A Purely Regular Approach to Non-Regular Core Spanners

Markus L. Schmid, Nicole Schweikardt

HU Berlin, Germany

ICDT 2021

Document Spanners

Document Spanners – Basic Definitions

- Σ is an alphabet.
- D is a document over Σ.
- $\mathcal{X}$ is a set of variables.

Document Spanners – Basic Definitions

- Σ is an alphabet.
- D is a document over Σ.
- $\mathcal{X}$ is a set of variables.

Spans: $$\text{Spans}(\mathrm{D}) = \{[i, j\rangle \mid i, j \in [|\,\mathrm{D}\,|+1]\}.$$

Document Spanners – Basic Definitions

- Σ is an alphabet.
- D is a document over Σ.
- $\mathcal{X}$ is a set of variables.

Spans: $\mathsf{Spans}(\mathsf{D}) = \{[i, j\rangle \mid i, j \in [|\mathsf{D}|+1]\}$.

Span-tuples: *partial* function $\mathcal{X} \to \mathsf{Spans}(\mathsf{D})$.

Document Spanners – Basic Definitions

- Σ is an alphabet.
- D is a document over Σ.
- $\mathcal{X}$ is a set of variables.

Spans:	$\mathsf{Spans}(\mathsf{D}) = \{[i, j\rangle \mid i, j \in [|\mathsf{D}|+1]\}$.
Span-tuples:	*partial* function $\mathcal{X} \to \mathsf{Spans}(\mathsf{D})$.
Span-relations:	set of span-tuples.

Document Spanners – Basic Definitions

- Σ is an alphabet.
- D is a document over Σ.
- $\mathcal{X}$ is a set of variables.

Spans:	$\text{Spans}(\text{D}) = \{[i, j\rangle \mid i, j \in [|\text{D}|+1]\}$.
Span-tuples:	*partial* function $\mathcal{X} \to \text{Spans}(\text{D})$.
Span-relations:	set of span-tuples.
(*Document*) *spanners*:	function from Σ^* to span-relations.

Document Spanners – Basic Definitions

Example

$\Sigma = \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$, $\mathcal{X} = \{\mathsf{x}, \mathsf{y}\}$.

Spanner S:

Map $D = d_1 d_2 \ldots d_n$ to all span-tuples t with

- $t(x) = [i, i+1\rangle$, *such that d_i is first occurrence of* $\mathtt{a}$,
- $t(y) = [j, \ell\rangle$, *such that $d_j d_{j+1} \ldots d_{\ell-1} \in \mathtt{c}^*$.*

Document Spanners – Basic Definitions

Example

$\Sigma = \{a, b, c\}$, $\mathcal{X} = \{x, y\}$.

Spanner S:

Map $D = d_1 d_2 \dots d_n$ *to all span-tuples* t *with*

- $t(x) = [i, i{+}1\rangle$, *such that* d_i *is first occurrence of* a,
- $t(y) = [j, \ell\rangle$, *such that* $d_j d_{j+1} \dots d_{\ell-1} \in c^*$.

$S(\text{a b c c a})$ =

x	y
$[1, 2\rangle$	$[3, 4\rangle$
$[1, 2\rangle$	$[4, 5\rangle$
$[1, 2\rangle$	$[3, 5\rangle$

Subword Marked Words

Meta Symbols

For every $x \in \mathcal{X}$, we use meta-symbols (markers) ${}^{\mathsf{x}}\triangleright, \triangleleft^{\mathsf{x}}$.

Subword Marked Words

Meta Symbols

For every $x \in \mathcal{X}$, we use meta-symbols (markers) ${}^{x}\triangleright$, $\triangleleft^{x}$.

Document	Span-Tuple x	Span-Tuple y	Subword Marked Word
abcca	$[1,2\rangle$	$[3,4\rangle$	${}^{x}\triangleright$ a $\triangleleft^{x}$ b ${}^{y}\triangleright$ c $\triangleleft^{y}$ ca
abcca	$[1,2\rangle$	$[4,5\rangle$	${}^{x}\triangleright$ a $\triangleleft^{x}$ bc ${}^{y}\triangleright$ c $\triangleleft^{y}$ a
abcca	$[1,2\rangle$	$[3,5\rangle$	${}^{x}\triangleright$ a $\triangleleft^{x}$ b ${}^{y}\triangleright$ cc $\triangleleft^{y}$ a

Subword Marked Words

Meta Symbols

For every $x \in \mathcal{X}$, we use meta-symbols (markers) ${}^{x}\triangleright$, $\triangleleft^{x}$.

