

Learning Semantic Parsers

Percy Liang

2014 JHU/CLSP Fred Jelinek Memorial PIRE Workshop

Stanford University



Paleobiology



Paleobiology



The transition between the Proterozoic and Phanerozoic eons, beginning 542 million years (Myr) ago, is distinguished by the diversification of multicellular animals and by their acquisition of mineralized skeletons during the Cambrian period¹. Considerable progress has been made in documenting and more precisely correlating biotic patterns in the Neoproterozoic–Cambrian fossil record with geochemical and physical environmental perturbations^{2–4}, but the mechanisms responsible for these perturbations remain uncertain⁵. Here we use new stratigraphic and geochemical data to show that early Paleozoic marine sediments deposited approximately 540–480 Myr ago record both an expansion in the area of shallow epicontinental seas and anomalous patterns of chemical sedimentation that are indicative of increased oceanic alkalinity and enhanced chemical weathering of continental crust. These geochemical conditions were caused by a protracted period of widespread continental denudation during the Neoproterozoic followed by extensive physical reworking of soil, regolith and basement rock during the first continental-scale marine transgression of the Phanerozoic. The resultant globally occurring stratigraphic surface, which in most regions separates continental crystalline basement rock from much younger Cambrian shallow marine sedimentary deposits, is known as the Great Unconformity⁶. Although Darwin and others have interpreted this widespread hiatus in sedimentation on the continents as a failure of the geologic record, this paleogeomorphic surface represents a unique physical environmental boundary condition that affected seawater chemistry during a time of profound expansion of shallow marine habitats. Thus, the formation of the Great Unconformity may have been an environmental trigger for the evolution of biomineralization and the ‘Cambrian explosion’ of ecologic and taxonomic diversity following the Neoproterozoic emergence of animals.

The term Great Unconformity was first used in the year 1869 to describe the prominent stratigraphic surface in the Grand Canyon that separates the shallow marine, ~525-Myr-old Cambrian Tapeats Sandstone from the underlying metamorphosed, 1,740-Myr-old Vishnu Schist and structurally filled sedimentary rocks of the 1,200–740-Myr-old Grand Canyon Supergroup⁷. The Great Unconformity is well exposed in the Grand Canyon, but this geomorphic surface, which records the erosion and weathering of continental crust followed by sediment accumulation, can be traced across Laurentia and globally, including Gondwana^{8,9}, Baltica¹⁰, Avalonia¹¹ and Siberia¹², making it the most widely recognized and distinctive stratigraphic surface in the rock record. It is also notable because the Cambrian sediments that overlie it in many regions preserve the first adductor-lens crown-group animals, a fact that some paleontologists have interpreted as evidence for stratigraphic bias and an incomplete record of early animal evolution¹³.

Here we use stratigraphic and lithologic data for 21,531 rock units from 830 geographic locations in North America, in conjunction with petrologic and geochemical data (Methods; see also Supplementary Information), to explore the hypothesis that the formation of the Great Unconformity is causally linked to the evolution of biomineralization; this linkage is proposed to occur by means of the geochemical

effects of prolonged continental denudation followed by enhanced physical and chemical weathering of continental crust during terminal Ediacaran and Cambrian time.

The Cambrian- to Early Ordovician-aged sediments of the Sauk Sequence^{14,15} that overlie the Great Unconformity are time-transgressive, such that Early Cambrian sediments occur on the margins of the paleocontinents and Late Cambrian sediments overlie the Great Unconformity in continental interiors (Fig. 1). The spatial extent of the Sauk Sequence is comparable to other Phanerozoic continent-scale sedimentary sequences^{16,17}, but its geological characteristics are unique. In most places, undeformed Cambrian sedimentary rocks deposited on Earth’s surface rest non-conformably on much older continental crystalline basement rocks, many of which were formed and/or metamorphosed within the Earth’s crust (Fig. 2a). Thus, the Great Unconformity marks the termination of an extended period of continental denudation that exhumed and exposed large areas of igneous and metamorphic rocks to subaerial weathering before marine transgression and subsequent sedimentation.

Continental-scale marine transgression during the Cambrian–Early Ordovician accentuated rates of weathering on the Great Unconformity by shifting landward the position of the erosive transgressive shoreline system, often called the ‘wave-base razor’¹⁸, as well as adjacent transitional backshore, aeolian and fluvial systems. As a result, much of the soil and weathered basement rock (regolith) that covered low-relief continental interiors¹⁹ was eroded and mobilized during the transgression, thereby exposing silicate mineral surfaces to weathering over an area that is unprecedented in the rock record (Fig. 2a). This is important because freshly exposed rock weathers chemically at rates more than three times faster than undisturbed soils and regoliths^{20,21}, and

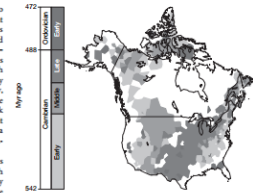


Figure 1 | Sauk Sequence in North America. Distribution and age of the oldest Phanerozoic sedimentary rocks in North America.

¹Department of Geosciences, University of Wisconsin, Madison, Wisconsin 53706, USA. ²Geology Department, Pomona College, Claremont, California 91711, USA.

Paleobiology



The transition between the Proterozoic and Phanerozoic eons, beginning 542 million years (Myr) ago, is distinguished by the diversification of multicellular animals and by their acquisition of mineralized skeletons during the Cambrian period¹. Considerable progress has been made in documenting and more precisely correlating biotic patterns in the Neoproterozoic–Cambrian fossil record with geochemical and physical environmental perturbations^{2,3}, but the mechanisms responsible for these perturbations remain uncertain⁴. Here we use new stratigraphic and geochemical data to show that early Palaeozoic marine sediments deposited approximately 540–480 Myr ago record both an expansion in the area of shallow epicontinental seas and anomalous patterns of chemical sedimentation that are indicative of increased oceanic alkalinity and enhanced chemical weathering of continental crust. These geochemical conditions were caused by a protracted period of widespread continental denudation during the Neoproterozoic, followed by extensive physical reworking of soil, regolith and basement rock during the first continental-scale marine transgression of the Phanerozoic. The resultant globally occurring stratigraphic surface, which in most regions separates continental crystalline basement rock from much younger Cambrian shallow marine sedimentary deposits, is known as the Great Unconformity⁵. Although Darwin and others have interpreted this widespread hiatus in sedimentation on the continents as a failure of the geologic record, this palaeogeomorphic surface represents a unique physical environmental boundary condition that affected seawater chemistry during a time of profound expansion of shallow marine habitats. Thus, the formation of the Great Unconformity may have been an environmental trigger for the evolution of biomineralization and the ‘Cambrian explosion’ of ecologic and taxonomic diversity following the Neoproterozoic emergence of animals.

The term Great Unconformity was first used in the year 1869 to describe the prominent stratigraphic surface in the Grand Canyon that separates the shallow marine, ~525-Myr-old Cambrian Tapeats Sandstone from the underlying metamorphosed, 1,740-Myr-old Vishnu Schist and structurally filled sedimentary rocks of the 1,200–740 Myr-old Grand Canyon Supergroup⁶. The Great Unconformity is well exposed in the Grand Canyon, but this geomorphic surface, which records the erosion and weathering of continental crust followed by sediment accumulation, can be traced across Laurentia and globally, including Gondwana^{7,8}, Baltica⁹, Avalonia¹⁰ and Siberia¹¹, making it the most widely recognized and distinctive stratigraphic surface in the rock record. It is also notable because the Cambrian sediments that overlie it in many regions preserve the first eukaryotic crown-group animals, a fact that some palaeontologists have interpreted as evidence for stratigraphic bias and an incomplete record of early animal evolution¹².

Here we use stratigraphic and lithologic data for 21,531 rock units from 830 geographic locations in North America, in conjunction with petrologic and geochemical data (Methods; see also Supplementary Information), to explore the hypothesis that the formation of the Great Unconformity is causally linked to the evolution of biomineralization; this linkage is proposed to occur by means of the geochemical

effects of prolonged continental denudation followed by enhanced physical and chemical weathering of continental crust during terminal Ediacaran and Cambrian time.

The Cambrian- to Early Ordovician-aged sediments of the Sauk Sequence^{13,14} that overlie the Great Unconformity are time-transgressive, such that Early Cambrian sediments occur on the margins of the palaeocontinents and Late Cambrian sediments overlie the Great Unconformity in continental interiors (Fig. 1). The spatial extent of the Sauk Sequence is comparable to other Phanerozoic continent-scale sedimentary sequences^{15,16}, but its geological characteristics are unique. In most places, undeformed Cambrian sedimentary rocks deposited on Earth’s surface rest non-conformably on much older continental crystalline basement rocks, many of which were formed and/or metamorphosed within the Earth’s crust (Fig. 2a). Thus, the Great Unconformity marks the termination of an extended period of continental denudation that exhumed and exposed large areas of igneous and metamorphic rocks to subaerial weathering before marine transgression and subsequent sedimentation.

Continental-scale marine transgression during the Cambrian–Early Ordovician accentuated rates of weathering on the Great Unconformity by shifting landward the position of the erosive transgressive shoreline system, often called the ‘wave-base razor’¹⁷, as well as adjacent transitional backshore, aeolian and fluvial systems. As a result, much of the soil and weathered basement rock (regolith) that covered low-relief continental interiors¹⁸ was eroded and mobilized during the transgression, thereby exposing silicate mineral surfaces to weathering over an area that is unprecedented in the rock record (Fig. 2a). This is important because freshly exposed rock weathers chemically at rates more than three times faster than undisturbed soils and regolith^{19,20}, and

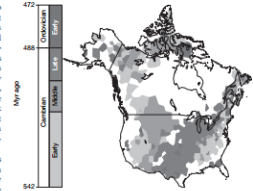


Figure 1 | Sauk Sequence in North America. Distribution and age of the oldest Phanerozoic sedimentary rocks in North America.

¹Department of Geosciences, University of Wisconsin, Madison, Wisconsin 53706, USA. ²Geology Department, Pomona College, Claremont, California 91711, USA.



paleobiodb.org

Paleobiology



The transition between the Proterozoic and Phanerozoic eons, beginning 542 million years (Myr) ago, is distinguished by the diversification of multicellular animals and by their acquisition of mineralized skeletons during the Cambrian period¹. Considerable progress has been made in documenting and more precisely correlating biotic patterns in the Neoproterozoic–Cambrian fossil record with geochemical and physical environmental perturbations^{2,3}, but the mechanisms responsible for these perturbations remain uncertain⁴. Here we use new stratigraphic and geochemical data to show that early Palaeozoic marine sediments deposited approximately 540–600 Myr ago record both an expansion in the area of shallow epicontinental seas and anomalous patterns of chemical sedimentation that are indicative of increased oceanic alkalinity and enhanced chemical weathering of continental crust. These geochemical conditions were caused by a protracted period of widespread continental denudation during the Neoproterozoic, followed by extensive physical reworking of soil, regolith and basement rock during the first continental-scale marine transgression of the Phanerozoic. The resultant globally occurring stratigraphic surface, which in most regions separates continental crystalline basement rock from much younger Cambrian shallow marine sedimentary deposits, is known as the Great Unconformity⁵. Although Darwin and others have interpreted this widespread hiatus in sedimentation on the continents as a failure of the geologic record, this palaeogeomorphic surface represents a unique physical environmental boundary condition that affected seawater chemistry during a time of profound expansion of shallow marine habitats. Thus, the formation of the Great Unconformity may have been an environmental trigger for the evolution of biomineralization and the ‘Cambrian explosion’ of ecologic and taxonomic diversity following the Neoproterozoic emergence of animals.

The term Great Unconformity was first used in the year 1869 to describe the prominent stratigraphic surface in the Grand Canyon that separates the shallow marine, ~525-Myr-old Cambrian Tapeats Sandstone from the underlying metamorphosed, 1,740-Myr-old Vishnu Schist and structurally filled sedimentary rocks of the 1,200–740 Myr-old Grand Canyon Supergroup⁶. The Great Unconformity is well exposed in the Grand Canyon, but this geomorphic surface, which records the erosion and weathering of continental crust followed by sediment accumulation, can be traced across Laurentia and globally, including Gondwana^{7,8}, Baltica⁹, Avalonia¹⁰ and Siberia¹¹, making it the most widely recognized and distinctive stratigraphic surface in the rock record. It is also notable because the Cambrian sediments that overlie it in many regions preserve the first eukaryotic crown-group animals, a fact that some palaeontologists have interpreted as evidence for stratigraphic hiatus and an incomplete record of early animal evolution¹².

Here we use stratigraphic and lithologic data for 21,521 rock units from 830 geographic locations in North America, in conjunction with petrologic and geochemical data (Methods; see also Supplementary Information), to explore the hypothesis that the formation of the Great Unconformity is causally linked to the evolution of biomineralization; this linkage is proposed to occur by means of the geochemical

effects of prolonged continental denudation followed by enhanced physical and chemical weathering of continental crust during terminal Ediacaran and Cambrian time.

The Cambrian- to Early Ordovician-aged sediments of the Sauk Sequence^{13,14} that overlie the Great Unconformity are time-transgressive, such that Early Cambrian sediments occur on the margins of the palaeocontinents and Late Cambrian sediments overlie the Great Unconformity in continental interiors (Fig. 1). The spatial extent of the Sauk Sequence is comparable to other Phanerozoic continent-scale sedimentary sequences^{15,16}, but its geological characteristics are unique. In most places, undeformed Cambrian sedimentary rocks deposited on Earth’s surface rest non-conformably on much older continental crystalline basement rocks, many of which were formed and/or metamorphosed within the Earth’s crust (Fig. 2a). Thus, the Great Unconformity marks the termination of an extended period of continental denudation that exhumed and exposed large areas of igneous and metamorphic rocks to subaerial weathering before marine transgression and subsequent sedimentation.

Continental-scale marine transgression during the Cambrian–Early Ordovician accentuated rates of weathering on the Great Unconformity by shifting landward the position of the erosive transgressive shoreline system, often called the ‘wave-base razor’¹⁷, as well as adjacent transitional backshore, aeolian and fluvial systems. As a result, much of the soil and weathered basement rock (regolith) that covered low-relief continental interiors¹⁸ was eroded and mobilized during the transgression, thereby exposing silicate mineral surfaces to weathering over an area that is unprecedented in the rock record (Fig. 2a). This is important because freshly exposed rock weathers chemically at rates more than three times faster than undisturbed soils and regolith^{19,20}, and

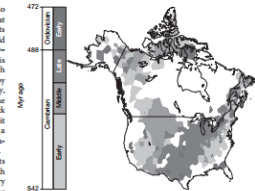


Figure 1 | Sauk Sequence in North America. Distribution and age of the oldest Phanerozoic sedimentary rocks in North America.

¹Department of Geosciences, University of Wisconsin, Madison, Wisconsin 53706, USA. ²Geology Department, Pomona College, Claremont, California 91711, USA.



paleobiodb.org

Where was the last American Mastadon found?

Paleobiology



The transition between the Proterozoic and Phanerozoic eons, beginning 542 million years (Myr) ago, is distinguished by the diversification of multicellular animals and by their acquisition of mineralized skeletons during the Cambrian period¹. Considerable progress has been made in documenting and more precisely correlating biotic patterns in the Neoproterozoic–Cambrian fossil record with geochemical and physical environmental perturbations^{2,3}, but the mechanisms responsible for these perturbations remain uncertain⁴. Here we use new stratigraphic and geochemical data to show that early Palaeozoic marine sediments deposited approximately 540–600 Myr ago record both an expansion in the area of shallow epicontinental seas and anomalous patterns of chemical sedimentation that are indicative of increased oceanic alkalinity and enhanced chemical weathering of continental crust. These geochemical conditions were caused by a protracted period of widespread continental denudation during the Neoproterozoic, followed by extensive physical reworking of soil, regolith and basement rock during the first continental-scale marine transgression of the Phanerozoic. The resultant globally occurring stratigraphic surface, which in most regions separates continental crystalline basement rock from much younger Cambrian shallow marine sedimentary deposits, is known as the Great Unconformity⁵. Although Darwin and others have interpreted this widespread hiatus in sedimentation on the continents as a failure of the geologic record, this palaeogeomorphic surface represents a unique physical environmental boundary condition that affected seawater chemistry during a time of profound expansion of shallow marine habitats. Thus, the formation of the Great Unconformity may have been an environmental trigger for the evolution of biomineralization and the ‘Cambrian explosion’ of ecologic and taxonomic diversity following the Neoproterozoic emergence of animals.

The term Great Unconformity was first used in the year 1869 to describe the prominent stratigraphic surface in the Grand Canyon that separates the shallow marine, ~525-Myr-old Cambrian Tapeats Sandstone from the underlying metamorphosed, 1,740-Myr-old Vishnu Schist and structurally filled sedimentary rocks of the 1,200–740 Myr-old Grand Canyon Supergroup⁶. The Great Unconformity is well exposed in the Grand Canyon, but this geomorphic surface, which records the erosion and weathering of continental crust followed by sediment accumulation, can be traced across Laurentia and globally, including Gondwana^{7,8}, Baltica⁹, Avalonia¹⁰ and Siberia¹¹, making it the most widely recognized and distinctive stratigraphic surface in the rock record. It is also notable because the Cambrian sediments that overlie it in many regions preserve the first eukaryotic crown-group animals, a fact that some palaeontologists have interpreted as evidence for stratigraphic hiatus and an incomplete record of early animal evolution¹².

Here we use stratigraphic and lithologic data for 21,531 rock units from 830 geographic locations in North America, in conjunction with petrologic and geochemical data (Methods; see also Supplementary Information), to explore the hypothesis that the formation of the Great Unconformity is causally linked to the evolution of biomineralization; this linkage is proposed to occur by means of the geochemical

effects of prolonged continental denudation followed by enhanced physical and chemical weathering of continental crust during terminal Ediacaran and Cambrian time.

