( \langle (M, I), \leq \rangle \), then \( \bigcup_{k \in K} \left( \bigcap_{j \in J} B_j \right) \leq \bigcap_{p \in P} B_p \).

Since the latter is an extent of \((G, M, I)\) by Theorem 2.1 and \( \left( \bigcup_{k \in K} \left( \bigcap_{j \in J} B_j \right) \right) B_0 = \) the least upper bound containing \( \bigcap_{k \in K} \left( \bigcap_{j \in J} B_j \right) \), we have \( \left( \bigcup_{k \in K} \left( \bigcap_{j \in J} B_j \right) \right) \langle \bigwedge_{p \in P} \gamma(B_p) \rangle = \) the least upper bound of \( \left\{ \bigwedge_{p \in P} \gamma(B_p) \right\}_{p \in \varnothing} \). It follows that \( \left( \bigwedge_{p \in P} \gamma(B_p) \right) \langle \bigwedge_{p \in P} \gamma(B_p) \rangle \rangle = \) is the least upper bound of \( \left\{ \bigwedge_{p \in P} \gamma(B_p) \right\}_{p \in \varnothing} \) in \( \langle (G, M, I), \leq \rangle \).

For the isomorphism between \( \langle (G, M, I), \leq \rangle \) and \( \langle (G, M, I), \leq \rangle \) define a mapping \( \phi : (G, M, I) \rightarrow (G, M, I) \) by \( \phi(A) = A^{\langle \gamma \rangle} \). It is not hard to check that both \( \phi \) and \( \psi \) are order-preserving, \( \phi \psi = \text{id}_{(G, M, I)} \) and \( \psi \phi = \text{id}_{(G, M, I)} \).

We turn to investigate the order-theoretic properties of the disjunctive query systems.

**Proposition 3.2.** Let \((G, M, I)\) be a context. Then for any \( B \subseteq M \) and disjunctive query \( \bigvee_{j \in J} \gamma(B_j) \) on \((G, M, I)\),

1. \( \left\lceil \bigwedge_{j \in J} \gamma(B_j) \right\rceil = \left\lceil \bigwedge_{j \in J} \gamma(B_j) \right\rceil = \bigwedge_{b \in B} \gamma(B) \rangle ; \)
2. \( \left\lceil \bigwedge_{j \in J} \gamma(B_j) \right\rceil = \bigwedge_{j \in J} \bigwedge_{j \in J} \gamma(B_j) \rangle ; \)
3. There exists a family of intents \( \{ \beta_t \}_{t \in T} \) such that \( \left\lfloor \bigvee_{j \in J} \gamma(B_j) \right\rfloor = \bigwedge_{t \in T} T(\gamma(B)) \rangle \).

**Proof.** (1) follows from \( B^\dagger = B^\dagger = \bigcap_{b \in B} B^\dagger \rangle \). (2) is immediate from Proposition 3.2. Finally, (3) is a consequence of (1) and (2).

Intuitively, Item (1) means that the documents retrieved by an elemental disjunctive query are exactly those retrieved by an elemental intential one and they are the common documents retrieved by elemental queries determined by the individual terms. From Proposition 3.2, we have that the retrieved documents associated with the query class \( \bigcup_{t \in T} T(\gamma(B)) \rangle \) can be obtained by merging the documents retrieved by each \( \gamma(B) \rangle \). Thus, Item (2) indicates that the documents retrieved by a disjunctive query is just the union of the documents retrieved by the elemental queries which compose the original disjunctive query. Item (3) means that the documents retrieved by a disjunctive query can be obtained by merging the documents retrieved by a family of individual disjunctive queries determined by intents. In addition, it also indicates that the disjunctive query system can be generated from the Galois lattice of the underlying context up to isomorphism.

**Example 3.4**

1. Following Example 2.1 and considering the disjunctive query \( Q_1 = (L \land K) \lor (L \land D) \lor (W \land F) \), it is easy to calculate that the corresponding collection of retrieved documents is \( \{D_1, D_2, D_3, D_4, D_5, D_6, D_7\} \). Consider another two disjunctive queries \( Q_2 = (L \land G) \lor (W \land D \land M) \) and \( Q_3 = (L \land G \land D) \lor (L \land G \land D \land M) \lor (W \land D \land M) \lor (W \land G \land D \land M) \). From Fig. 1, we can see that the subsets \( \{L, G, W, F, D, M\} \), \( \{L, G, W, D, M\} \), \( \{L, G, W, D, M, SW\} \) and \( \{W, F, D, M\} \) are all intents. Moreover, it is trivial to check that \( \varepsilon(Q_1) = \varepsilon(Q_2) = \varepsilon(Q_3) \).