Document	Span-Tuple		Subword Marked Word
	x	y	
abcca	$[1,2\rangle$	$[3,4\rangle$	${}^{x}\triangleright$ a $\triangleleft^{x}$ b ${}^{y}\triangleright$ c $\triangleleft^{y}$ ca
abcca	$[1,2\rangle$	$[4,5\rangle$	${}^{x}\triangleright$ a $\triangleleft^{x}$ bc ${}^{y}\triangleright$ c $\triangleleft^{y}$ a
abcca	$[1,2\rangle$	$[3,5\rangle$	${}^{x}\triangleright$ a $\triangleleft^{x}$ b ${}^{y}\triangleright$ cc $\triangleleft^{y}$ a
$\mathfrak{e}(w)$	$\mathsf{st}(w)$		w

Subword Marked Words

Meta Symbols

For every $x \in \mathcal{X}$, we use meta-symbols (markers) ${}^{\mathsf{x}}\!\triangleright$, $\triangleleft^{\mathsf{x}}$.

Document	Span-Tuple x	Span-Tuple y	Subword Marked Word
abcca	$[1,2\rangle$	$[3,4\rangle$	${}^{\mathsf{x}}\!\triangleright$ a $\triangleleft^{\mathsf{x}}$ b ${}^{\mathsf{y}}\!\triangleright$ c $\triangleleft^{\mathsf{y}}$ ca
abcca	$[1,2\rangle$	$[4,5\rangle$	${}^{\mathsf{x}}\!\triangleright$ a $\triangleleft^{\mathsf{x}}$ bc ${}^{\mathsf{y}}\!\triangleright$ c $\triangleleft^{\mathsf{y}}$ a
abcca	$[1,2\rangle$	$[3,5\rangle$	${}^{\mathsf{x}}\!\triangleright$ a $\triangleleft^{\mathsf{x}}$ b ${}^{\mathsf{y}}\!\triangleright$ cc $\triangleleft^{\mathsf{y}}$ a
$\mathfrak{e}(w)$	$\mathsf{st}(w)$		w

Central Observation

Any subword marked language L represents spanner
$[\![L]\!](\mathsf{D}) = \{\mathsf{st}(w) \mid w \in L, \mathfrak{e}(w) = \mathsf{D}\}$.

Subword Marked Words

Meta Symbols

For every $x \in \mathcal{X}$, we use meta-symbols (markers) ${}^{\mathsf{x}}\triangleright$, $\triangleleft^{\mathsf{x}}$.

Document	Span-Tuple		Subword Marked Word
	x	y	
abcca	$[1,2\rangle$	$[3,4\rangle$	${}^{\mathsf{x}}\triangleright$ a $\triangleleft^{\mathsf{x}}$ b ${}^{\mathsf{y}}\triangleright$ c $\triangleleft^{\mathsf{y}}$ ca
abcca	$[1,2\rangle$	$[4,5\rangle$	${}^{\mathsf{x}}\triangleright$ a $\triangleleft^{\mathsf{x}}$ bc ${}^{\mathsf{y}}\triangleright$ c $\triangleleft^{\mathsf{y}}$ a
abcca	$[1,2\rangle$	$[3,5\rangle$	${}^{\mathsf{x}}\triangleright$ a $\triangleleft^{\mathsf{x}}$ b ${}^{\mathsf{y}}\triangleright$ cc $\triangleleft^{\mathsf{y}}$ a
$\mathfrak{e}(w)$	$\mathrm{st}(w)$		w

Central Observation

Any subword marked language L represents spanner
$[\![L]\!](\mathrm{D}) = \{\mathrm{st}(w) \mid w \in L, \mathfrak{e}(w) = \mathrm{D}\}$.

For every spanner S there is a subw. marked lang. L with $S = [\![L]\!]$.