The Cambrian- to Early Ordovician-aged sediments of the Sauk Sequence^{13,14} that overlie the Great Unconformity are time-transgressive, such that Early Cambrian sediments occur on the margins of the palaeocontinents and Late Cambrian sediments overlie the Great Unconformity in continental interiors (Fig. 1). The spatial extent of the Sauk Sequence is comparable to other Phanerozoic, continent-scale sedimentary sequences^{15,16}, but its geological characteristics are unique. In most places, undeformed Cambrian sedimentary rocks deposited on Earth’s surface rest non-conformably on much older continental crystalline basement rocks, many of which were formed and/or metamorphosed within the Earth’s crust (Fig. 2a). Thus, the Great Unconformity marks the termination of an extended period of continental denudation that exhumed and exposed large areas of igneous and metamorphic rocks to subaerial weathering before marine transgression and subsequent sedimentation.

Continental-scale marine transgression during the Cambrian–Early Ordovician accentuated rates of weathering on the Great Unconformity by shifting landward the position of the erosive transgressive shoreline system, often called the ‘wave-base razor’¹⁷, as well as adjacent tectonically backshore, aeolian and fluvial systems. As a result, much of the soil and weathered basement rock (regolith) that covered low-relief continental interiors¹⁸ was eroded and mobilized during the transgression, thereby exposing silicate mineral surfaces to weathering over an area that is unprecedented in the rock record (Fig. 2a). This is important because freshly exposed rock weathers chemically at rates more than three times faster than undisturbed soils and regoliths^{19,20}, and

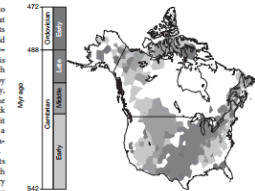


Figure 1 | Sauk Sequence in North America. Distribution and age of the oldest Phanerozoic sedimentary rocks in North America.

¹Department of Geosciences, University of Wisconsin, Madison, Wisconsin 53706, USA. ²Geology Department, Pomona College, Claremont, California 91711, USA.



paleobiodb.org

Where was the last American Mastadon found?

How long do species exist on average before going extinct?

Paleobiology



The transition between the Proterozoic and Phanerozoic eons, beginning 542 million years (Myr) ago, is distinguished by the diversification of multicellular animals and by their acquisition of mineralized skeletons during the Cambrian period¹. Considerable progress has been made in documenting and more precisely correlating biotic patterns in the Neoproterozoic-Cambrian fossil record with geochemical and physical environmental perturbations²⁻⁴, but the mechanisms responsible for these perturbations remains uncertain⁵. Here we use new stratigraphic and geochemical data to show that early Paleozoic marine sediments deposited approximately 540-600 Myr ago record both an expansion in the area of shallow epicontinental seas and anomalous patterns of chemical sedimentation that are indicative of increased oceanic alkalinity and enhanced chemical weathering of continental crust. These geochemical conditions were caused by a protracted period of widespread continental denudation during the Neoproterozoic, followed by extensive physical reworking of soil, regolith and basement rock during the first continental-scale marine transgression of the Phanerozoic. The resultant globally occurring stratigraphic surface, which in most regions separates continental crystalline basement rock from much younger Cambrian shallow marine sedimentary deposits, is known as the Great Unconformity⁶. Although Darwin and others have interpreted this widespread hiatus in sedimentation on the continents as a failure of the geologic record, this paleogeomorphic surface represents a unique physical environmental boundary condition that affected seawater chemistry during a time of profound expansion of shallow marine habitats. Thus, the formation of the Great Unconformity may have been an environmental trigger for the evolution of biomineralization and the 'Cambrian explosion' of ecologic and taxonomic diversity following the Neoproterozoic emergence of animals.

The term Great Unconformity was first used in the year 1869 to describe the prominent stratigraphic surface in the Grand Canyon that separates the shallow marine, ~525-Myr-old Cambrian Tapeats Sandstone from the underlying metamorphosed, 1,740-Myr-old Vishnu Schist and structurally filled sedimentary rocks of the 1,200-740 Myr-old Grand Canyon Supergroup⁷. The Great Unconformity is well exposed in the Grand Canyon, but this geomorphic surface, which records the erosion and weathering of continental crust followed by sediment accumulation, can be traced across Laurentia and globally, including Gondwana^{8,9}, Baltica¹⁰, Avalonia¹¹ and Siberia¹², making it the most widely recognized and distinctive stratigraphic surface in the rock record. It is also notable because the Cambrian sediments that overlie it in many regions preserve the first eukaryotic crown-group animals, a fact that some paleontologists have interpreted as evidence for stratigraphic hiatus and an incomplete record of early animal evolution¹³.

Here we use stratigraphic and lithologic data for 21,521 rock units from 830 geographic locations in North America, in conjunction with petrologic and geochemical data (Methods; see also Supplementary Information), to explore the hypothesis that the formation of the Great Unconformity is causally linked to the evolution of biomineralization; this linkage is proposed to occur by means of the geochemical

effects of prolonged continental denudation followed by enhanced physical and chemical weathering of continental crust during terminal Ediacaran and Cambrian time.

The Cambrian- to Early Ordovician-aged sediments of the Sauk Sequence^{14,15} that overlie the Great Unconformity are time-transgressive, such that Early Cambrian sediments occur on the margins of the palaeocontinents and Late Cambrian sediments overlie the Great Unconformity in continental interiors (Fig. 1). The spatial extent of the Sauk Sequence is comparable to other Phanerozoic continent-scale sedimentary sequences^{16,17}, but its geological characteristics are unique. In most places, undeformed Cambrian sedimentary rocks deposited on Earth's surface rest non-conformably on much older continental crystalline basement rocks, many of which were formed and/or metamorphosed within the Earth's crust (Fig. 2a). Thus, the Great Unconformity marks the termination of an extended period of continental denudation that exhumed and exposed large areas of igneous and metamorphic rocks to subaerial weathering before marine transgression and subsequent sedimentation.

Continental-scale marine transgression during the Cambrian-Early Ordovician accentuated rates of weathering on the Great Unconformity by shifting landward the position of the erosive transgressive shoreline system, often called the 'wave-base razor'¹⁸, as well as adjacent tectonically backshore, aeolian and fluvial systems. As a result, much of the soil and weathered basement rock (regolith) that covered low-relief continental interiors¹⁹ was eroded and mobilized during the transgression, thereby exposing silicate mineral surfaces to weathering over an area that is unprecedented in the rock record (Fig. 2a). This is important because freshly exposed rock weathers chemically at rates more than three times faster than undisturbed soils and regoliths^{20,21}, and

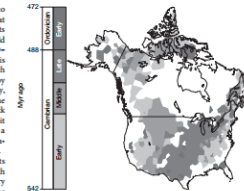


Figure 1 | Sauk Sequence in North America. Distribution and age of the oldest Phanerozoic sedimentary rocks in North America.

¹Department of Geosciences, University of Wisconsin, Madison, Wisconsin 53706, USA. ²Geology Department, Pomona College, Claremont, California 91711, USA.



paleobiodb.org

Where was the last American Mastadon found?

How long do species exist on average before going extinct?

Goal: help scientists answer macro-questions

Challenge: requires computation / aggregation

Question answering via semantic parsing

Where was the last American Mastadon found?

Question answering via semantic parsing

Where was the last American Mastadon found?



semantic parsing

LocationOf.argmax(Type.Occurrence \sqcap Genus.Mammut, Period)

Question answering via semantic parsing

Where was the last American Mastadon found?



semantic parsing

`LocationOf.argmax(Type.Occurrence \sqcap Genus.Mammut, Period)`



execute

New Mexico

Question answering via semantic parsing

Where was the last American Mastadon found?



semantic parsing



execute

New Mexico

Email assistant via semantic parsing

Send a reminder to all authors who haven't sent an abstract.

Email assistant via semantic parsing

Send a reminder to all authors who haven't sent an abstract.



semantic parsing

$\forall x \in (\text{Author} \sqcap \neg \text{Sent.Subject.Abstract}) : \text{Remind}(x)$

Email assistant via semantic parsing

Send a reminder to all authors who haven't sent an abstract.



semantic parsing

$\forall x \in (\text{Author} \sqcap \neg \text{Sent.Subject.Abstract}) : \text{Remind}(x)$



execute

[5 emails sent]

Email assistant via semantic parsing

Send a reminder to all authors who haven't sent an abstract.



semantic parsing



execute

[5 emails sent]

Semantic parsing

[utterance: user input]



semantic parsing

[program]

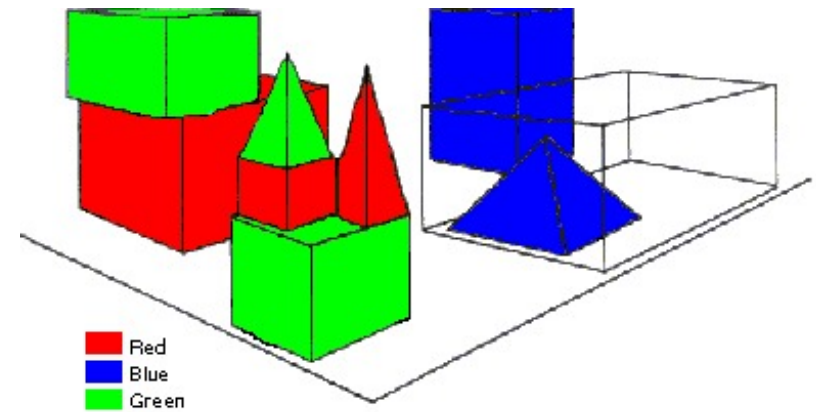


execute

[behavior: user output]

Semantics = how language connects with the world

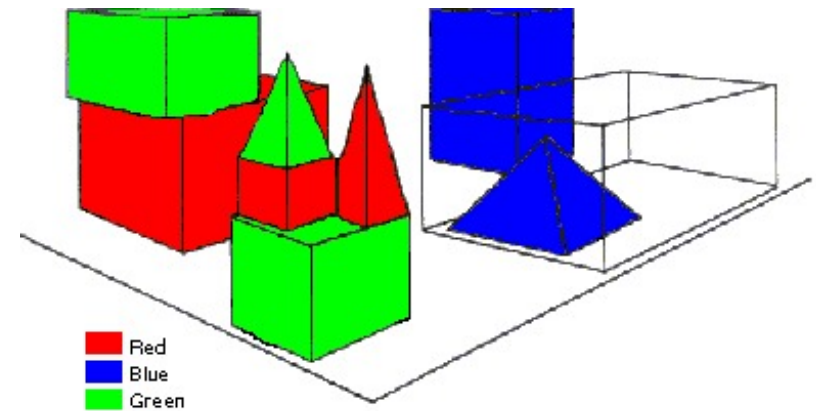
SHRDLU [1971]



SHRDLU [1971]

Person: Pick up a big red block.

Computer: OK.



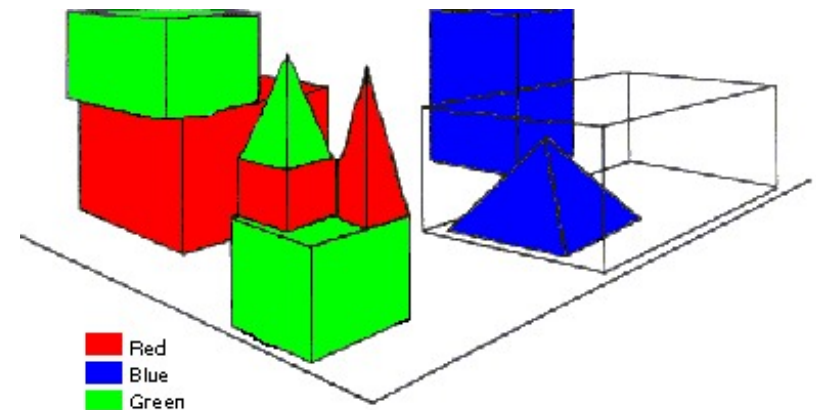
SHRDLU [1971]

Person: Pick up a big red block.

Computer: OK.

Person: Grasp the pyramid.

Computer: I don't understand which pyramid you mean.



SHRDLU [1971]

Person: Pick up a big red block.

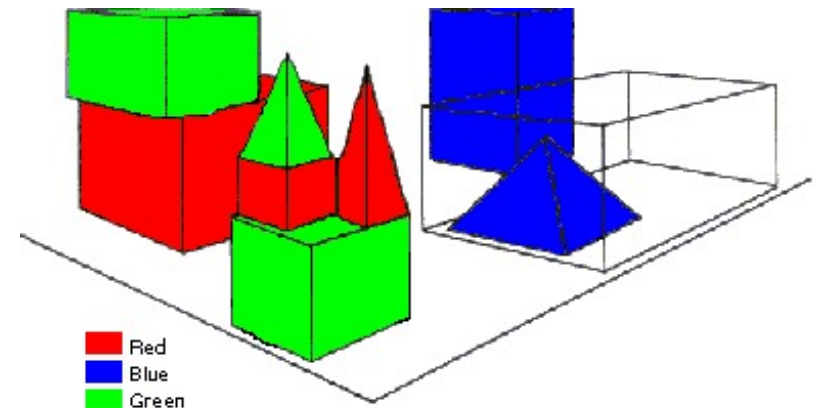
Computer: OK.

Person: Grasp the pyramid.

Computer: I don't understand which pyramid you mean.

Person (changing their mind): Find a block which is taller than the one you are holding and put it into the box.

Computer: By "it", I assume you mean the block which is taller than the one I am holding.



SHRDLU [1971]

Person: Pick up a big red block.

Computer: OK.

Person: Grasp the pyramid.

Computer: I don't understand which pyramid you mean.

Person (changing their mind): Find a block which is taller than the one you are holding and put it into the box.

Computer: By "it", I assume you mean the block which is taller than the one I am holding.

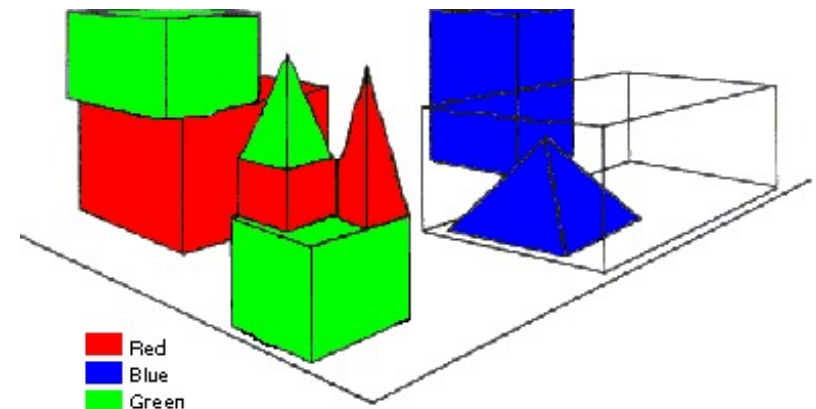
Computer: OK.

Person: What does the box contain?

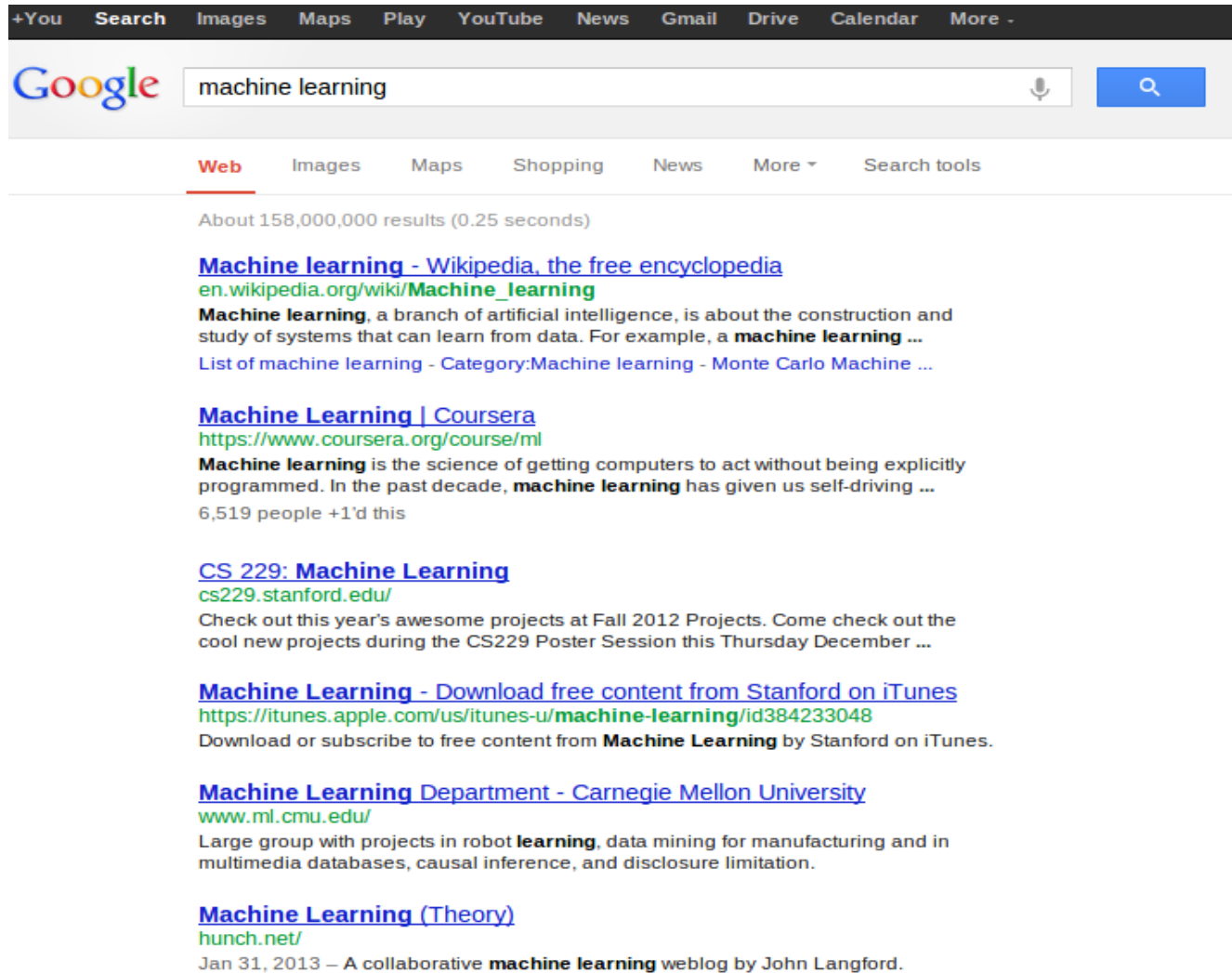
Computer: The blue pyramid and the blue block.