2. The equation \( \left\lfloor \bigvee_{j \in J} \gamma(B_j) \right\rfloor = \bigwedge_{b \in B} \gamma(B) \rangle \) does not hold necessarily. For instance, considering the disjunctive queries \( (L \land A \land G) \lor (L \land G \land K) \) and \( L \land G \land A \), it is self-evident that their corresponding collections of retrieved documents are \( \{D_1, D_2, D_3\} \) and \( \{D_1, D_2, D_3, D_4\} \), respectively.

Before investigating more properties of disjunctive query systems, we recall a notation: Let \( \{J_k\}_{k \in K} \) be a family of nonempty sets where \( K \) is the index set, then the collection of choice functions for the family \( \{J_k\}_{k \in K} \) is denoted by \( \Pi J_k \) i.e., \( \Pi J_k = \{f : K \rightarrow \bigcup_{k \in K} J_k \mid f(k) \in J_k \text{ for each } k \in K\} \). By the axiom of choice [24], such functions exist.

**Theorem 3.2.** Let \((G, M, I)\) be a context. Then \( \langle (G, M, I), \leq \rangle \) is a complete lattice. In particular, for any family of disjunctive queries \( \left\{ \bigvee_{j \in J} \gamma(B_j) \right\}_{k \in K} \) on \((G, M, I)\),

\[
\bigcup_{k \in K} \left\lceil \bigvee_{j \in J_k} \gamma(B_j) \right\rceil = \left\lceil \bigvee_{k \in K} \bigvee_{j \in J_k} \gamma(B_j) \right\rceil, \tag{5}
\]

\[
\bigwedge_{k \in K} \left\lceil \bigvee_{j \in J_k} \gamma(B_j) \right\rceil = \left\lceil \bigvee_{k \in K} \gamma \left( \bigcup_{j \in J_k} B_j \right) \right\rceil \tag{6}
\]

Moreover, \( \langle (G, M, I), \leq \rangle \) is completely distributive. Particularly, for any family of disjunctive queries \( \{Q_{kh}\}_{k \in H, k \in K} \) on \((G, M, I)\) where \( H \) and \( K \) are nonempty index sets,
\[
\bigcap_{h \in H} \bigcup_{k \in K} [Q_{h_{k}(h)}]_{\infty} = \bigcup_{h \in H} \bigcap_{k \in K} [Q_{h_{k}(h)}]_{\infty}.
\] (7)

**Proof.** Eq. (5) is obvious from Definition 3.2. For Eq. (6), we show \(\bigcap_{k \in K} \left( \bigcup_{j \in J} B'_{j} \right) = \bigcup_{j \in J} \left( \bigcap_{k \in K} B_{j(k)} \right) \) in the following: In fact, suppose \(g \in G\), we have \(g \in \bigcap_{k \in K} \left( \bigcup_{j \in J} B'_{j} \right)\) if and only if \(g \in \bigcup_{j \in J} B'_{j}\) for any \(k \in K\). This is equal to that for any \(k \in K\) there exists \(j \in J\) such that \(g \in B'_{j}\), i.e., there exists \(j \in J\) such that \(g \in \bigcup_{k \in K} B_{j(k)}\), which means \(g \in \bigcup_{j \in J} \left( \bigcap_{k \in K} B_{j(k)} \right)\).

From the above proof, we can see that \(\psi(\bigcap_{k \in K} [Q_{k}]_{\infty}) = \bigcap_{k \in K} \psi([Q_{k}]_{\infty})\) and \(\psi(\bigcup_{k \in K} [Q_{k}]_{\infty}) = \bigcup_{k \in K} \psi([Q_{k}]_{\infty})\) for any family of disjunctive queries \(\{Q_{k}\}_{k \in K}\). Therefore, to prove Eq. (7), it is sufficient to check that \(\bigcap_{h \in H} \left( \bigcup_{k \in K} \psi([Q_{h_{k}(h)}]_{\infty}) \right) = \bigcup_{h \in H} \left( \bigcap_{k \in K} \psi([Q_{h_{k}(h)}]_{\infty}) \right)\) which is trivial. \(\square\)

In the rest of this section, we discuss the connection between the disjunctive query system and the Galois lattice of the underlying context. We note that every disjunctive query \(\bigvee_{t \in T} \gamma_{t}(B_{t})\) on \((G, M, I)\) always corresponds to a family of extents \(\{B'_{t}\}_{t \in T}\). This prompts us to relate every disjunctive query to a subset of \(\mathcal{G}(G)\). However, different disjunctive queries having relation \(\infty\) may correspond to different families of extents in the above sense. Therefore, we need other types of subsets of \(\mathcal{G}(G)\) to investigate the connection between \(\mathcal{D}(G, M, I)\) and \(\mathcal{G}(G)\). To this end, we introduce the following notion.