Regular Spanners (Fagin et al. JACM 2015)

Regular Spanners

A spanner S is regular if $S = [\![L]\!]$ for a *regular* subword marked language L.

Regular Spanners (Fagin et al. JACM 2015)

Regular Spanners

A spanner S is regular if $S = [\![L]\!]$ for a *regular* subword marked language L.

Example

$$(\mathtt{b} \vee \mathtt{c})^{*} \; {}^{\mathsf{x}}\!\triangleright \; \mathtt{a} \; \triangleleft^{\mathsf{x}} \; \Sigma^{*} \; {}^{\mathsf{y}}\!\triangleright \; \mathtt{c}^{+} \; \triangleleft^{\mathsf{y}} \; \Sigma^{*}$$

We use regular expressions for illustrations!

Regular Spanners (Fagin et al. JACM 2015)

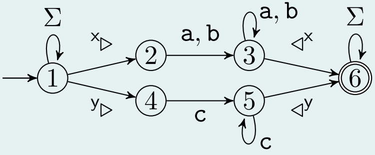
Regular Spanners

A spanner S is regular if $S = [\![L]\!]$ for a *regular* subword marked language L.

Example

$$(\mathtt{b} \vee \mathtt{c})^* \; {}^{x}\!\triangleright \; \mathtt{a} \; \triangleleft^{x} \; \Sigma^* \; {}^{y}\!\triangleright \; \mathtt{c}^+ \; \triangleleft^{y} \; \Sigma^*$$

We use regular expressions for illustrations!



We assume NFA as inputs for algorithms.

Core Spanners (Fagin et al. JACM 2015)

String-Equality Selection

Let R be a span-relation and let $\mathcal{Y} \subseteq \mathcal{X}$.

$\varsigma^{=}_{\mathcal{Y}}(R)$ selects those $t \in R$ such that the y-entries for all $\mathsf{y} \in \mathcal{Y}$ refers to the same subword of D.

Core Spanners (Fagin et al. JACM 2015)

String-Equality Selection

Let R be a span-relation and let $\mathcal{Y} \subseteq \mathcal{X}$.

$\varsigma^{=}_{\mathcal{Y}}(R)$ selects those $t \in R$ such that the y-entries for all y $\in \mathcal{Y}$ refers to the same subword of D.

For spanner S, $\varsigma^{=}_{\mathcal{Y}}(S)$ is the spanner with $(\varsigma^{=}_{\mathcal{Y}}(S))(\mathrm{D}) = \varsigma^{=}_{\mathcal{Y}}(S(\mathrm{D}))$.

Core Spanners (Fagin et al. JACM 2015)

String-Equality Selection

Let R be a span-relation and let $\mathcal{Y} \subseteq \mathcal{X}$.

$\varsigma^{=}_{\mathcal{Y}}(R)$ selects those $t \in R$ such that the y-entries for all $\mathrm{y} \in \mathcal{Y}$ refers to the same subword of D.

For spanner S, $\varsigma^{=}_{\mathcal{Y}}(S)$ is the spanner with $(\varsigma^{=}_{\mathcal{Y}}(S))(\mathrm{D}) = \varsigma^{=}_{\mathcal{Y}}(S(\mathrm{D}))$.

Example

$\mathrm{D} = \mathtt{abccabbc}$, $\mathcal{X} = \{\mathrm{x}, \mathrm{y}, \mathrm{z}\}$, $\mathcal{Y} = \{\mathrm{x}, \mathrm{z}\}$

$\varsigma^{=}_{\mathcal{Y}}($

x	y	z
$[1,2\rangle$	$[3,4\rangle$	$[1,5\rangle$
$[1,3\rangle$	$[5,6\rangle$	$[5,7\rangle$
$[3,4\rangle$	$[8,9\rangle$	$[4,5\rangle$
$[3,4\rangle$	$[8,9\rangle$	$[3,5\rangle$
$[2,3\rangle$	$[1,2\rangle$	$[7,9\rangle$

$)$

Core Spanners (Fagin et al. JACM 2015)

String-Equality Selection

Let R be a span-relation and let $\mathcal{Y} \subseteq \mathcal{X}$.

$\varsigma^{=}_{\mathcal{Y}}(R)$ selects those $t \in R$ such that the y-entries for all $\mathrm{y} \in \mathcal{Y}$ refers to the same subword of D.