Person: What is the pyramid supported by?

Computer: The box.



Web search [1990s-]



The image shows a screenshot of a Google search results page. At the top, there is a navigation bar with links for '+You', 'Search', 'Images', 'Maps', 'Play', 'YouTube', 'News', 'Gmail', 'Drive', 'Calendar', and 'More -'. Below this is the Google logo and a search bar containing the text 'machine learning'. To the right of the search bar is a microphone icon and a blue search button with a magnifying glass icon. Below the search bar, there is a secondary navigation bar with links for 'Web', 'Images', 'Maps', 'Shopping', 'News', 'More -', and 'Search tools'. The 'Web' link is underlined in red. Below the navigation bars, the search results are displayed. The first result is 'Machine learning - Wikipedia, the free encyclopedia' with the URL 'en.wikipedia.org/wiki/Machine_learning'. The second result is 'Machine Learning | Coursera' with the URL 'https://www.coursera.org/course/ml'. The third result is 'CS 229: Machine Learning' with the URL 'cs229.stanford.edu/'. The fourth result is 'Machine Learning - Download free content from Stanford on iTunes' with the URL 'https://itunes.apple.com/us/itunes-u/machine-learning/id384233048'. The fifth result is 'Machine Learning Department - Carnegie Mellon University' with the URL 'www.ml.cmu.edu/'. The sixth result is 'Machine Learning (Theory)' with the URL 'hunch.net/'.

+You Search Images Maps Play YouTube News Gmail Drive Calendar More -

Google machine learning

Web Images Maps Shopping News More - Search tools

About 158,000,000 results (0.25 seconds)

[Machine learning - Wikipedia, the free encyclopedia](https://en.wikipedia.org/wiki/Machine_learning)
en.wikipedia.org/wiki/Machine_learning
Machine learning, a branch of artificial intelligence, is about the construction and study of systems that can learn from data. For example, a **machine learning** ...
List of machine learning - Category:Machine learning - Monte Carlo Machine ...

[Machine Learning | Coursera](https://www.coursera.org/course/ml)
<https://www.coursera.org/course/ml>
Machine learning is the science of getting computers to act without being explicitly programmed. In the past decade, **machine learning** has given us self-driving ...
6,519 people +1'd this

[CS 229: Machine Learning](https://cs229.stanford.edu/)
cs229.stanford.edu/
Check out this year's awesome projects at Fall 2012 Projects. Come check out the cool new projects during the CS229 Poster Session this Thursday December ...

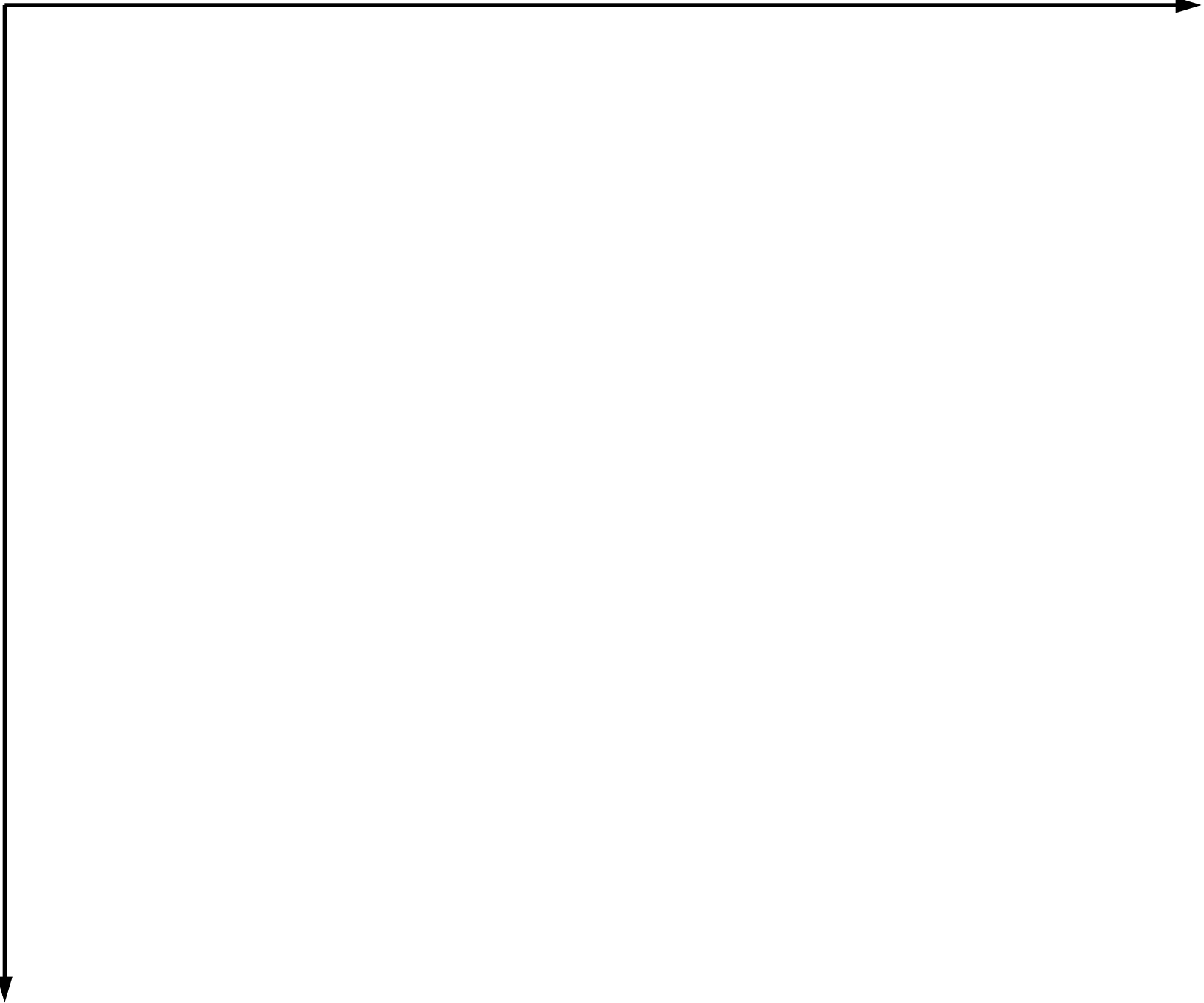
[Machine Learning - Download free content from Stanford on iTunes](https://itunes.apple.com/us/itunes-u/machine-learning/id384233048)
<https://itunes.apple.com/us/itunes-u/machine-learning/id384233048>
Download or subscribe to free content from **Machine Learning** by Stanford on iTunes.

[Machine Learning Department - Carnegie Mellon University](http://www.ml.cmu.edu/)
www.ml.cmu.edu/
Large group with projects in robot **learning**, data mining for manufacturing and in multimedia databases, causal inference, and disclosure limitation.

[Machine Learning \(Theory\)](http://hunch.net/)
hunch.net/
Jan 31, 2013 – A collaborative **machine learning** weblog by John Langford.

Breadth

Depth



Breadth

information retrieval

Depth

Breadth

Depth

information retrieval

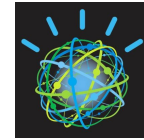
bing™ Google™

Breadth

Depth

information retrieval

bing™ Google™

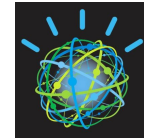


Breadth

Depth

information retrieval

bing™ Google™

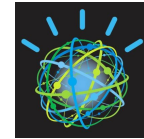


semantic parsing

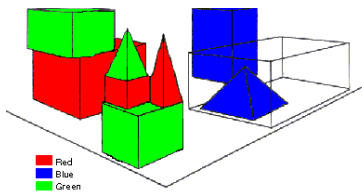
Breadth

information retrieval

bing™ Google™



Depth



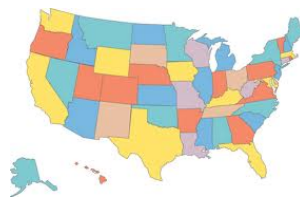
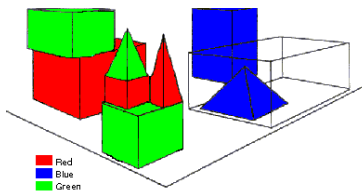
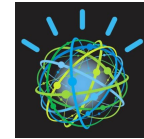
semantic parsing

Breadth

Depth

information retrieval

bing™ Google™

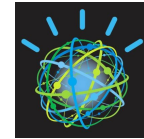


semantic parsing

Breadth

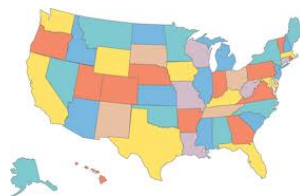
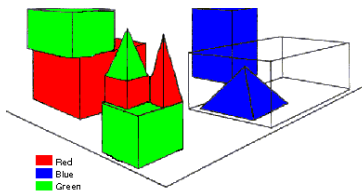
information retrieval

bing™ Google™

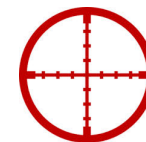


Objective: to develop semantic parsers with the modern sensibilities of web search.

Depth



semantic parsing



Training data for semantic parsing

Heavy supervision

What's Bulgaria's capital?

Capital.Bulgaria

When was Walmart started?

DateFounded.Walmart

What movies has Tom Cruise been in?

Type.Movie \square Starring.TomCruise

...

Training data for semantic parsing

Heavy supervision

What's Bulgaria's capital?

Capital.Bulgaria

When was Walmart started?

DateFounded.Walmart

What movies has Tom Cruise been in?

Type.Movie \square Starring.TomCruise

...



Training data for semantic parsing

Heavy supervision

What's Bulgaria's capital?

Capital.Bulgaria

When was Walmart started?

DateFounded.Walmart

What movies has Tom Cruise been in?

Type.Movie \square Starring.TomCruise

...

Light supervision

What's Bulgaria's capital?

Sofia

When was Walmart started?

1962

What movies has Tom Cruise been in?

TopGun, VanillaSky, ...

...



Training data for semantic parsing

Heavy supervision

What's Bulgaria's capital?

Capital.Bulgaria

When was Walmart started?

DateFounded.Walmart

What movies has Tom Cruise been in?

Type.Movie \square Starring.TomCruise

...

Light supervision

What's Bulgaria's capital?

Sofia

When was Walmart started?

1962

What movies has Tom Cruise been in?

TopGun, VanillaSky, ...

...



Training data for semantic parsing

Heavy supervision

What's Bulgaria's capital?

Capital.Bulgaria

When was Walmart started?

DateFounded.Walmart

What movies has Tom Cruise been in?

Type.Movie \square Starring.TomCruise

...

Light supervision

What's Bulgaria's capital?

Sofia

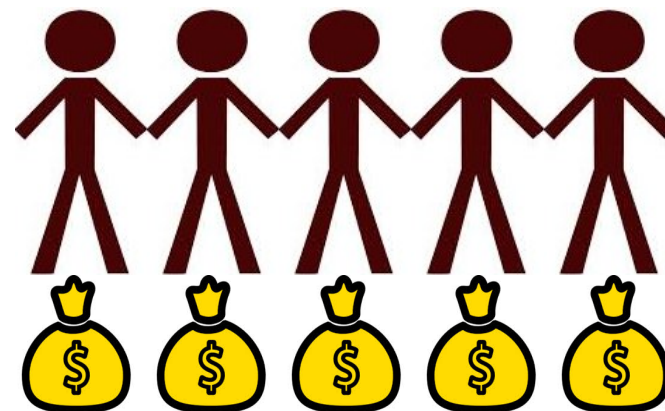
When was Walmart started?

1962

What movies has Tom Cruise been in?

TopGun, VanillaSky, ...

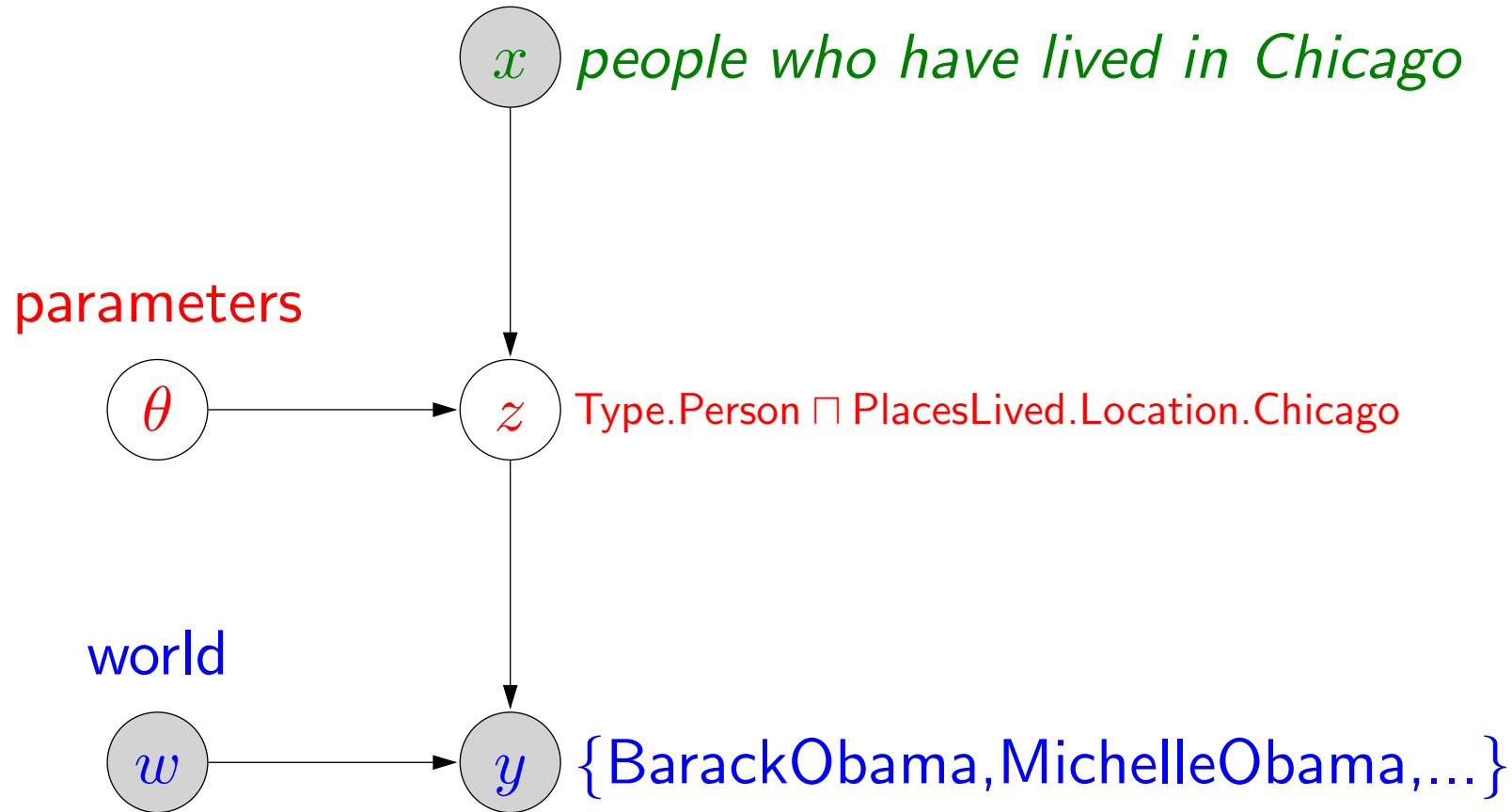
...