**Definition 3.4.** Let \(\mathcal{A} = \{x_{t}\}_{t \in T}\) be a family of extents of \((G, M, I)\). We say \(\mathcal{A}\) satisfies the \(\sigma\)-condition if for any extent \(x\) of \((G, M, I)\), \(x \subseteq \bigcup_{t \in T} \mathcal{A}\) always implies \(x \in \mathcal{A}\).

If \(\mathcal{A}\) is a down-set, we can check that \(\mathcal{A}\) satisfies the \(\sigma\)-condition if and only if for any extent \(x\) of \((G, M, I)\), \(x \subseteq \bigcup_{t \in T} \mathcal{A}\) always implies the existence of some \(t \in T\) with \(x \subseteq x_{t}\). For convenience, we use \(\mathcal{D}_{\sigma}(\mathcal{G}(G))\) to denote the set of all down-sets of \(\mathcal{G}(G)\) which satisfy the \(\sigma\)-condition. The following result provides a new perspective to disjunctive query systems via the notion of \(\sigma\)-condition.

**Theorem 3.3.** Let \((G, M, I)\) be a context. Then \((\mathcal{G}(G), \subseteq)\) is join-dense in \((\mathcal{D}_{\infty}(G, M, I), \subseteq)\) up to isomorphism. Moreover, \((\mathcal{D}_{\infty}(G, M, I), \subseteq)\) is isomorphic to \((\mathcal{D}_{\sigma}(\mathcal{G}(G)), \subseteq)\).

**Proof.** Define a mapping \(\psi: \mathcal{G}(G) \rightarrow \mathcal{D}_{\infty}(G, M, I)\) by \(\psi(x) = \bigvee \{x'\}_{\infty}\). It is clear that \(\psi\) is an order-embedding. By Proposition 3.2(3), \(\psi(\mathcal{G}(G))\) is join-dense in \(\mathcal{D}_{\infty}(G, M, I)\).

For any disjunctive query \(Q\) on \((G, M, I)\), we can naturally get a family of extents \(\{x \in \mathcal{G}(G) | x \subseteq \gamma(Q)\}\). Obviously, every such family is a down-set of \(\mathcal{G}(G)\) and satisfies the \(\sigma\)-condition. This allows us to define a mapping \(\psi: \mathcal{D}_{\infty}(G, M, I) \rightarrow \mathcal{D}_{\sigma}(\mathcal{G}(G))\) by \(\psi((Q)_{\infty}) = \{x \in \mathcal{G}(G) | x \subseteq \gamma(Q)\}\). To finish the proof, define another mapping \(\varphi: \mathcal{D}_{\sigma}(\mathcal{G}(G)) \rightarrow \mathcal{D}_{\infty}(G, M, I)\) by \(\varphi(\mathcal{A}) = \bigvee_{x \in \mathcal{A}} \gamma(x)\). It is easy to check that both \(\psi\) and \(\varphi\) are order-preserving and \(\psi \circ \varphi = \text{id}_{\mathcal{D}_{\sigma}(\mathcal{G}(G))}\). For \(\psi \circ \varphi = \text{id}_{\mathcal{D}_{\infty}(G, M, I)}\), we only need to prove that \(\mathcal{A} = \{x' \in \mathcal{G}(G) | x' \subseteq \bigcup_{t \in T} \mathcal{A}\}\) which immediately follows from the fact that \(\mathcal{A}\) satisfies the \(\sigma\)-condition. \(\square\)

![Fig. 2. The Hasse diagram of disjunctive query system of Example 3.5.](image-url)
Example 3.5. Fig. 2 demonstrates the Hasse diagram of the disjunctive query system on the context given in Example 2.1. Since the disjunctive queries are classified according to their retrieved documents, we only present the collections of documents associated with every class of disjunctive queries. In addition, we highlight the nodes which can be join-densely generated from the concepts of the underlying context.

4. Directed disjunctive query systems and algebraic lattices

The vocabulary problem for short queries is one of the primary challenges of the IR systems [8]. It is often difficult for the user to formulate the exact query for the relevant resource. Consequently, finding relevant information by only using the initial query is often unusual. For instance, towards the documents about the topic of formal concept analysis, the query terms employed by users may be “formal concept analysis” or “Galois lattice”, or even only the abbreviation “FCA”. To tackle this problem, the systems are often expected to contain a query reformulation process in order to provide more relevant documents which not only match the initial queries. In this section, we consider a special type of disjunctive queries. Particularly, we assume that the elemental queries contained in such disjunctive queries constitute a directed paradigm in the sense that the relevant resource can be achieved through a successive approximation procedure. To formalize such procedure, we propose the following notion.