For spanner S, $\varsigma^{=}_{\mathcal{Y}}(S)$ is the spanner with $(\varsigma^{=}_{\mathcal{Y}}(S))(\mathrm{D}) = \varsigma^{=}_{\mathcal{Y}}(S(\mathrm{D}))$.

Example

$\mathrm{D} = \mathtt{abccabbc}$, $\mathcal{X} = \{\mathrm{x}, \mathrm{y}, \mathrm{z}\}$, $\mathcal{Y} = \{\mathrm{x}, \mathrm{z}\}$

$\varsigma^{=}_{\mathcal{Y}}($

x	y	z
$[1,2\rangle$	$[3,4\rangle$	$[1,5\rangle$
$[1,3\rangle$	$[5,6\rangle$	$[5,7\rangle$
$[3,4\rangle$	$[8,9\rangle$	$[4,5\rangle$
$[3,4\rangle$	$[8,9\rangle$	$[3,5\rangle$
$[2,3\rangle$	$[1,2\rangle$	$[7,9\rangle$

$) =$

x	y	z
$[1,3\rangle$	$[5,6\rangle$	$[5,7\rangle$
$[3,4\rangle$	$[8,9\rangle$	$[4,5\rangle$
$[2,3\rangle$	$[1,2\rangle$	$[7,9\rangle$

Core Spanners

Core Spanners

A spanner S is a core-spanner if, for some regular spanner S' and $\mathcal{Z}, \mathcal{Y}_1, \ldots, \mathcal{Y}_m \subseteq \mathcal{X}$,

$$S = \pi_{\mathcal{Z}}\, \varsigma^{=}_{\mathcal{Y}_1} \ldots \varsigma^{=}_{\mathcal{Y}_m}(S')$$

Core Spanners

Core Spanners

A spanner S is a core-spanner if, for some regular spanner S' and $\mathcal{Z}, \mathcal{Y}_1, \ldots, \mathcal{Y}_m \subseteq \mathcal{X}$,

$$S = \pi_{\mathcal{Z}} \varsigma^{=}_{\mathcal{Y}_1} \ldots \varsigma^{=}_{\mathcal{Y}_m}(S')$$

Example

reg. spanner: $r = \Sigma^* \, {}^{\mathsf{x}}\!\triangleright \, (\mathtt{a} \vee \mathtt{b})^+ \triangleleft^{\mathsf{x}} \, \Sigma^* \mathtt{c} \Sigma^* \, {}^{\mathsf{y}}\!\triangleright \, (\mathtt{a} \vee \mathtt{b})^+ \triangleleft^{\mathsf{y}} \, \Sigma^*$

core spanner: $\varsigma^{=}_{\{\mathsf{x},\mathsf{y}\}}(\llbracket \mathcal{L}(r) \rrbracket)$

Complexity of Evaluation and Static Analysis

Problem	Question	Regular sp.	Core sp.
Testing	$t \in S(\mathrm{D})$?	?	NP-c
NonEmptiness	$S(\mathrm{D}) \neq \emptyset$?	$O(\lvert M\rvert \lvert \mathrm{D} \rvert)$	NP-c
Satisfiability	$\exists \mathrm{D} : S(\mathrm{D}) \neq \emptyset$	$O(\lvert M\rvert)$	PSpace-c
Containment	$\forall \mathrm{D} : S_1(\mathrm{D}) \subset S_2(\mathrm{D})$?	PSpace-c	undec.
Equivalence	$\forall \mathrm{D} : S_1(\mathrm{D}) = S_2(\mathrm{D})$?	PSpace-c	undec.
Hierarchicality	Does S extract overlapping spans?	$O(\lvert M\rvert \lvert \mathcal{X}\rvert^3)$	PSpace-c

Maturana, Riveros, Vrgoc, PODS 2018

Freydenberger, Holldack, ICDT 2016, ToCS 2018

Ref-Languages and Refl-Spanners

Refl-Languages – Intuition

Main Idea

Regular spanner: $r = \Sigma^* \, {}^{\mathsf{x}}\!\triangleright \, (\mathtt{a} \vee \mathtt{b})^+ \triangleleft^{\mathsf{x}} \, \Sigma^* \mathtt{c} \Sigma^* \, {}^{\mathsf{y}}\!\triangleright \, (\mathtt{a} \vee \mathtt{b})^+ \triangleleft^{\mathsf{y}} \, \Sigma^*$

String equ. sel.: $\varsigma^{=}_{\{\mathsf{x},\mathsf{y}\}}$.