Outline

- **A semantic parsing framework**
- A closer look at the elements
 - Logical forms: lambda DCS
 - Lexical coverage
 - Grammar: building logical forms
 - Learning via bootstrapping
 - Leveraging denotations
 - Datasets/results
- Beyond Freebase
- Final remarks

Probabilistic framework

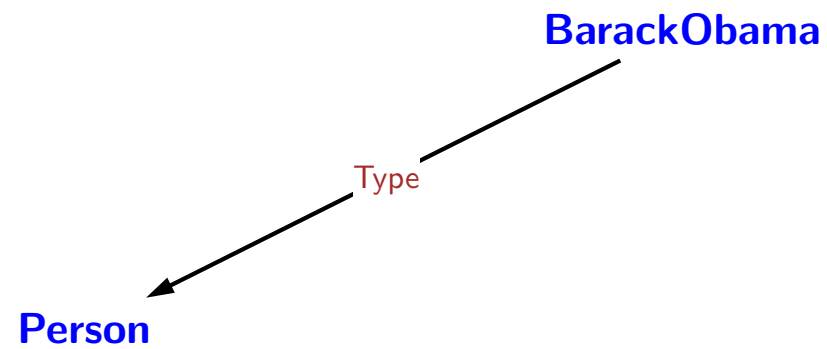


World: Freebase

100M **entities** (nodes) 1B **assertions** (edges)

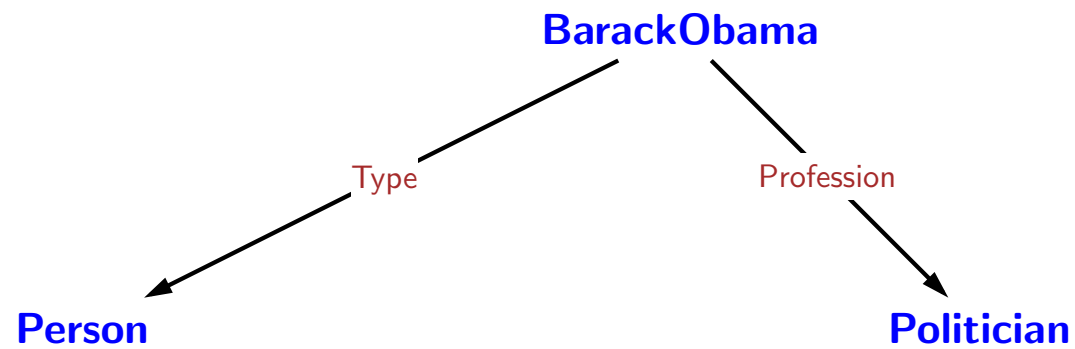
World: Freebase

100M **entities** (nodes) 1B **assertions** (edges)



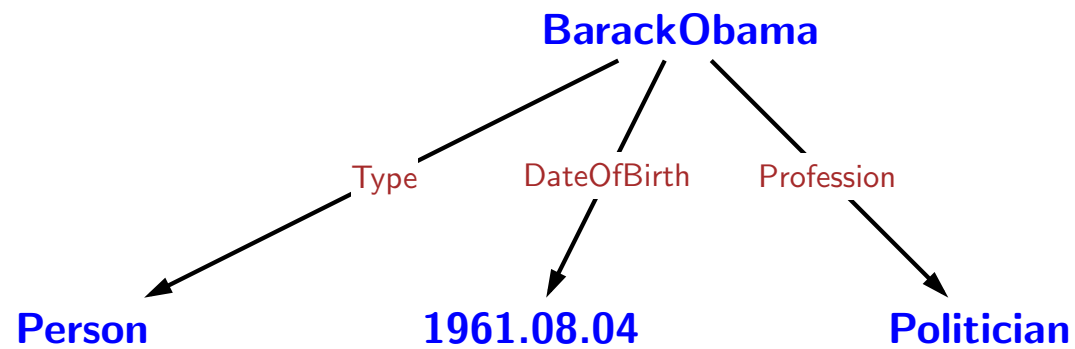
World: Freebase

100M **entities** (nodes) 1B **assertions** (edges)



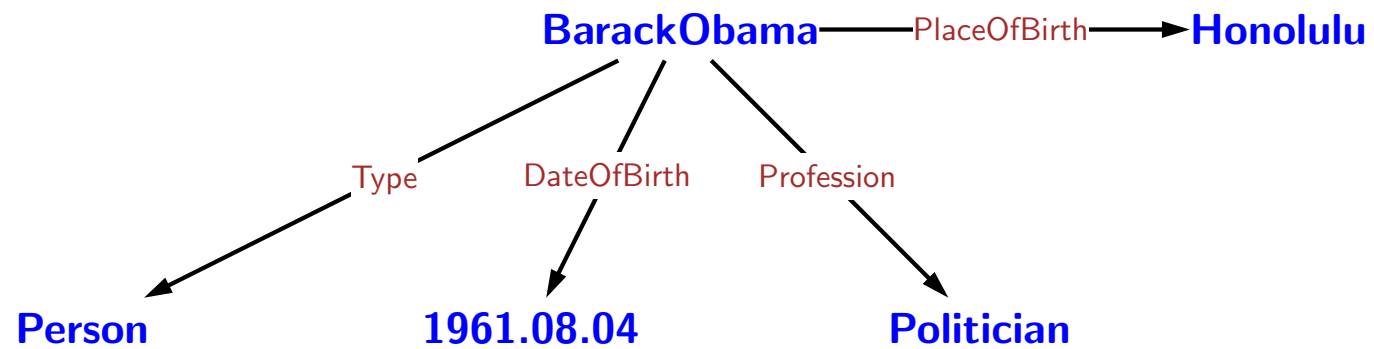
World: Freebase

100M **entities** (nodes) 1B **assertions** (edges)



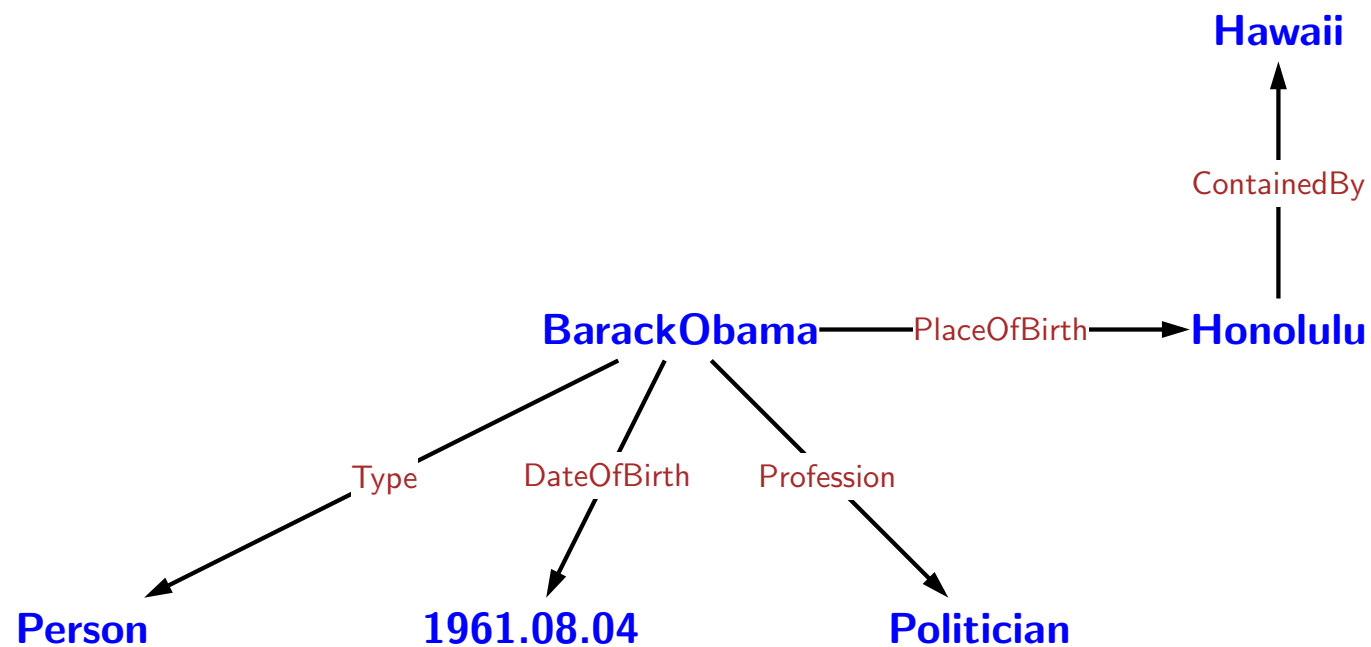
World: Freebase

100M **entities** (nodes) 1B **assertions** (edges)



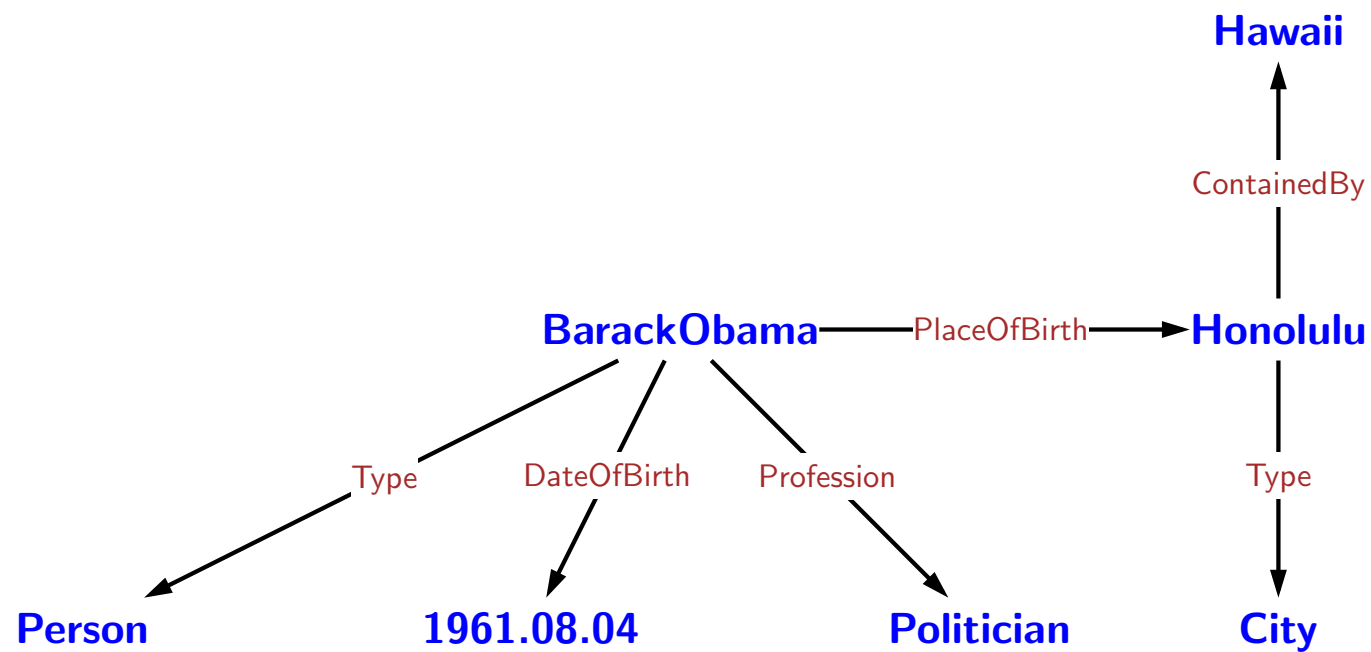
World: Freebase

100M **entities** (nodes) 1B **assertions** (edges)



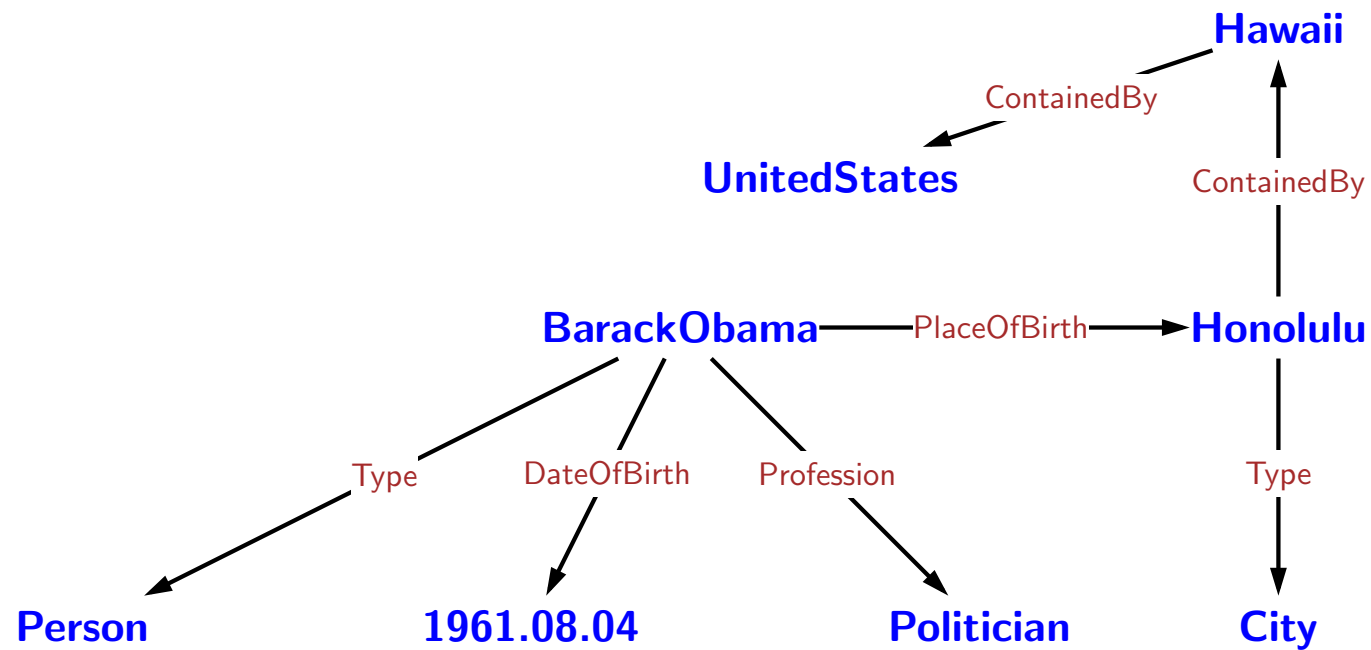
World: Freebase

100M **entities** (nodes) 1B **assertions** (edges)



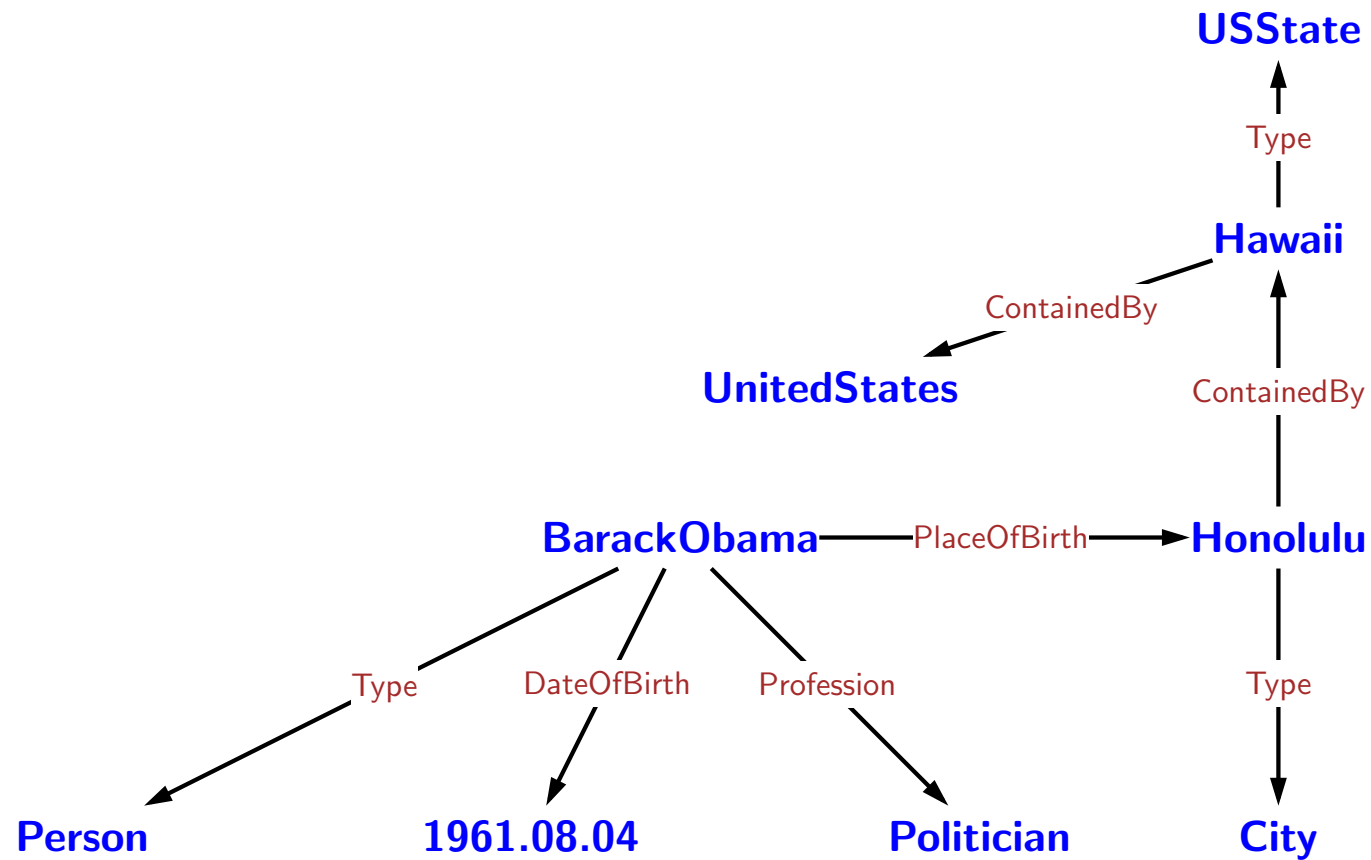
World: Freebase

100M **entities** (nodes) 1B **assertions** (edges)



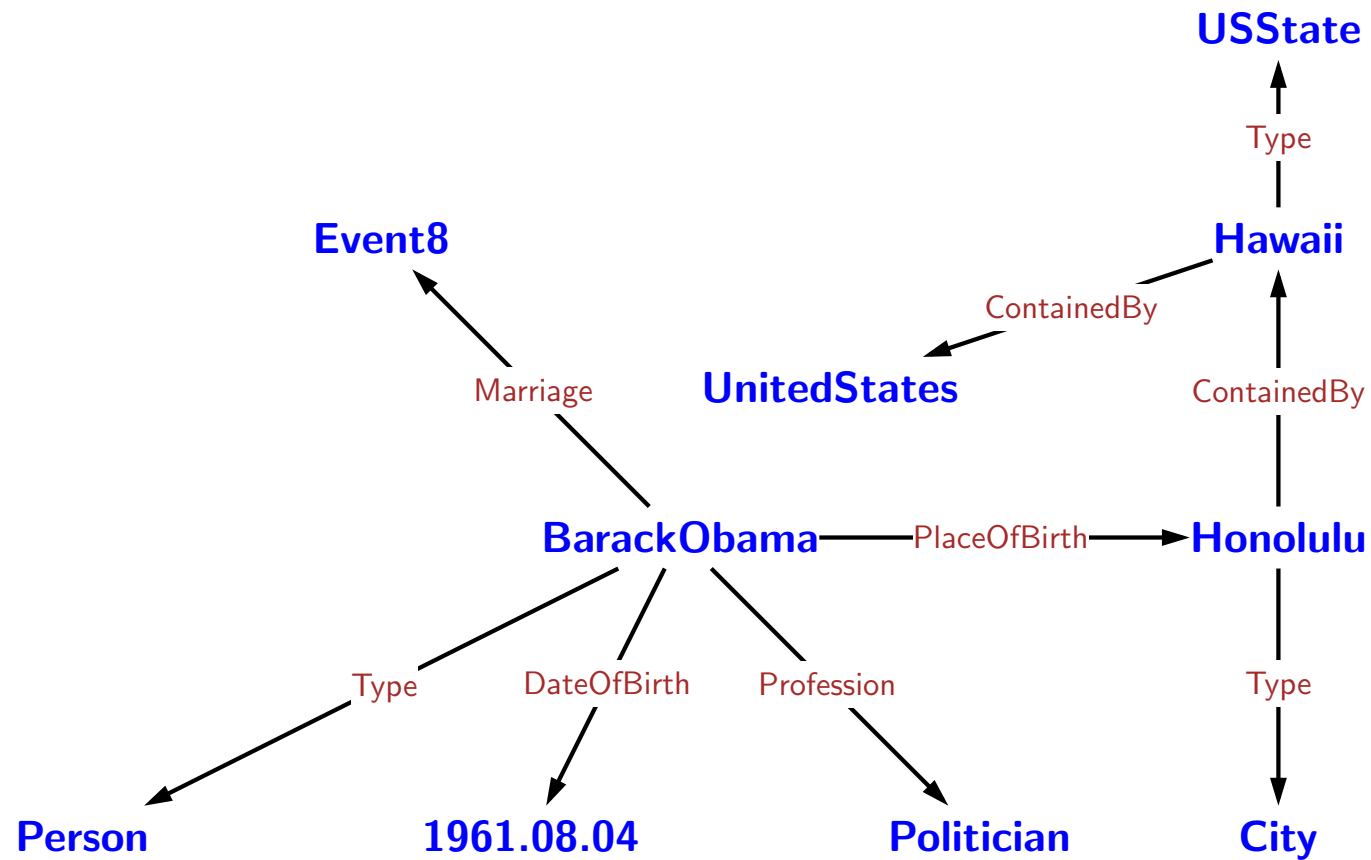
World: Freebase

100M **entities** (nodes) 1B **assertions** (edges)