Definition 4.1. Let \((G,M,I)\) be a context. A disjunctive query \(\bigvee_{j \in J} \gamma(B_j)\) on \((G,M,I)\) is said to be directed if \(\{B_j\}_{j \in J}\) is directed with respect to \(\leq\), i.e., for any \(j_1,j_2 \in J\), there exists \(j_3 \in J\) such that \(B_{j_1} \cup B_{j_2} \subseteq B_{j_3}\).

Example 4.1. Following Example 2.1 and considering the elemental queries \(\text{AR} \land \text{GL}, \text{LG} \land \text{DM} \land \text{SW} \land \text{LG} \land \text{GL}\), we can compute that their corresponding retrieved document collections are \(\{\text{D2}\}\), \(\{\text{D3}\}\), and \(\{\text{D1}, \text{D2}, \text{D3}\}\), respectively. Therefore, this problem follows immediately from Proposition 2.1.

Example 4.2. Now we give an example to show that some notions in order theory can be described by means of directed disjunctive queries. Let \((L,\leq)\) be a complete lattice and \(1\) is its greatest element. We can construct a context \((G_1, M_1, I_1)\) by letting \(G_1 = M_1 = L\) and \(I_1 = \leq\). Suppose \(D\) is a directed subset of \(L\) which is upper bounded by \(x \in L\). Consider the family of intervals \(\mathcal{A} = \{(d,1) \mid d \in D\}\). Because \(\mathcal{A}\) is filtered with respect to \(\leq\), \(\bigvee_{A \in \mathcal{A}} \gamma(A)\) is a directed disjunctive query on \((G_1, M_1, I_1)\) and \(x \in A\) holds for any \(A \in \mathcal{A}\). Conversely, suppose that \(\bigvee_{B \in \mathcal{B}} \gamma(B)\) is a directed disjunctive query on \((G_2, M_2, I_2)\) which satisfies that \(x \in B\) holds for any \(B \in \mathcal{B}\). One can check that \(D = \{\bigcup B \mid B \in \mathcal{B}\}\) is a directed subset of \(L\) and \(x\) is an upper bound of \(D\).

The following result indicates that a directed disjunctive query has the same retrieved documents as an intential one.

Proposition 4.1. Let \((G,M,I)\) be a context. Then for any directed disjunctive query \(Q\) on \((G,M,I)\), there exists an intential disjunctive query \(\bigvee_{e \in E} \gamma(B_e)\) such that \(\{B_e \mid e \in E\}\) is filtered in \((\mathfrak{P}(G), \subseteq)\) and \(e(Q) = \bigcup_{e \in E} B_e\).

Proof. Suppose \(Q\) has the form \(\bigvee_{j \in J} \gamma(B_j)\). By Definition 4.1, the family \(\{B_j \mid j \in J\}\) is directed in \((\mathfrak{P}(G), \subseteq)\). By Proposition 2.1, the family \(\{B_{j} \mid j \in J\}\) is filtered in \((\mathfrak{P}(L), \subseteq)\). Construct an intential disjunctive query \(\bigvee_{j \in J} \gamma(B_j)\). The equation \(\bigvee_{j \in J} B_j = \bigcup_{j \in J} B_j\) follows immediately from Proposition 2.1. □

The following result presents an equivalent characterization of the directed disjunctive queries. Without stated otherwise, we always use the notation \(F \subseteq X\) to mean that \(F\) is a finite subset of \(X\) in the following.

Proposition 4.2. Let \((G,M,I)\) be a context. For any directed disjunctive query \(Q\) on \((G,M,I)\), the subset \(e(Q)\) satisfies that \(F^\varepsilon \subseteq e(Q)\) for any \(F \subseteq e(Q)\). Conversely, for any \(A \subseteq G\) with the property that \(F^\varepsilon \subseteq A\) for any \(F \subseteq A\), there exists a directed disjunctive query \(Q\) on \((G,M,I)\) such that \(A = e(Q)\).

Proof. Suppose \(Q\) has the form \(\bigvee_{j \in J} \gamma(B_j)\) and \(F \subseteq \bigcup_{j \in J} B_j\). As \(\{B_j \mid j \in J\}\) is directed with respect to \(\leq\), there exists \(j_0 \in J\) such that \(F \subseteq B_{j_0}\). It follows from Proposition 2.1 that \(F^\varepsilon \subseteq B_{j_0}\). Therefore, \(F^\varepsilon \subseteq e(Q)\).