Refl-Languages – Intuition

Main Idea

Regular spanner: $r = \Sigma^* \; {}^{\mathsf{x}}\!\triangleright \; (\mathtt{a} \vee \mathtt{b})^+ \; \triangleleft^{\mathsf{x}} \; \Sigma^* \mathtt{c} \Sigma^* \; {}^{\mathsf{y}}\!\triangleright \; (\mathtt{a} \vee \mathtt{b})^+ \; \triangleleft^{\mathsf{y}} \; \Sigma^*$

String equ. sel.: $\varsigma^{=}_{\{\mathsf{x},\mathsf{y}\}}$.

$$\Downarrow$$

Refl-spanner: $r = \Sigma^* \; {}^{\mathsf{x}}\!\triangleright \; (\mathtt{a} \vee \mathtt{b})^+ \; \triangleleft^{\mathsf{x}} \; \Sigma^* \mathtt{c} \Sigma^* \; {}^{\mathsf{y}}\!\triangleright \; \mathsf{x} \; \triangleleft^{\mathsf{y}} \; \Sigma^*$

x = "whatever is enclosed in ${}^{\mathsf{x}}\!\triangleright \ldots \triangleleft^{\mathsf{x}}$".

Refl-Languages – Intuition

Main Idea

Regular spanner: $r = \Sigma^* \; {}^{\mathsf{x}}\!\triangleright \, (\mathtt{a} \vee \mathtt{b})^+ \triangleleft^{\mathsf{x}} \; \Sigma^* \mathtt{c} \Sigma^* \; {}^{\mathsf{y}}\!\triangleright \, (\mathtt{a} \vee \mathtt{b})^+ \triangleleft^{\mathsf{y}} \; \Sigma^*$

String equ. sel.: $\varsigma^{=}_{\{\mathsf{x},\mathsf{y}\}}$.

$$\Downarrow$$

Refl-spanner: $r = \Sigma^* \; {}^{\mathsf{x}}\!\triangleright \, (\mathtt{a} \vee \mathtt{b})^+ \triangleleft^{\mathsf{x}} \; \Sigma^* \mathtt{c} \Sigma^* \; {}^{\mathsf{y}}\!\triangleright \, \mathsf{x} \triangleleft^{\mathsf{y}} \; \Sigma^*$

x = "whatever is enclosed in ${}^{\mathsf{x}}\!\triangleright \ldots \triangleleft^{\mathsf{x}}$".

Ref-Words

A *ref-word over* Σ *and* $\mathcal{X}$ is a subword-marked word over $\Sigma \cup \mathcal{X}$ and $\mathcal{X}$ with

$$\forall \mathsf{x} \in \mathcal{X} : w = w_1 \, \mathsf{x} \, w_2 \implies \left(w_1 = v_1 \; {}^{\mathsf{x}}\!\triangleright \, v_2 \triangleleft^{\mathsf{x}} \; v_3\right) \wedge \left(\mathsf{x} \notin v_2\right).$$

Refl-Languages – Formal Definition

Dereference-Function

Let w be a ref-word (over Σ and $\mathcal{X}$).
$\mathfrak{d}(w) =$ "replace all x by whatever is in ${}^{\mathsf{x}}\triangleright \ldots \triangleleft^{\mathsf{x}}$".

Refl-Languages – Formal Definition

Dereference-Function

Let w be a ref-word (over Σ and $\mathcal{X}$).
$\mathfrak{d}(w) =$ "replace all x by whatever is in ${}^{\mathrm{x}}\triangleright \ldots \triangleleft^{\mathrm{x}}$".