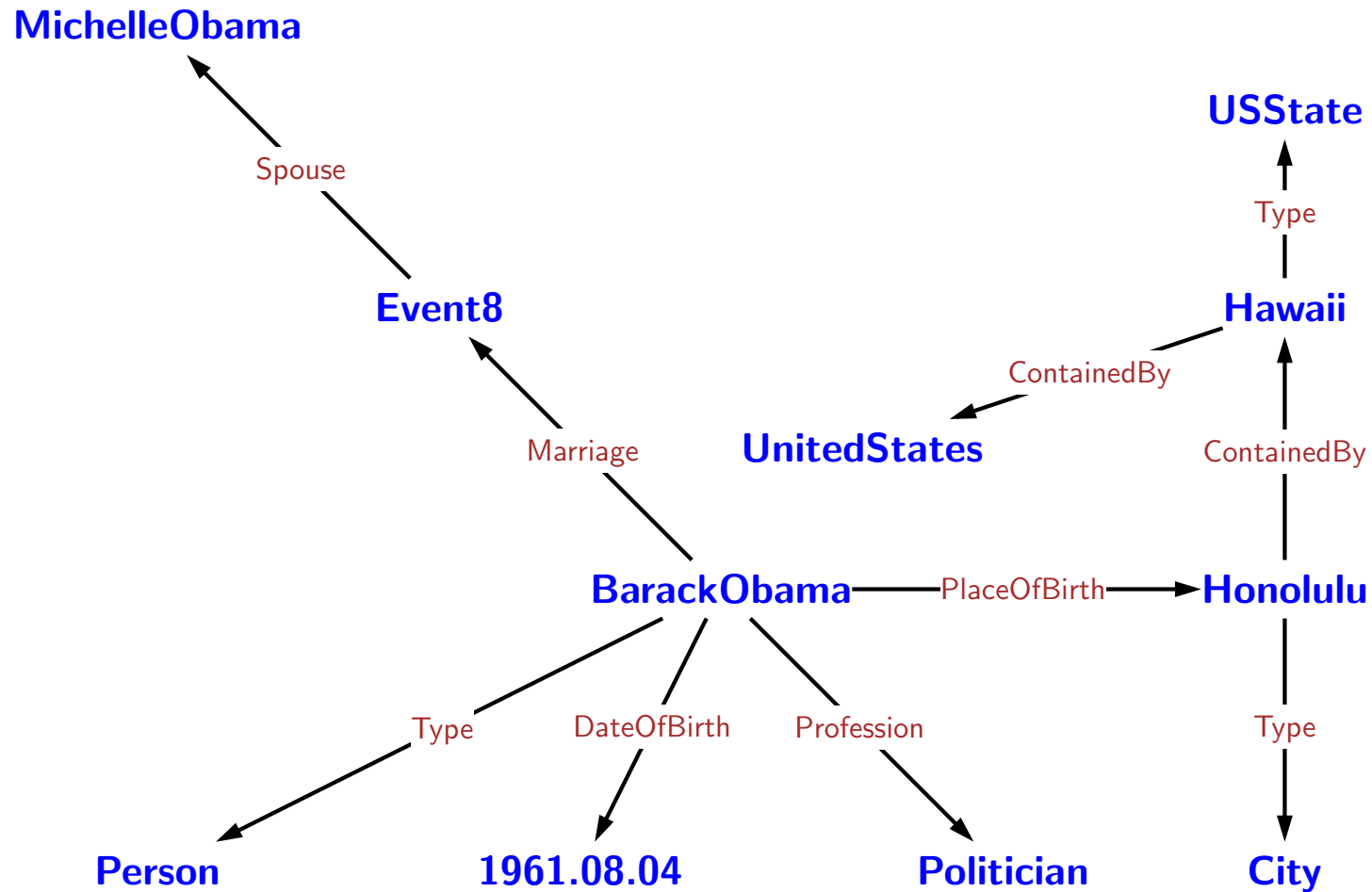
World: Freebase

100M **entities** (nodes) 1B **assertions** (edges)



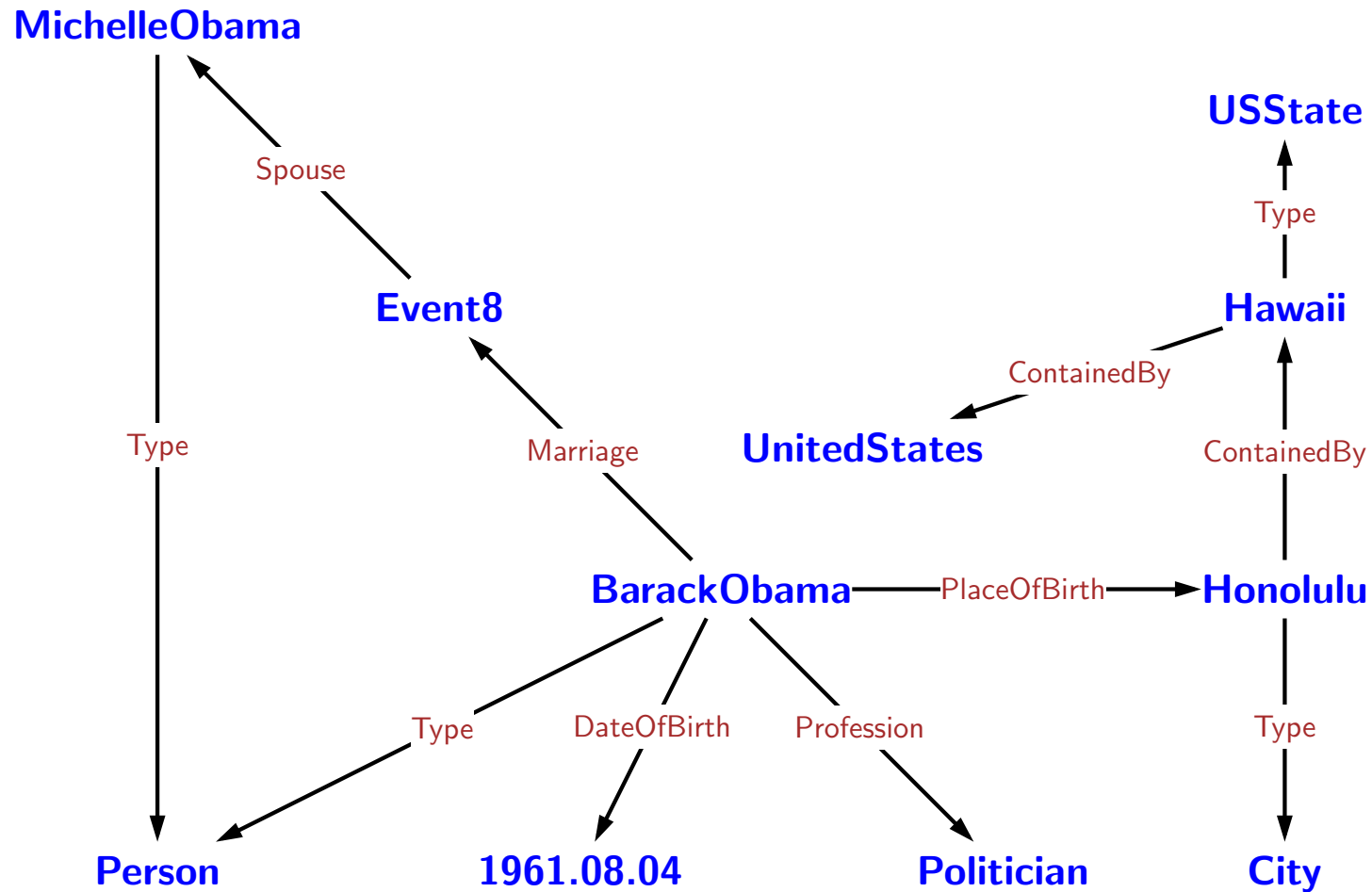
World: Freebase

100M **entities** (nodes) 1B **assertions** (edges)



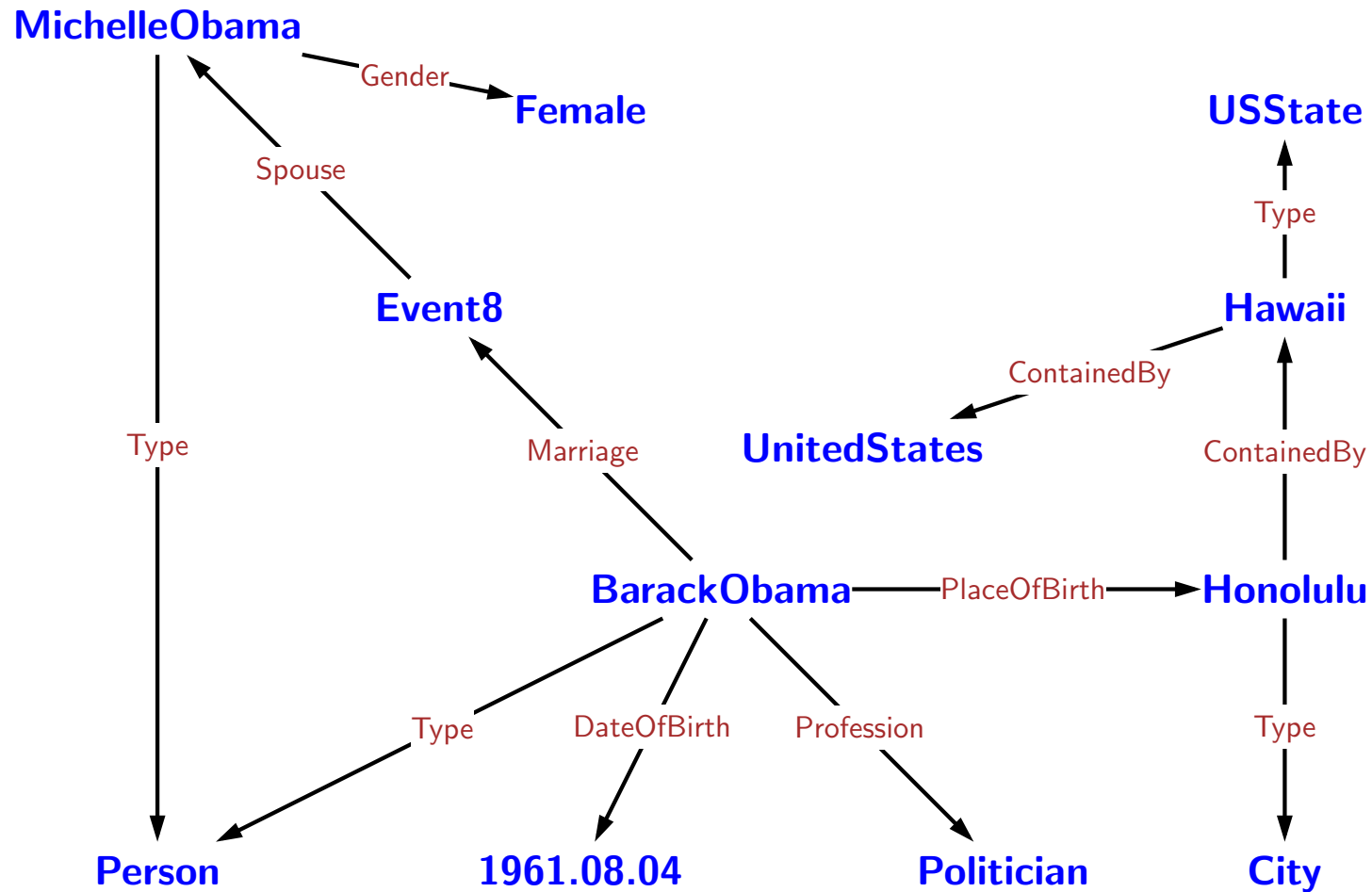
World: Freebase

100M **entities** (nodes) 1B **assertions** (edges)



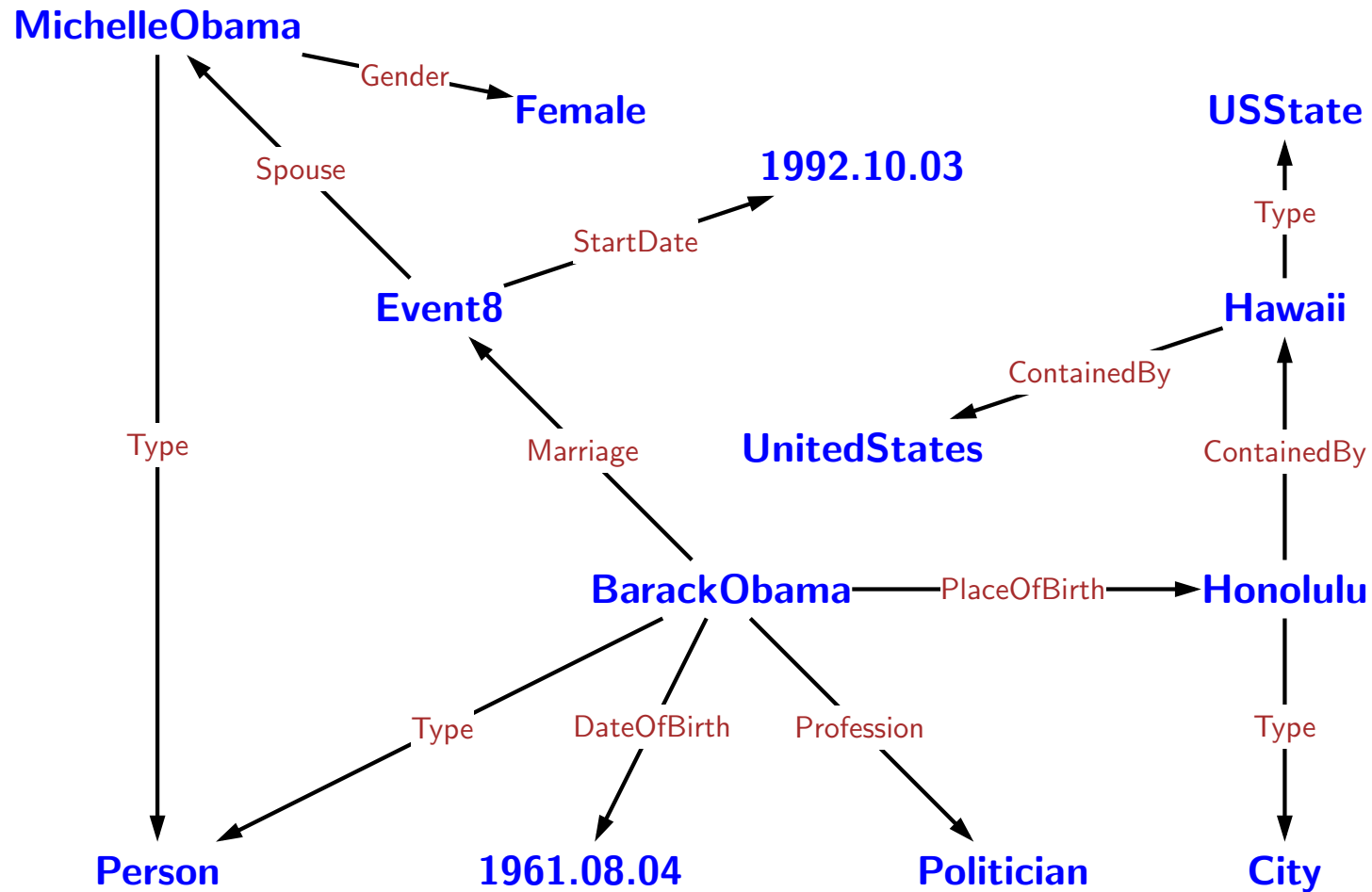
World: Freebase

100M **entities** (nodes) 1B **assertions** (edges)