For the converse direction, we show that \(Q = \bigvee_{F \subseteq A} \gamma(F^\varepsilon)\) is the desired disjunctive query: For any \(F_1, F_2 \subseteq A\), let \(F_3 = F_1 \cup F_2\). Then by Proposition 2.1, it holds that \(F_3^\varepsilon \subseteq F_1^\varepsilon \cup F_2^\varepsilon \subseteq F_3^\varepsilon\). This means that \(F^\varepsilon \subseteq A\) is directed. Thus \(\bigvee_{F \subseteq A} \gamma(F^\varepsilon)\) is a directed disjunctive query. Furthermore, for any \(a \in A\), by Proposition 2.1, we have \(a^\varepsilon \subseteq \bigcup_{F \subseteq A} F^\varepsilon\), which implies \(A \subseteq \bigcup_{F \subseteq A} F^\varepsilon\). On the other hand, for any \(g \in \bigcup_{F \subseteq A} F^\varepsilon\), there exists \(F \subseteq A\) such that \(g \in F^\varepsilon\). Since \(F^\varepsilon \subseteq A\) for any \(F \subseteq A\), we have \(g \in A\). This implies that \(\bigcup_{F \subseteq A} F^\varepsilon \subseteq A\). Therefore, we obtain that \(A = \bigcup_{F \subseteq A} F^\varepsilon\), i.e., \(A = e(Q)\). □
Remark 4.1. In [49], Zhang and Shen discussed the connection between FCA, Chu spaces and Domain theory. Observing the system, they introduced the notion of approximable concept as follows: given a context $(G,A)$ where $A$ is a self-contained but also reveal some efficient computational mechanisms in the possible applications. which is different from Zhang’s. Therefore, the detailed discussion on directed disjunctive queries cannot only make this pa-
directed disjunctive query systems provide a new approach to representing algebraic lattices.

fact that approximable concepts provide a representation of algebraic lattices, it immediately follows that directed disjunctive query systems provide a new approach to representing algebraic lattices.

Obviously, the motivation behind the notion of directed disjunctive query is from the FCA-based documents retrieval, which is different from Zhang’s. Therefore, the detailed discussion on directed disjunctive queries cannot only make this paper self-contained but also reveal some efficient computational mechanisms in the possible applications.

In the sequel, we investigate the order-theoretic properties of $(\mathcal{T}_\infty(G,M,I), \preceq)$. The following result shows that $(\mathcal{T}_\infty(G,M,I), \preceq)$ is always a complete lattice.

Proposition 4.3. Let $(G,M,I)$ be a context and $\{Q_t\}_{t \in T}$ a family of directed disjunctive queries on $(G,M,I)$. Then

$$\bigcap_{t \in T} \overline{Q_t} = \left[ \bigvee_{F \subseteq A} \gamma(F) \right]_{\infty_d}$$

where $A = \bigcap_{t \in T} \ell(Q_t)$.

Proof. It is trivial to check that $\bigvee_{F \subseteq A} \gamma(F)$ is a directed disjunctive query. To end the proof, it is sufficient to check that $A = \bigcup_{F \subseteq A} d^\mu$. Indeed, suppose $F \subseteq A$, then $F \subseteq \ell(Q_t)$ holds for any $t \in T$. By Proposition 4.2, it holds that $F^\mu \subseteq \ell(Q_t)$. This implies that $d^\mu \subseteq A$ and thus $\bigcup_{F \subseteq A} d^\mu \subseteq A$. On the other hand, for any $a \in A$, we have $a \in \bigcup_{F \subseteq A} d^\mu$ since $a \in d^\mu$ and $\{a\} \subseteq A$. Thus, $A \subseteq \bigcup_{F \subseteq A} d^\mu$. \hfill $\square$

The following result presents the formulation of the directed join in $(\mathcal{T}_\infty(G,M,I), \preceq)$.

Proposition 4.4. Let $(G,M,I)$ be a context and $\{Q_t\}_{t \in T}$ a family of directed disjunctive queries on $(G,M,I)$ such that $\{\overline{Q_t}\}_{t \in T}$ is directed in $(\mathcal{T}_\infty(G,M,I), \preceq)$. Then

$$\bigcup_{t \in T} \overline{Q_t} = \left[ \bigvee_{F \subseteq D} \gamma(F) \right]_{\infty_d}$$

where $D = \bigcup_{t \in T} \ell(Q_t)$.