Example – Dereference-Function

${}^{\mathrm{x}_1}\triangleright\ \mathtt{a}\ {}^{\mathrm{x}_2}\triangleright\ \mathtt{aa}\ \triangleleft^{\mathrm{x}_1}\ \mathtt{c}\mathrm{x}_1\ {}^{\mathrm{x}_3}\triangleright\ \mathtt{ac}\ \triangleleft^{\mathrm{x}_2}\ \mathrm{x}_2\mathtt{a}\mathrm{x}_1 \triangleleft^{\mathrm{x}_3}$

Refl-Languages – Formal Definition

Dereference-Function

Let w be a ref-word (over Σ and $\mathcal{X}$).
$\mathfrak{d}(w)$ = "replace all x by whatever is in ${}^{\mathsf{x}}\triangleright \ldots \triangleleft^{\mathsf{x}}$".

Example – Dereference-Function

${}^{\mathsf{x}_1}\triangleright\ \mathtt{a}\ {}^{\mathsf{x}_2}\triangleright\ \mathtt{aa}\ \triangleleft^{\mathsf{x}_1}\ \mathtt{caaa}\ {}^{\mathsf{x}_3}\triangleright\ \mathtt{ac}\ \triangleleft^{\mathsf{x}_2}\ \mathsf{x}_2\mathtt{aaaa}\triangleleft^{\mathsf{x}_3}$

Refl-Languages – Formal Definition

Dereference-Function

Let w be a ref-word (over Σ and $\mathcal{X}$).
$\mathfrak{d}(w) =$ "replace all x by whatever is in $^{\mathsf{x}}\triangleright \ldots \triangleleft^{\mathsf{x}}$".

Example – Dereference-Function

$^{\mathsf{x}_1}\triangleright$ a $^{\mathsf{x}_2}\triangleright$ aa $\triangleleft^{\mathsf{x}_1}$ caaa $^{\mathsf{x}_3}\triangleright$ ac $\triangleleft^{\mathsf{x}_2}$ aacaaaacaaaa$\triangleleft^{\mathsf{x}_3}$

Refl-Languages – Formal Definition

Dereference-Function

Let w be a ref-word (over Σ and $\mathcal{X}$).
$\mathfrak{d}(w) =$ “replace all x by whatever is in ${}^{\mathsf{x}}\triangleright \ldots \triangleleft^{\mathsf{x}}$”.

Example – Dereference-Function

${}^{\mathsf{x}_1}\triangleright\ \mathsf{a}\ {}^{\mathsf{x}_2}\triangleright\ \mathsf{aa}\ \triangleleft^{\mathsf{x}_1}\ \mathsf{caaa}\ {}^{\mathsf{x}_3}\triangleright\ \mathsf{ac}\ \triangleleft^{\mathsf{x}_2}\ \mathsf{aacaaaacaaaa}\triangleleft^{\mathsf{x}_3}$

Ref-word	subword marked word	document	span-tuple
w	$\mathfrak{d}(w)$	$\mathfrak{e}(\mathfrak{d}(w))$	$\mathrm{st}(\mathfrak{d}(w))$

Refl-Spanners

Ref-Languages

A set L of ref-words (over Σ and $\mathcal{X}$) is a ref-language.

For a ref-language L (over Σ and $\mathcal{X}$), we define $[\![L]\!] = [\![\mathfrak{d}(L)]\!]$.

Refl-Spanners

Ref-Languages

A set L of ref-words (over Σ and $\mathcal{X}$) is a ref-language.

For a ref-language L (over Σ and $\mathcal{X}$), we define $[\![L]\!] = [\![\mathfrak{d}(L)]\!]$.

Refl-Spanners

A spanner S is a refl-spanner if $S = [\![L]\!]$ for a *regular* ref-language L.

Results about Refl-Spanners

Complexity of Refl-Spanners

Problem	Regular sp.	Refl-spanners	Core sp.
Testing	?	poly($\lvert M \rvert$)($\lvert$ D $\rvert + (2\lvert\mathcal{X}\rvert)!$)	NP-c
NonEmptiness	O($\lvert M \rvert \lvert$ D $\rvert$)	NP-c	NP-c
Satisfiability	O($\lvert M \rvert$)	O($\lvert M \rvert$)	PSpace-c
Containment	PSpace-c	ExpSpace (for str. ref.)	undec.
Equivalence	PSpace-c	ExpSpace (for str. ref.)	undec.
Hierarchicality	O($\lvert M \rvert \lvert \mathcal{X} \rvert^3$)	O($\lvert M \rvert \lvert \mathcal{X} \rvert^3$)	PSpace-c

Containment and Equivalence for Regular Spanner

Normalisation (Doleschal et al. PODS 2019, Fagin et al. JACM 2015)

Subword marked languages are *normalised* if all factors over $\{{}^{\mathsf{x}}\triangleright, \triangleleft^{\mathsf{x}} \mid \mathsf{x} \in \mathcal{X}\}$ are ordered in the same way.