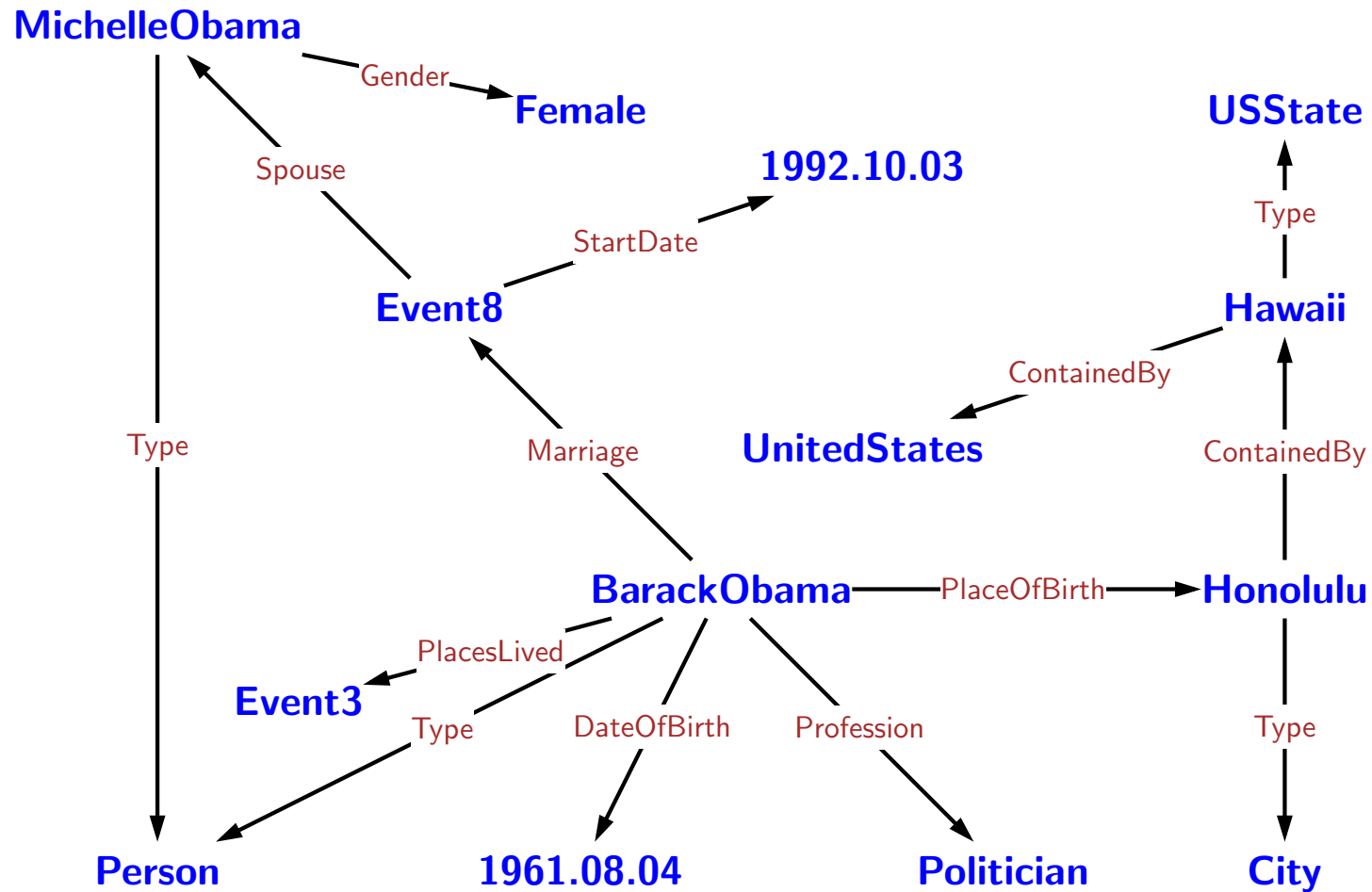
World: Freebase

100M entities (nodes) 1B assertions (edges)



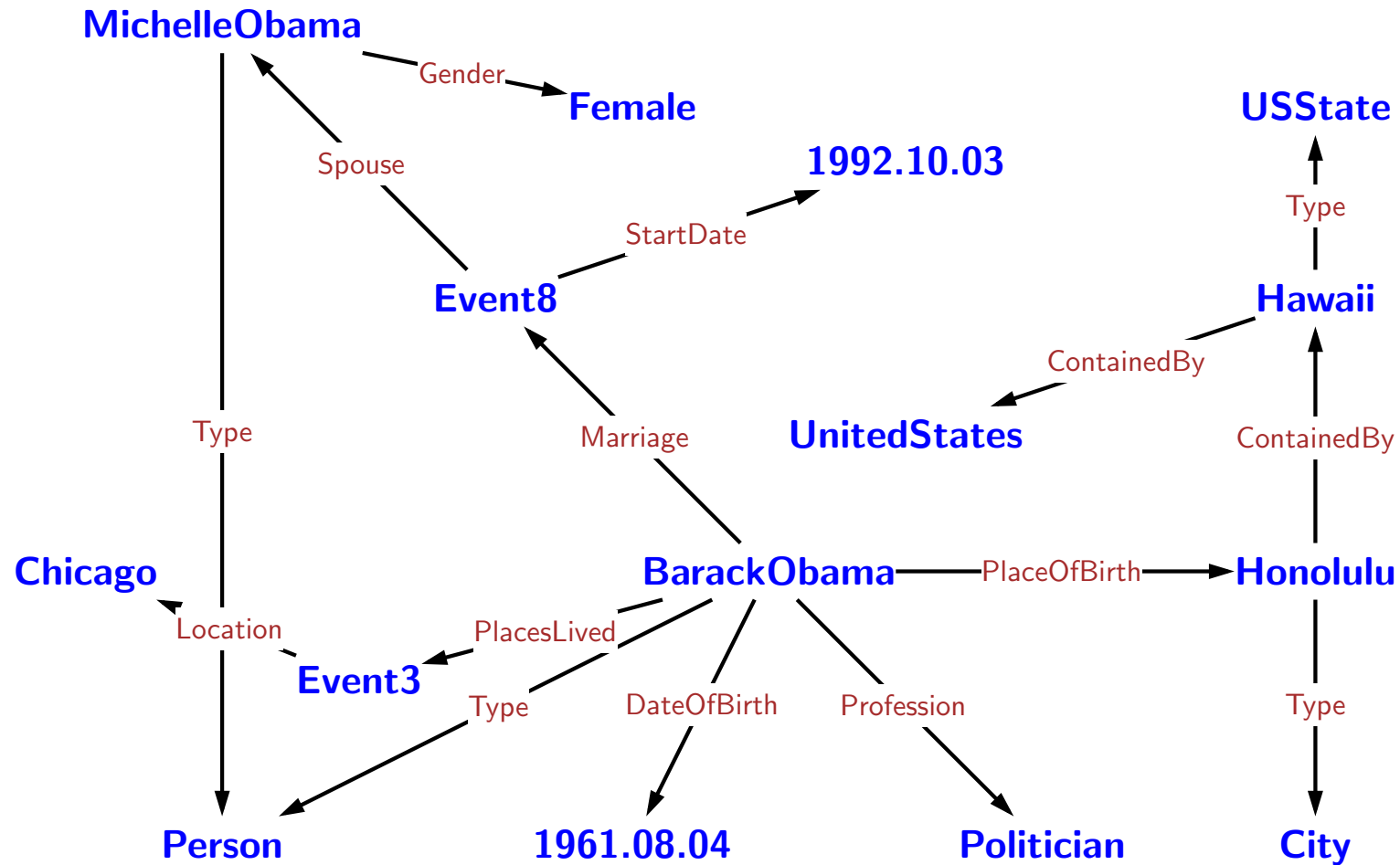
World: Freebase

100M entities (nodes) 1B assertions (edges)



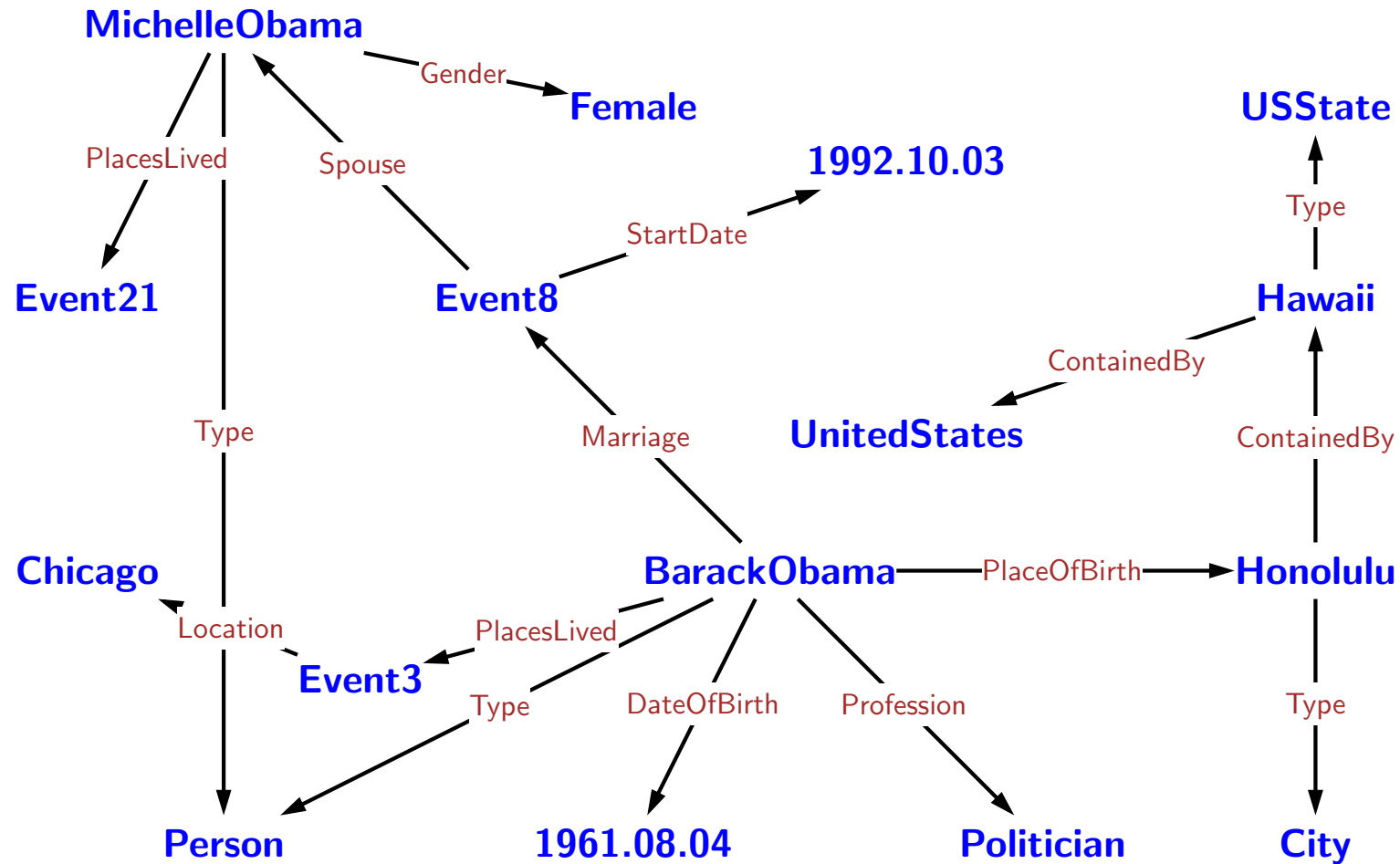
World: Freebase

100M **entities** (nodes) 1B **assertions** (edges)



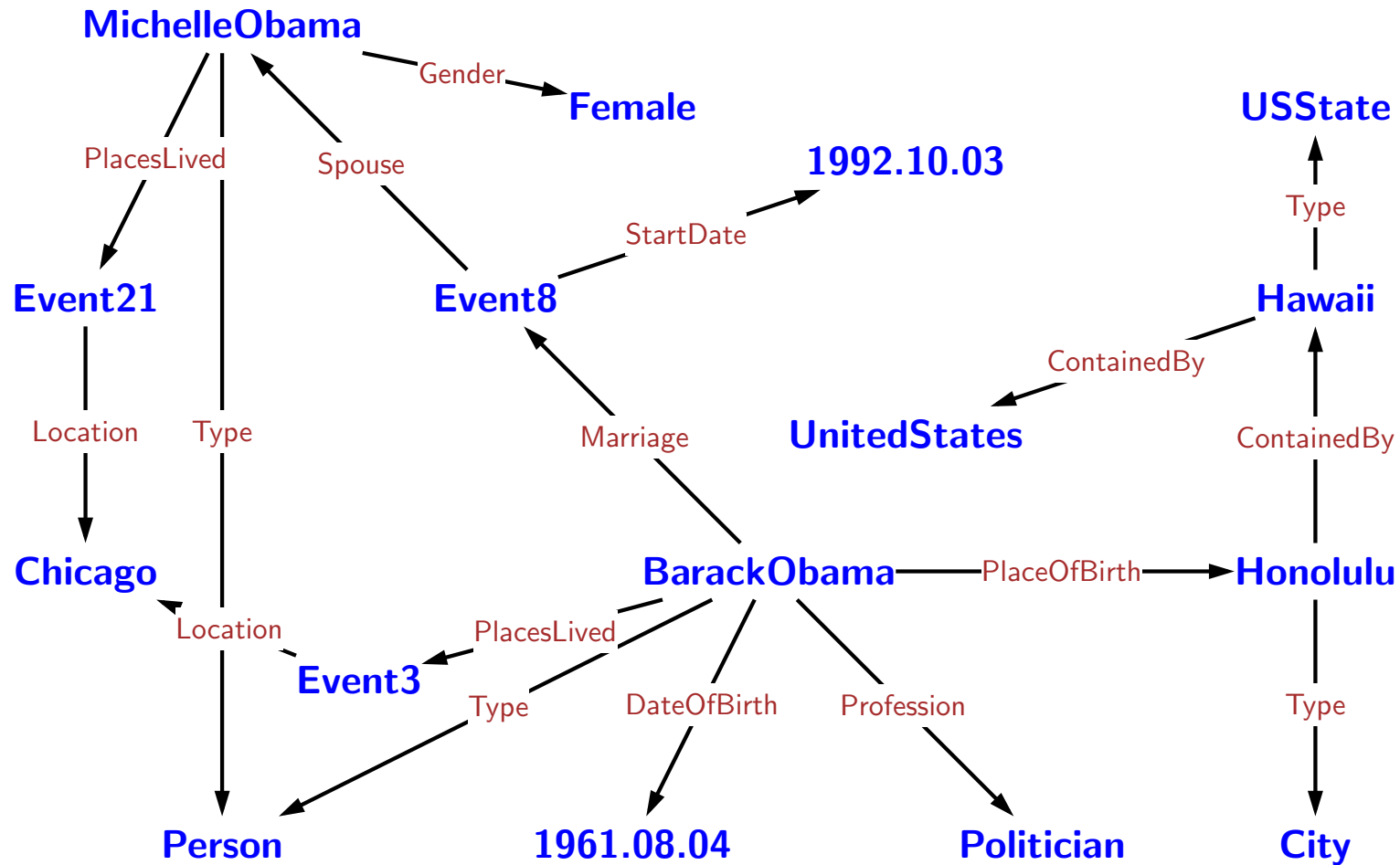
World: Freebase

100M **entities** (nodes) 1B **assertions** (edges)



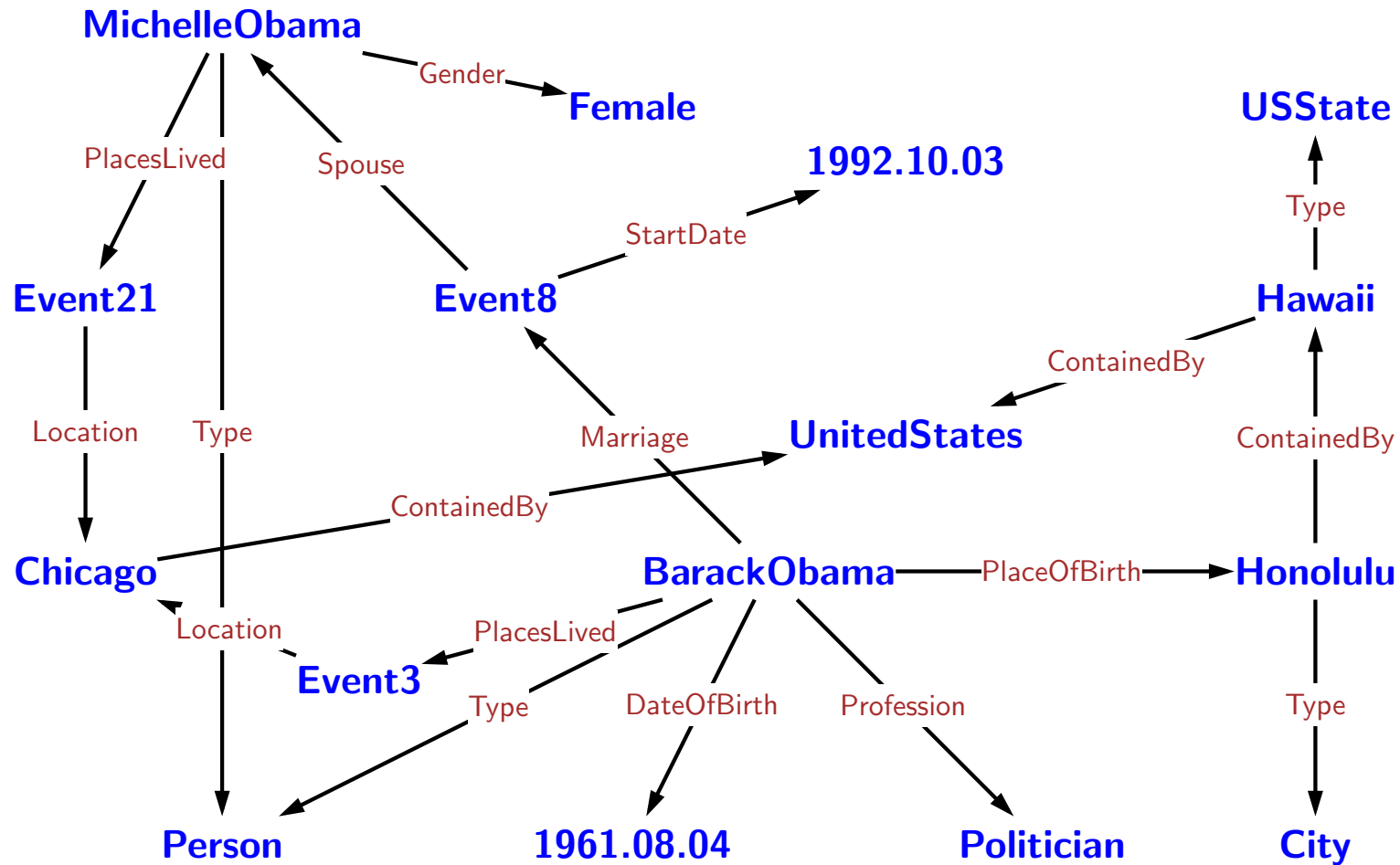
World: Freebase

100M entities (nodes) 1B assertions (edges)