Proof. Obviously, $\left[ \bigvee_{F \subseteq D} \gamma(F) \right]_{\infty_d}$ is a directed disjunctive query. To end the proof, we only need to check that $D = \bigcup_{F \subseteq D} d^\mu$. On the one hand, $D \subseteq \bigcup_{F \subseteq D} d^\mu$ is obvious since $d \in d^\mu$ for any $d \in D$. On the other hand, as $\{\overline{Q_t}\}_{t \in T}$ is directed in $(\mathcal{T}_\infty(G,M,I), \preceq)$, it follows that $\overline{\ell(Q_t)}_{t \in T}$ is directed with respect to $\preceq$. Thus, for any $F \subseteq D$, there exists $t_0 \in T$ such that $F \subseteq \ell(Q_{t_0})$. Since $Q_{t_0}$ is a directed disjunctive query, by Proposition 4.2, we have $F^\mu \subseteq \ell(Q_{t_0})$. Thus, $\bigcup_{F \subseteq D} d^\mu \subseteq D$. \hfill $\square$

Based on Proposition 4.4 and the definition of way below relation, we obtain that given two directed disjunctive queries $Q_1$ and $Q_2$ on $(G,M,I)$, $Q_1 \preceq Q_2$ if and only if for any family $\{\overline{Q_t}\}_{t \in T}$ such that $\overline{\ell(Q_t)}_{t \in T}$ is directed with respect to $\preceq$, the condition $\overline{\ell(Q_t)}_{t \in T}$ always implies the existence of $t_0 \in T$ such that $\ell(Q_{t_0}) \subseteq \ell(Q_t)$. The following proposition presents more properties about the way below relation in $(\mathcal{T}_\infty(G,M,I), \preceq)$.

Proposition 4.5. Let $(G,M,I)$ be a context, $F_1$ and $F_2$ finite subsets of $G$. Then

1. $F_1 \subseteq F_2 \Rightarrow [\gamma(F_1')]_{\infty_d} \preceq [\gamma(F_2')]_{\infty_d}$;
2. $[\gamma(F_1')]_{\infty_d} \sqcup [\gamma(F_2')]_{\infty_d} = [\gamma((F_1 \cup F_2)')]_{\infty_d}$.
Theorem 4.2. Let \((G, M, I)\) be a context and \(Q\) a directed disjunctive query on \((G, M, I)\). Then \(|Q|_\prec\) is a compact element in \((\mathfrak{S}_\prec(G, M, I), \preceq)\) if and only if there exists \(F \sqsubseteq G\) such that \(\varepsilon(Q) = F^0\).

Proof. Suppose \(|Q|_\prec\) is a compact element in \((\mathfrak{S}_\prec(G, M, I), \preceq)\). Obviously, \(\{\gamma(F^0)\}_{F \sqsubseteq \varepsilon(Q)}\) is directed in \((\mathfrak{S}_\prec(G, M, I), \preceq)\). For any \(g \in \varepsilon(Q)\), we have \(g \in \bigcup_{F \sqsubseteq \varepsilon(Q)} F^0\) since \(g \in g^0\) and \(g \in \varepsilon(Q)\). Thus, \(|Q|_\prec \subseteq \bigcup_{F \sqsubseteq \varepsilon(Q)} \{\gamma(F^0)\}_{\varepsilon(Q)}\).

In Proposition 4.3, we have shown that \((|Q|_\prec, \subseteq)\) is a directed disjunctive query system such that \(|Q|_\prec\) is another upper bound of \(\{\gamma(F^0)\}_{F \sqsubseteq \varepsilon(Q)}\) and \(\{\gamma(F^0)\}_{F \sqsubseteq \varepsilon(Q)}\). Then \(F^0 \sqsubseteq \varepsilon(Q)\), and \(F^0 \sqsubseteq \varepsilon(Q)\), which immediately follows that \(F^0 \sqsubseteq \varepsilon(Q)\) by Proposition 4.2. It holds that \((F^0 \sqsubseteq \varepsilon(Q))\), which means that \(|Q|_\prec \subseteq \varepsilon(Q)\). Therefore, \(|Q|_\prec \subseteq \varepsilon(Q)\) is the least upper bound of \(|F^0|_\prec\) and \(|F^0|_\prec\) in \((\mathfrak{S}_\prec(G, M, I), \preceq)\).

The following theorem shows that every compact element in \((\mathfrak{S}_\prec(G, M, I), \preceq)\) is determined by a finite subset of \(G\).

Theorem 4.1. Let \((G, M, I)\) be a context and \(Q\) a directed disjunctive query on \((G, M, I)\). Then \(|Q|_\prec\) is a compact element in \((\mathfrak{S}_\prec(G, M, I), \preceq)\) if and only if there exists \(F \sqsubseteq G\) such that \(\varepsilon(Q) = F^0\).

Proof. Suppose \(|Q|_\prec\) is a compact element in \((\mathfrak{S}_\prec(G, M, I), \preceq)\). Obviously, \(\{\gamma(F^0)\}_{F \sqsubseteq \varepsilon(Q)}\) is directed in \((\mathfrak{S}_\prec(G, M, I), \preceq)\). For any \(g \in \varepsilon(Q)\), we have \(g \in \bigcup_{F \sqsubseteq \varepsilon(Q)} F^0\) since \(g \in g^0\) and \(g \in \varepsilon(Q)\). Thus, \(|Q|_\prec \subseteq \bigcup_{F \sqsubseteq \varepsilon(Q)} \{\gamma(F^0)\}_{\varepsilon(Q)}\).