Containment and Equivalence for Regular Spanner

Normalisation (Doleschal et al. PODS 2019, Fagin et al. JACM 2015)

Subword marked languages are *normalised* if all factors over $\{{}^{\mathsf{x}}\triangleright, \triangleleft^{\mathsf{x}} \mid \mathsf{x} \in \mathcal{X}\}$ are ordered in the same way.

Theorem

For normalised subword marked languages L and K,

$$L \subseteq K \iff [\![L]\!] \subseteq [\![K]\!] \, .$$

($\Rightarrow$ Containment decidable for regular spanners.)

Containment and Equivalence for Refl-Spanners

Example

$$\mathfrak{d}(\,{}^{\mathsf{x}}\!\triangleright\, \mathtt{ab} \triangleleft^{\mathsf{x}}\, \mathtt{b}\, {}^{\mathsf{y}}\!\triangleright\, \mathtt{abb} \triangleleft^{\mathsf{y}}\, \mathsf{y}\mathtt{ab}) = \mathfrak{d}(\,{}^{\mathsf{x}}\!\triangleright\, \mathtt{ab} \triangleleft^{\mathsf{x}}\, \mathtt{b}\, {}^{\mathsf{y}}\!\triangleright\, \mathsf{x}\mathtt{b} \triangleleft^{\mathsf{y}}\, \mathsf{x}\mathtt{b}\mathsf{x})$$

Containment and Equivalence for Refl-Spanners

Example

$$\mathfrak{d}(\ {}^{x}\!\triangleright\ \mathtt{ab}\ \triangleleft^{x}\ \mathtt{b}\ {}^{y}\!\triangleright\ \mathtt{abb}\ \triangleleft^{y}\ \mathtt{yab}) = \mathfrak{d}(\ {}^{x}\!\triangleright\ \mathtt{ab}\ \triangleleft^{x}\ \mathtt{b}\ {}^{y}\!\triangleright\ \mathtt{xb}\ \triangleleft^{y}\ \mathtt{xbx})$$

Strongly Reference Extracting Ref-Languages

All variable references x occur in the form ${}^{y_x}\!\triangleright\ \mathsf{x} \triangleleft^{y_x}$
(where y_x's only purpose is to extract the span represented by x).

Containment and Equivalence for Refl-Spanners

Example

$$\mathfrak{d}(\ {}^{\mathsf{x}}\!\triangleright\ \mathsf{ab} \triangleleft^{\mathsf{x}}\ \mathsf{b}\ {}^{\mathsf{y}}\!\triangleright\ \mathsf{abb} \triangleleft^{\mathsf{y}}\ \mathsf{yab}) = \mathfrak{d}(\ {}^{\mathsf{x}}\!\triangleright\ \mathsf{ab} \triangleleft^{\mathsf{x}}\ \mathsf{b}\ {}^{\mathsf{y}}\!\triangleright\ \mathsf{xb} \triangleleft^{\mathsf{y}}\ \mathsf{xbx})$$

Strongly Reference Extracting Ref-Languages

All variable references x occur in the form ${}^{\mathsf{y_x}}\!\triangleright\ \mathsf{x} \triangleleft^{\mathsf{y_x}}$
(where $\mathsf{y_x}$'s only purpose is to extract the span represented by x).

Theorem

For strongly reference extracting and normalised ref-languages L and K,

$$L \subseteq K \iff [\![L]\!] \subseteq [\![K]\!] .$$

($\Rightarrow$ Containment decidable for refl-spanners.)

Expressive Power

Refl-Spanners $\rightarrow$ Core Spanners

Every reference-bounded refl-spanner is a core spanner.