World: Freebase

100M entities (nodes) 1B assertions (edges)

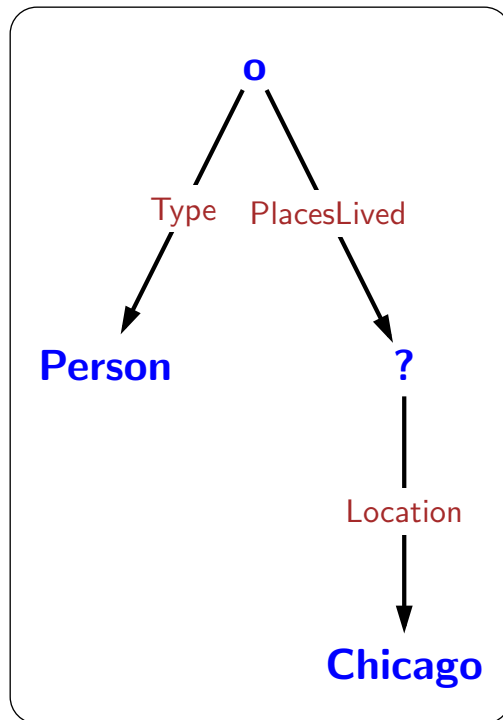


Logical forms: lambda DCS

Type.Person \sqcap PlacesLived.Location.Chicago

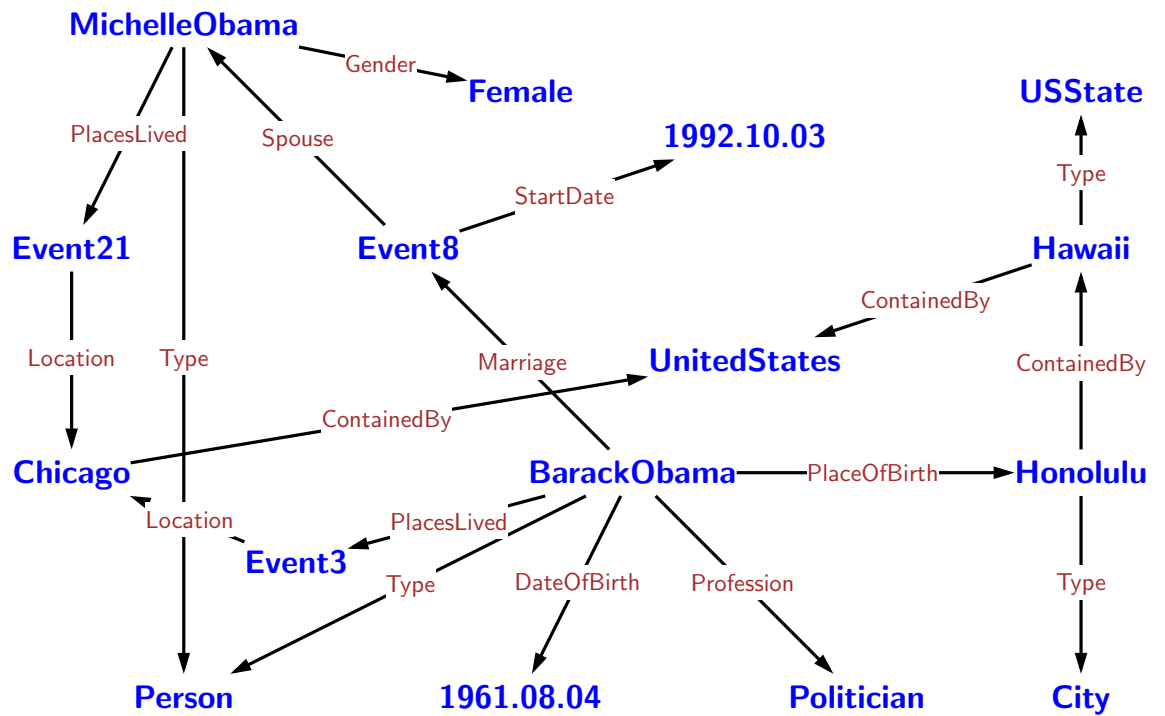
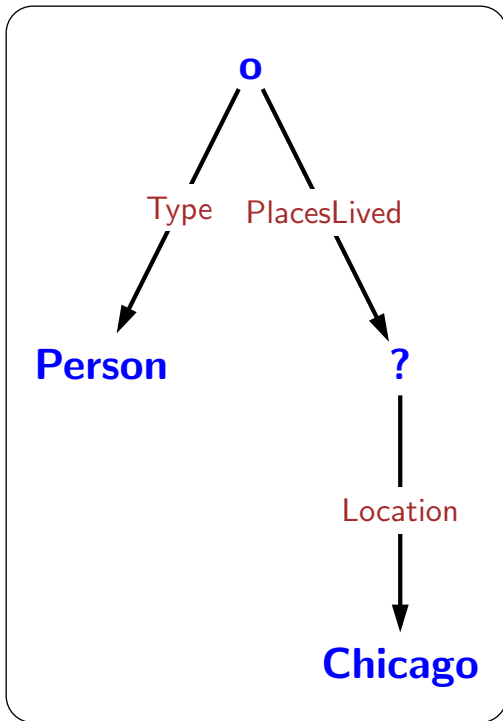
Logical forms: lambda DCS

Type.Person \sqcap PlacesLived.Location.Chicago



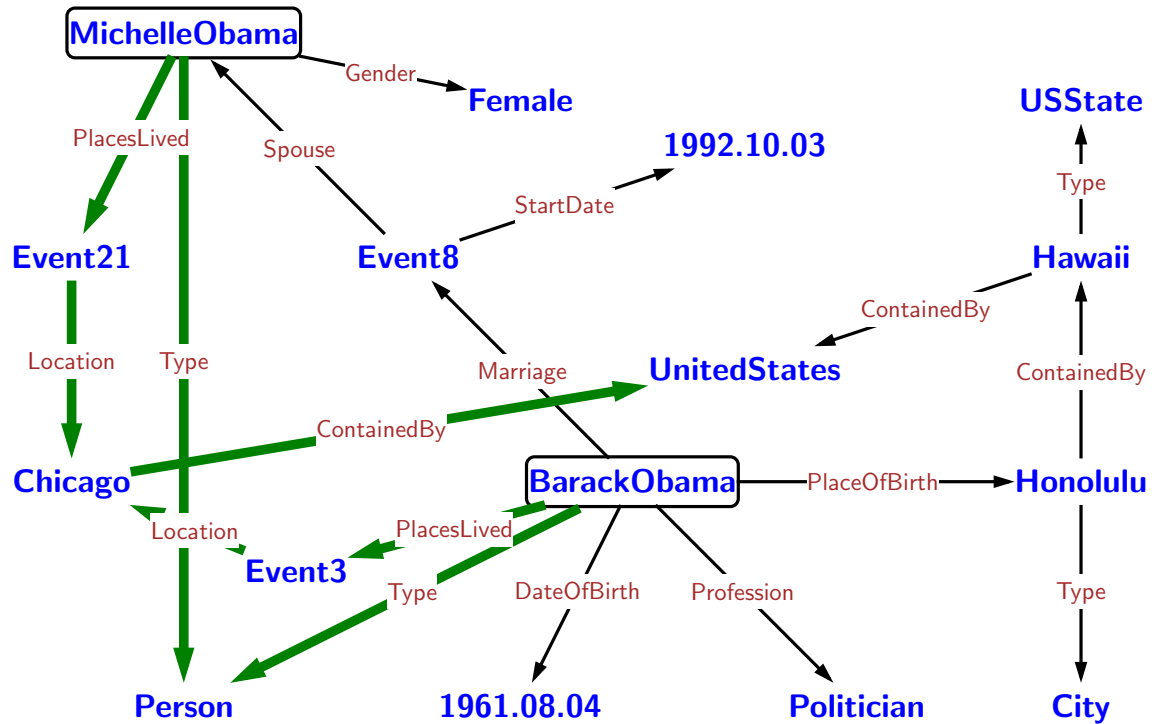
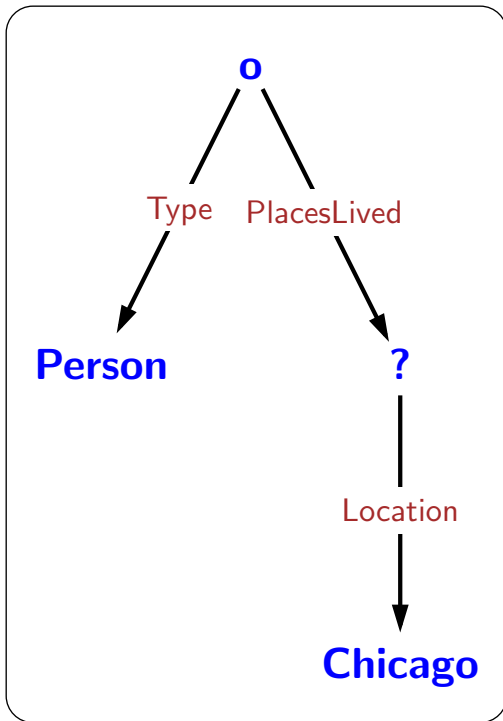
Logical forms: lambda DCS