Conversely, suppose there exists \(F \sqsubseteq G\) such that \(|Q|_\prec \subseteq \gamma(F^0)\). By Proposition 4.5(1), \(|Q|_\prec \subseteq \gamma(F^0)\) which means \(|Q|_\prec\) is a compact element.

Theorem 4.2. For any context \((G, M, I)\), \((\mathfrak{S}_\prec(G, M, I), \preceq)\) is an algebraic lattice. Conversely, for any algebraic lattice \((L, \preceq)\), there exists a context \((G, M, I)\) such that \((L, \preceq)\) is isomorphic to \((\mathfrak{S}_\prec(G, M, I), \preceq)\).

Proof. In Proposition 4.3, we have shown that \((\mathfrak{S}_\prec(G, M, I), \preceq)\) is a complete lattice. To end the proof, we only need to verify that \((\mathfrak{S}_\prec(G, M, I), \preceq)\) satisfies the Axiom of Compact Approximation: For any directed disjunctive query \(Q\) on \((G, M, I)\), consider the family \(\{\gamma(F^0)\}_{F \sqsubseteq \varepsilon(Q)}\). By Theorem 4.1, every \(\gamma(F^0)\) is a compact element in \((\mathfrak{S}_\prec(G, M, I), \preceq)\). Moreover, it is obvious that \(\{\gamma(F^0)\}_{F \sqsubseteq \varepsilon(Q)}\) is directed in \((\mathfrak{S}_\prec(G, M, I), \preceq)\). Additionally, it follows from Proposition 4.4 that \(|Q|_\prec \subseteq \bigcup_{F \sqsubseteq \varepsilon(Q)} \{\gamma(F^0)\}_{\varepsilon(Q)}\). This means that \((\mathfrak{S}_\prec(G, M, I), \preceq)\) is an algebraic lattice.

Conversely, consider the context \((K(L), L)\) where \(L \subseteq K(L) \times L\) is defined by \((k, x) \in L\) if and only if \(k \leq x\). In the following, we show that every element of \(\mathfrak{S}_\prec(K(L), L)\) has the form \(\gamma([x])\) for some \(x \in L\): Suppose \(Q\) is a directed disjunctive query on \((K(L), L)\). It is easy to check that \(\varepsilon(Q)\) is a down-set of \(K(L)\). Let \(x = \bigcup \{F \subseteq \varepsilon(Q)\}\). We shall show that this \(x\) is exactly what we desired. To this end, it is sufficient to prove that \(\varepsilon(Q) = [x] \cap K(L)\). Since \(|x|^0 = [x] \cap K(L)\), it is easy to check that \(\varepsilon(Q) \subseteq [x] \cap K(L)\). On the other hand, suppose \(y \in [x] \cap K(L)\). Then, it is clear that \(y \leq x\). Since \(\bigcup \{F \subseteq \varepsilon(Q)\}\) is a directed subset of \(L\), there exists \(F \subseteq \varepsilon(Q)\) such that \(y \leq \bigcup F\). By Proposition 4.2, we have \(\bigcup F \subseteq \varepsilon(Q)\). Additionally, it is crucial to check that \(\bigcup F \subseteq \varepsilon(Q)\). As \(\bigcup F = \bigcup \{F \subseteq \varepsilon(Q)\}\), we can calculate that \(\bigcup F \subseteq \varepsilon(Q)\). Thus, \(\bigcup F \subseteq \varepsilon(Q)\). Because \(\varepsilon(Q)\) is a down-set of \(K(L)\), we have \(y \in \varepsilon(Q)\). It thus implies that \([x] \subseteq K(L) \subseteq \varepsilon(Q)\). Therefore, we obtain \(\bigcup F \subseteq \varepsilon(Q)\) and thus \(|Q|_\prec \subseteq \gamma([x])\) as desired.

Finally, define a mapping \(\phi : L \rightarrow \mathfrak{S}_\prec(K(L), L)\) by \(\phi(x) = \gamma([x])\). It is trivial to check that \(\phi\) is an isomorphism between \((L, \preceq)\) and \((\mathfrak{S}_\prec(K(L), L), \preceq)\).
(i) for any extent $x$ of $(G, M, I)$, $x \subseteq \bigcup \mathcal{A}$ always implies $x \in \mathcal{A}$;
(ii) there exists a sub-family $\mathcal{B} \subseteq \mathcal{A}$ such that $\mathcal{B}$ is directed with respect to $\subseteq$ and $\bigcup \mathcal{B} = \bigcup \mathcal{A}$.