Expressive Power

Refl-Spanners → Core Spanners

Every reference-bounded refl-spanner is a core spanner.

How to transform core-spanners into refl-spanners?

$\varsigma^{=}_{\{x,y\}} [\![\mathcal{L}(\,^{x}\triangleright \; a^* \triangleleft^{x} \;\; ^{y}\triangleright \;\; ^{z}\triangleright \; a^* \triangleleft^{z} \; a^* \triangleleft^{y})]\!]$?

Expressive Power

Refl-Spanners $\to$ Core Spanners

Every reference-bounded refl-spanner is a core spanner.

How to transform core-spanners into refl-spanners?

$\varsigma^{=}_{\{x,y\}}[\![\mathcal{L}(\ {}^{x}\!\triangleright\ a^*\ \triangleleft^{x}\ \ {}^{y}\!\triangleright\ \ {}^{z}\!\triangleright\ a^*\ \triangleleft^{z}\ a^*\triangleleft^{y})]\!]$?

$\varsigma^{=}_{\{x,y\}}([\![\mathcal{L}(\ {}^{x}\!\triangleright \ldots\ {}^{y}\!\triangleright \ldots\ \triangleleft^{x} \ldots \triangleleft^{y})]\!])$?

Expressive Power – From Core Spanners to Refl-Spanners

Span-Fusion

$\biguplus_{\{x_1,x_2\}\to z}($

x_1	x_2	x_3
$[1,4\rangle$	$[4,7\rangle$	$[1,10\rangle$
$[3,5\rangle$	$[5,8\rangle$	$[4,7\rangle$
$[1,3\rangle$	$[3,10\rangle$	$[2,4\rangle$

$) =$

z	x_3
$[1,7\rangle$	$[1,10\rangle$
$[3,8\rangle$	$[4,7\rangle$
$[1,10\rangle$	$[2,4\rangle$

Expressive Power – From Core Spanners to Refl-Spanners

Span-Fusion

$$\biguplus_{\{x_1,x_2\}\to z}\left(\begin{array}{|c|c|c|}\hline x_1 & x_2 & x_3 \\ \hline [1,4\rangle & [4,7\rangle & [1,10\rangle \\ \hline [3,5\rangle & [5,8\rangle & [4,7\rangle \\ \hline [1,3\rangle & [3,10\rangle & [2,4\rangle \\ \hline \end{array}\right) = \begin{array}{|c|c|}\hline z & x_3 \\ \hline [1,7\rangle & [1,10\rangle \\ \hline [3,8\rangle & [4,7\rangle \\ \hline [1,10\rangle & [2,4\rangle \\ \hline \end{array}$$

Theorem

Let $S = \varsigma^{=}_{\mathcal{Y}_1} \ldots \varsigma^{=}_{\mathcal{Y}_k}(S')$ be a core spanner, where $\varsigma^{=}_{\mathcal{Y}_1} \ldots \varsigma^{=}_{\mathcal{Y}_k}$ are *non-overlapping* string-equality selections.

Expressive Power – From Core Spanners to Refl-Spanners

Span-Fusion

$\biguplus_{\{x_1, x_2\} \to z}($

x_1	x_2	x_3
$[1,4\rangle$	$[4,7\rangle$	$[1,10\rangle$
$[3,5\rangle$	$[5,8\rangle$	$[4,7\rangle$
$[1,3\rangle$	$[3,10\rangle$	$[2,4\rangle$

$) =$

z	x_3
$[1,7\rangle$	$[1,10\rangle$
$[3,8\rangle$	$[4,7\rangle$
$[1,10\rangle$	$[2,4\rangle$

Theorem

Let $S = \varsigma^{=}_{\mathcal{Y}_1} \ldots \varsigma^{=}_{\mathcal{Y}_k}(S')$ be a core spanner, where $\varsigma^{=}_{\mathcal{Y}_1} \ldots \varsigma^{=}_{\mathcal{Y}_k}$ are *non-overlapping* string-equality selections.

Then there is a refl-spanner S'' such that
$S = \biguplus_{\mathcal{Z}_1 \to z_1} \ldots \biguplus_{\mathcal{Z}_m \to z_m}(S'')$.

Thank you very much for your attention.