Remark 4.2

(1) Obviously, the condition (i) in Definition 4.2 guarantees that every family of extents satisfying the $q$-condition must satisfy the $\sigma$-condition.
(2) Given a directed family of extents $\mathcal{B}$ of $(G, M, I)$, we can construct a family $\mathcal{A} = \{x \in \mathcal{B}(G) | x \subseteq \bigcup \mathcal{B}\}$. It is trivial to check that $\mathcal{A}$ satisfies the $q$-condition with $\mathcal{B}$ being the directed sub-family in the sense of Definition 4.2.

We use $\mathcal{D}_q(\mathcal{B}(G))$ to denote the set of all down-sets of extents of $(G, M, I)$ which satisfy the $q$-condition. The following result shows that the notion of $\sigma$-condition provides a new perspective to the directed disjunctive query systems.

Theorem 4.3. Let $(G, M, I)$ be a context. Then $(\mathcal{D}_\sigma(G, M, I), \subseteq)$ is isomorphic to $(\mathcal{D}_q(\mathcal{B}(G)), \subseteq)$.

Proof. Define a mapping $\mu : \mathcal{D}_\sigma(G, M, I) \rightarrow \mathcal{D}_q(\mathcal{B}(G))$ by $\mu(Q_{\subseteq}) = \{x \in \mathcal{B}(G) | x \subseteq \bigcup \mathcal{B} \}$ and a mapping $\nu : \mathcal{D}_q(\mathcal{B}(G)) \rightarrow \mathcal{D}_\sigma(G, M, I)$ by $\nu(\mathcal{A}) = \left\{ \bigcup_{x \in \mathcal{A}} x \right\}_{x \in \mathcal{A}}$, where $\mathcal{B}$ is the directed sub-family of $\mathcal{A}$ in the sense of Definition 4.2. It is not hard to verify that both $\mu$ and $\nu$ are order-preserving, $\forall \mu = \text{id}_{\mathcal{D}_\sigma(G, M, I)}$ and $\nu \mu = \text{id}_{\mathcal{D}_q(\mathcal{B}(G))}$. \qed

5. Conclusions and future work

In this paper, we considered the problem of query classification in the FCA-based document retrieval paradigm and investigated the order-theoretic properties of the derived query systems. We first proposed the notions of formal conjunctive and disjunctive query on a formal context. They can be viewed as the formalization of two approaches of generating queries from elemental ones by Boolean operators AND and OR. By introducing equivalences on conjunctive and disjunctive queries according to their corresponding retrieved documents, we developed the notions of conjunctive and disjunctive query system, and then investigated their order-theoretic properties in detail. Our results show that the former one is isomorphic to the Galois lattice of the underlying context and the latter can be join-densely generated from the Galois lattice of the underlying context. Moreover, we proposed a special type of disjunctive query named directed disjunctive query. It is shown that this notion provides a new approach to restructuring algebraic lattices in the sense that the directed disjunctive query system on any context is an algebraic lattice, and conversely, every algebraic lattice is isomorphic to the directed disjunctive query system on an appropriate context.

The notion of continuity is the central idea of Domain theory [18]. From the viewpoint of computing, continuity enables us to successively approximate partial information states by computing information states that contain less information. Historically, it was initiated by Scott during the introduction of continuous lattices, which is a generalization of algebraic lattices [39]. Therefore, algebraic lattice is just a special type of domain structure. Since this paper provides a new approach to restructuring algebraic lattices, it is of interest to study the restructuring of other domain structures such as continuous lattices, algebraic domains and continuous domains by incorporating the idea of continuity in our proposed framework.

Different approaches to generalizing the classical FCA framework into the fuzzy setting have been proposed to explore the potential of FCA in handling uncertain or incomplete information [1,2,5,28,29,32]. Particularly, in [9], the first author of this paper has discussed the fuzzification of approximable concepts developed by Zhang and Shen. However, it is still not satisfactory mainly because the proposed notion therein relies on the fuzzification of finite subsets. From the results of Section 4, we can see that the notion of directed disjunctive query also has the capability of representing algebraic lattices. Moreover, this notion is not defined on the basis of finite subsets. Therefore, it may provide a new approach to the fuzzification of approximable concepts. In future, we shall study the fuzzification of our results.

Acknowledgements

We would like to thank the anonymous referees for their professional comments and valuable suggestions. We also appreciate the communication with Professor Petko Valtchev from Département d’Informatique of Université du Québec à Montréal, especially on the linguistic quality improvement. This work is supported by the National Natural Science Foundation of China (No. 11071061), the National Basic Research Program of China (No. 2011CB311808), and China Postdoctoral Science Foundation (No. 2012M521515).