Type.Person \sqcap PlacesLived.Location.Chicago

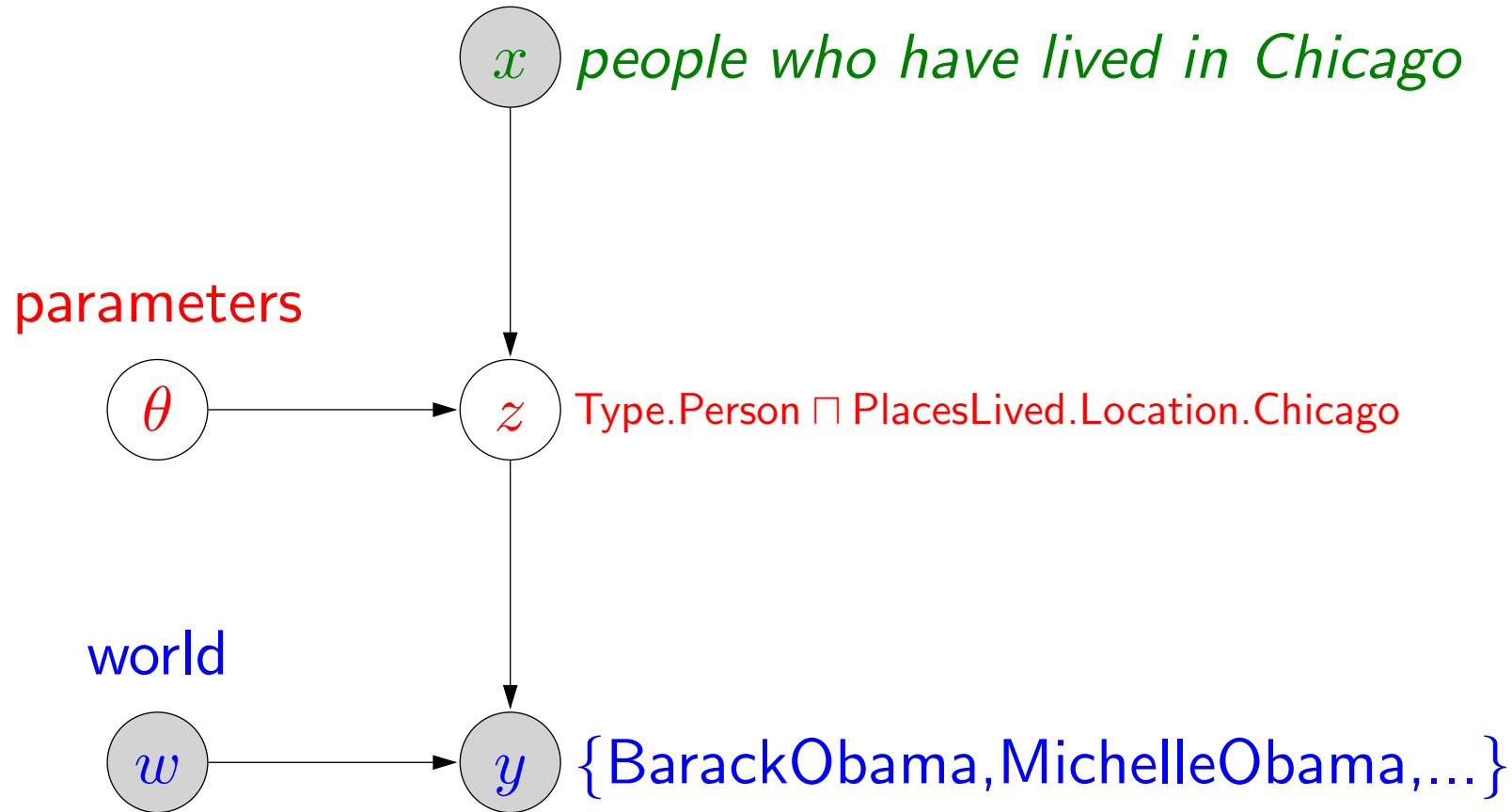


Logical forms: lambda DCS

Type.Person \sqcap PlacesLived.Location.Chicago

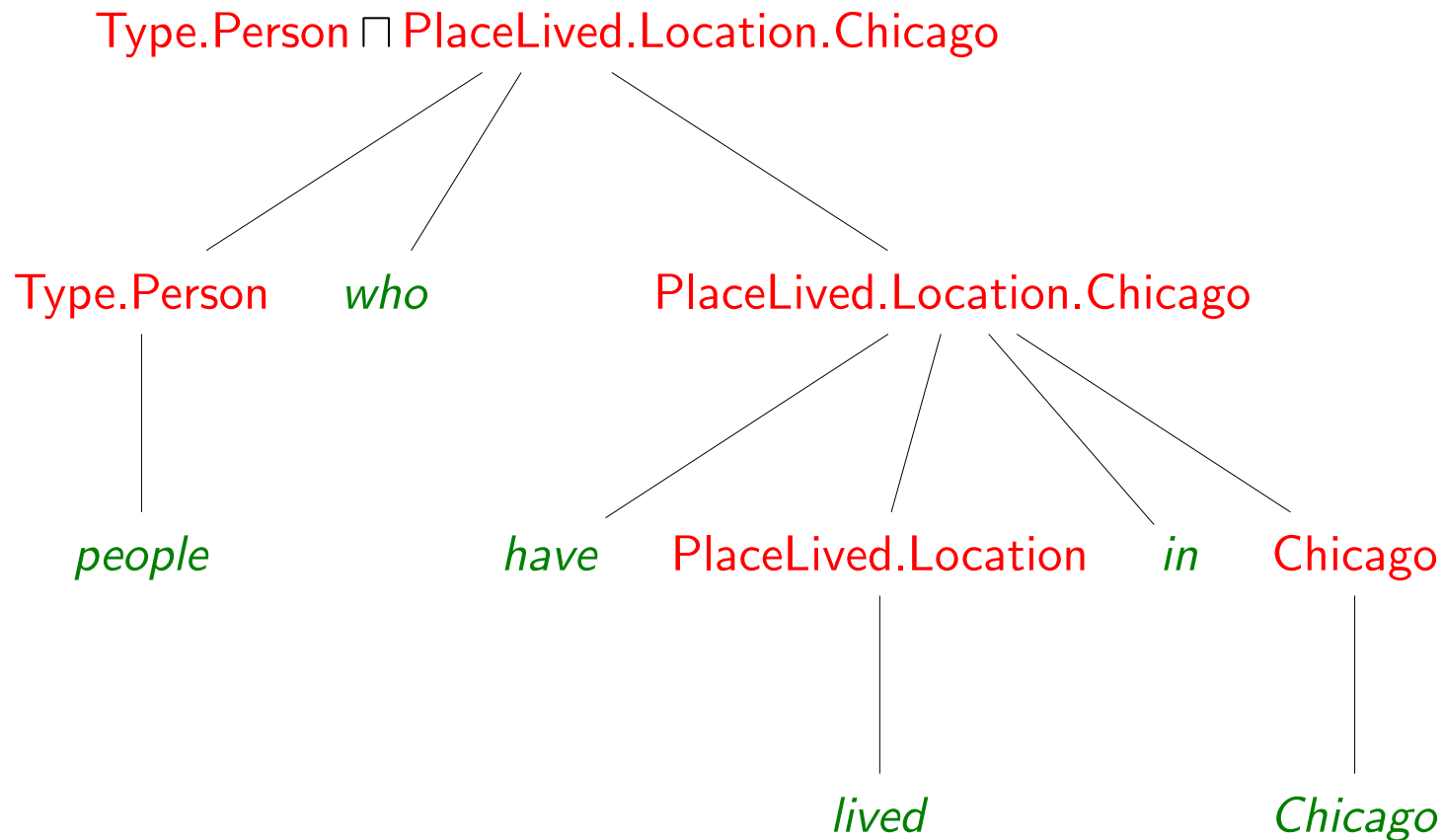


Probabilistic framework



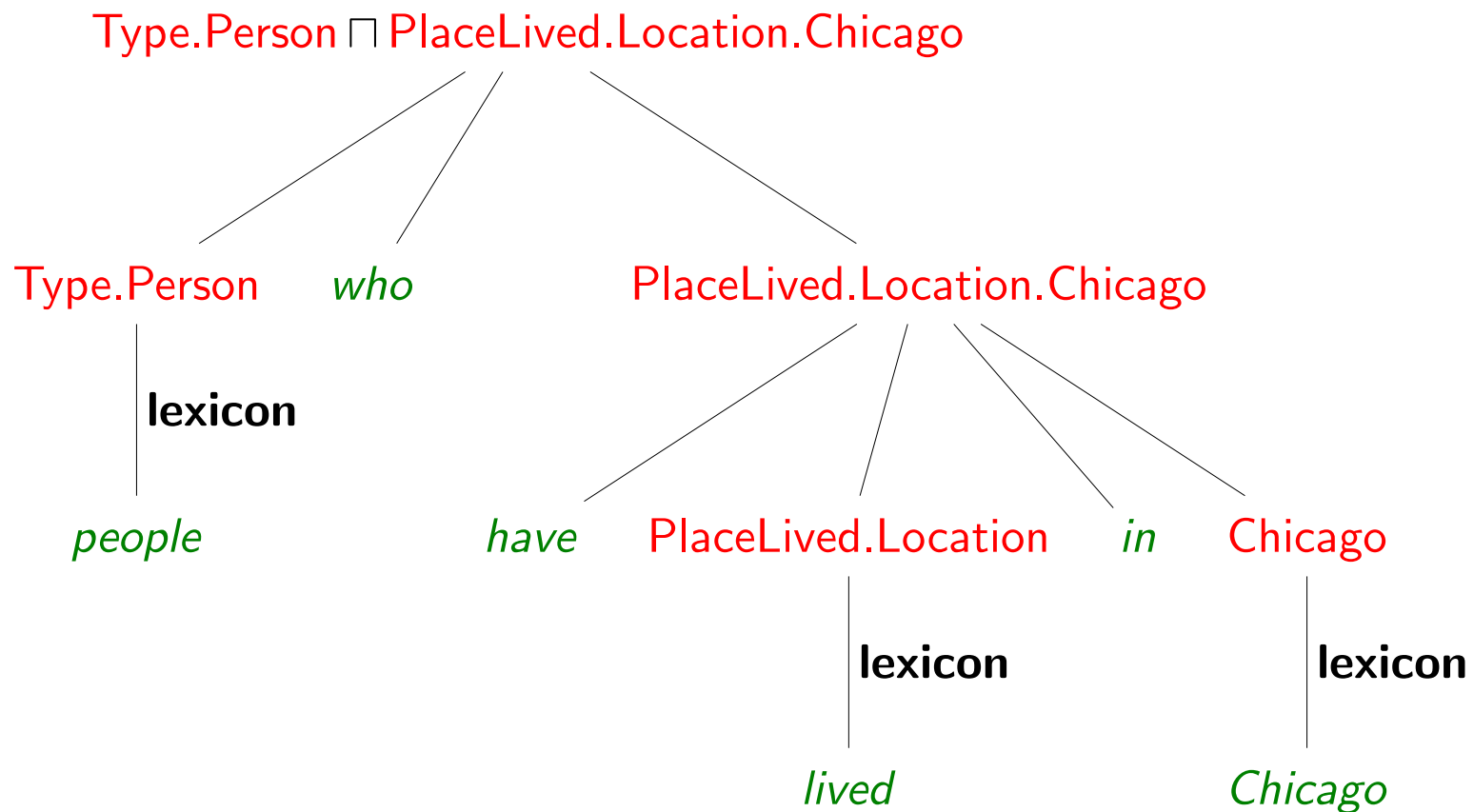
Derivations

Derivation: construction of logical form given utterance



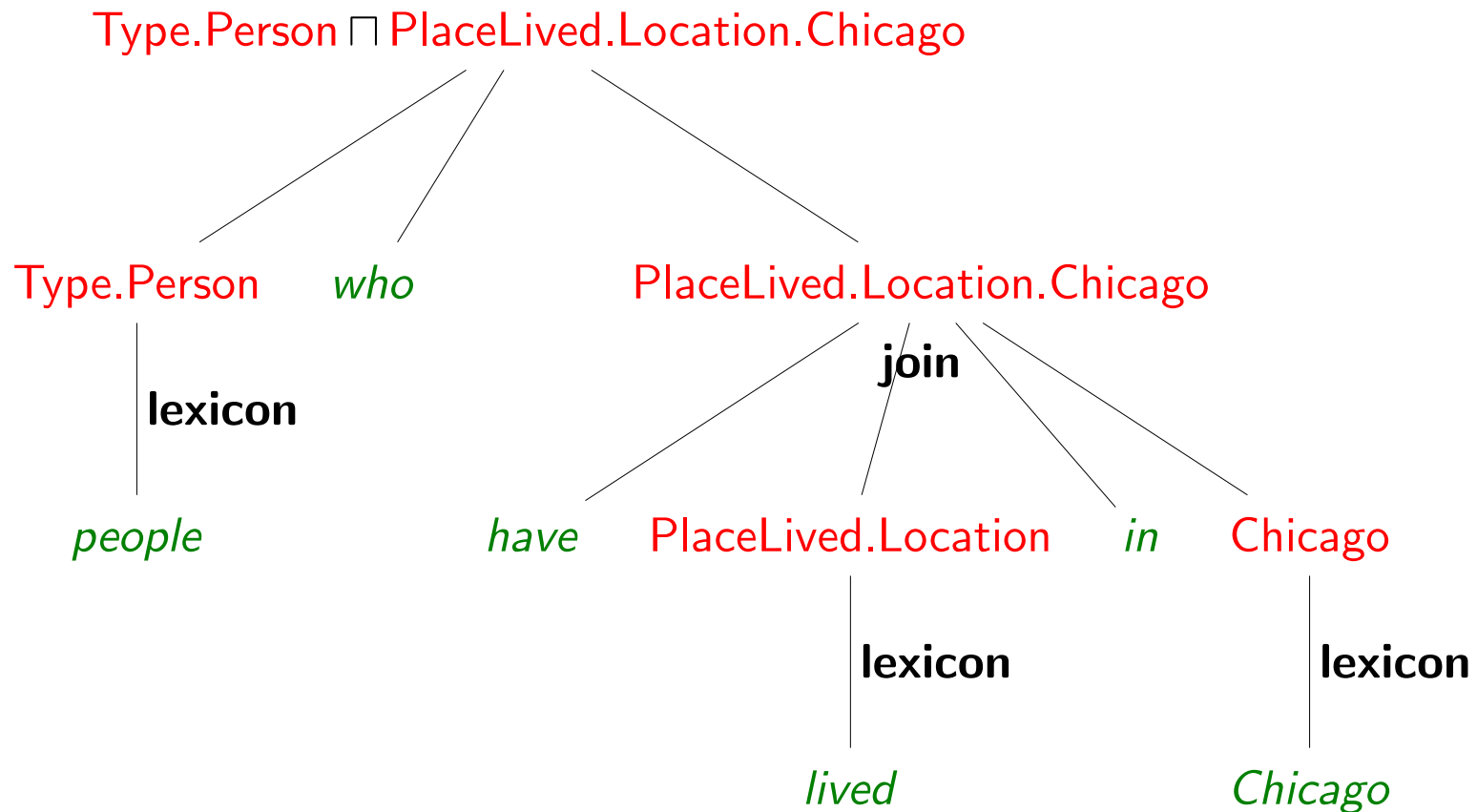
Derivations

Derivation: construction of logical form given utterance



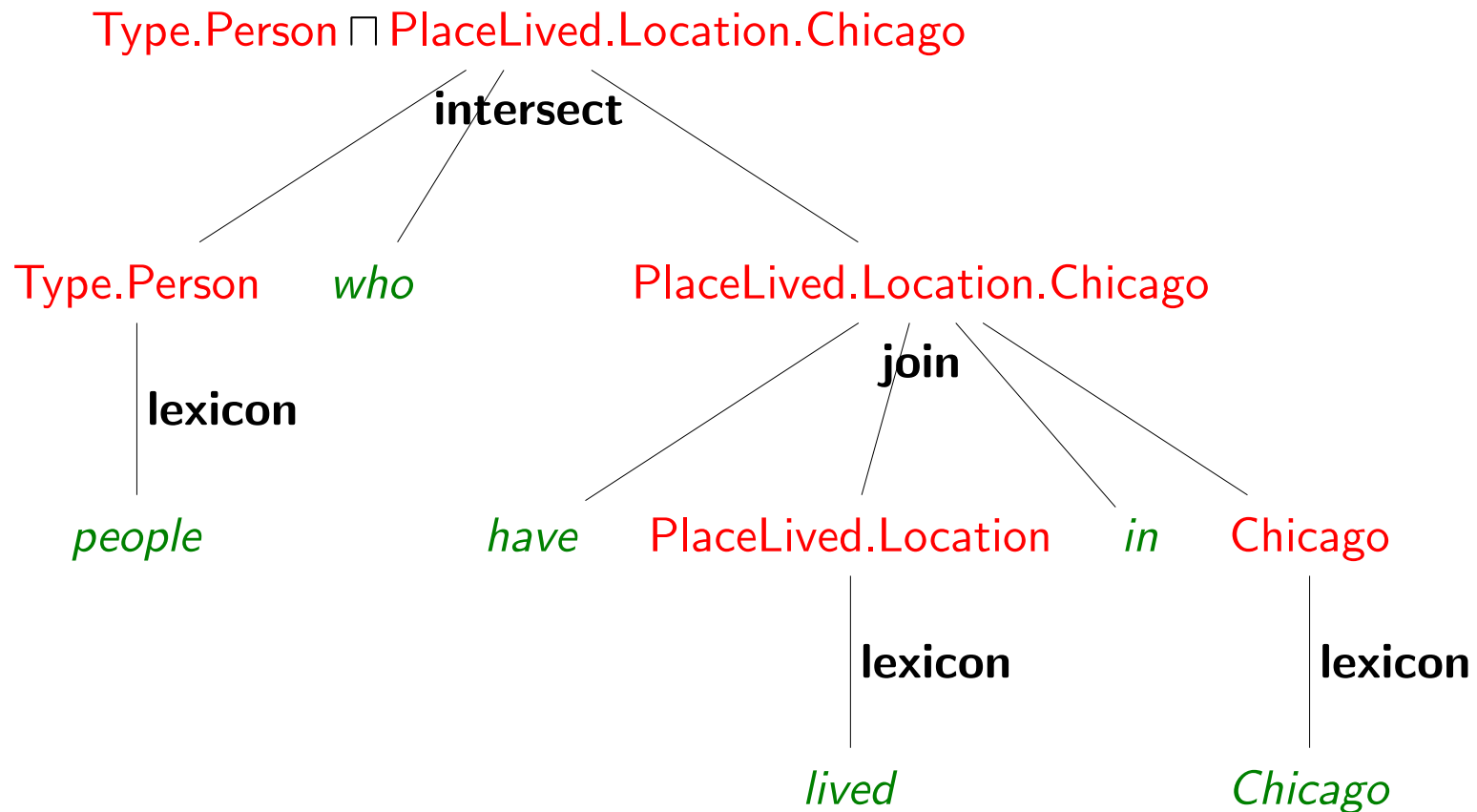
Derivations

Derivation: construction of logical form given utterance

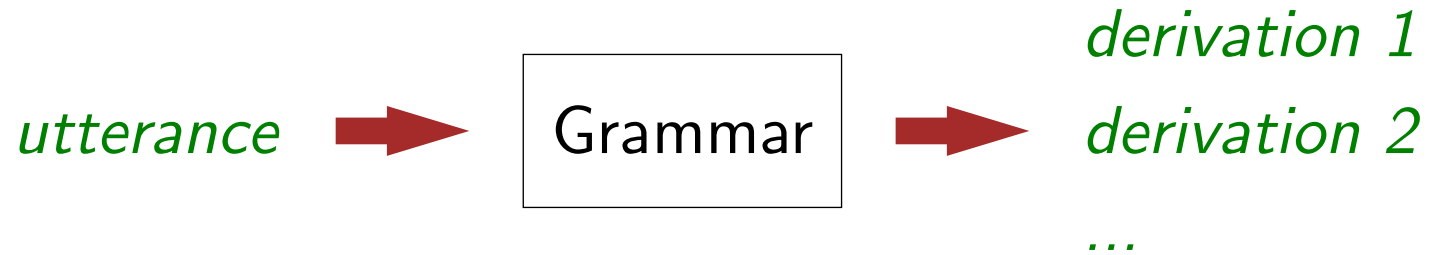


Derivations

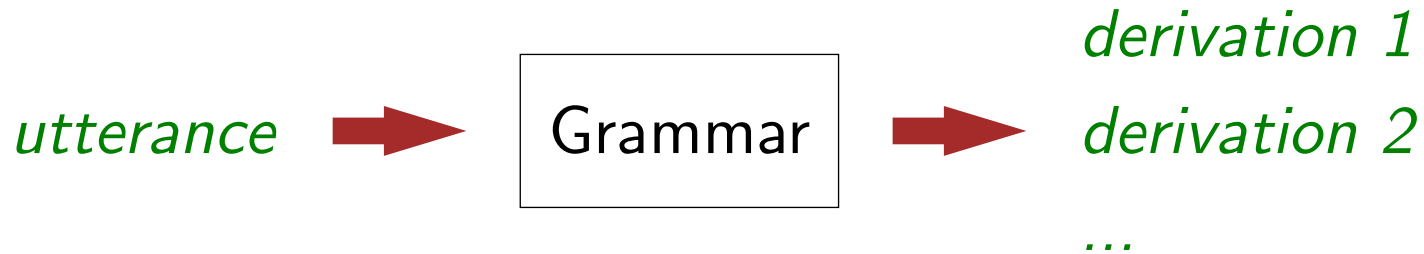
Derivation: construction of logical form given utterance



(Over)-generating derivations



(Over)-generating derivations



A Real Dumb Grammar

(lexicon) *Chicago* \Rightarrow $E : \text{Chicago}$

(lexicon) *people* \Rightarrow $E : \text{Type.Person}$

(lexicon) *live* \Rightarrow $E \times E : \text{PlacesLived}$

...

(join) $E \times E : b \quad E : u \quad \Rightarrow \quad E : b.u$

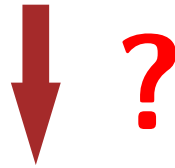
(intersect) $E : u \quad E : v \quad \Rightarrow \quad E : u \sqcap v$

Many possible derivations!

people who have lived in Chicago

Many possible derivations!

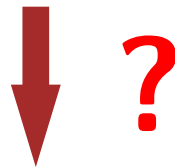
people who have lived in Chicago



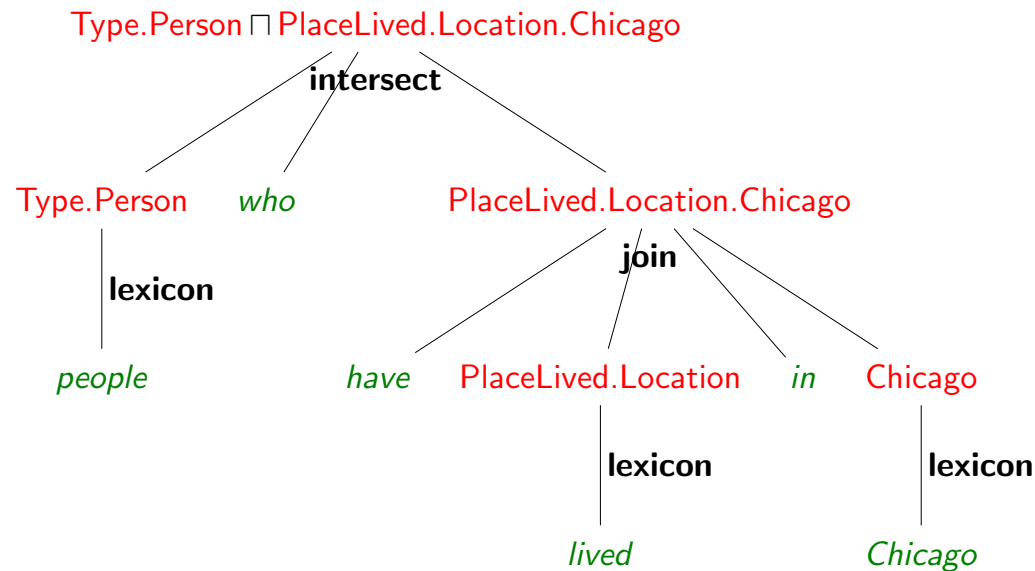
set of candidate derivations $\mathcal{D}(x)$

Many possible derivations!

people who have lived in Chicago

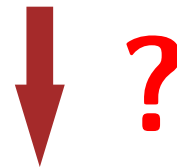


set of candidate derivations $\mathcal{D}(x)$

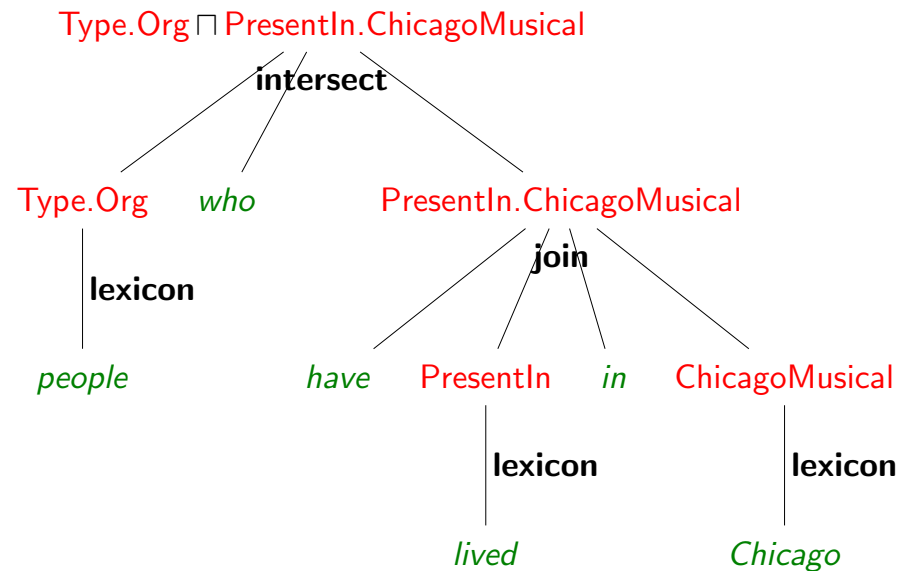


Many possible derivations!

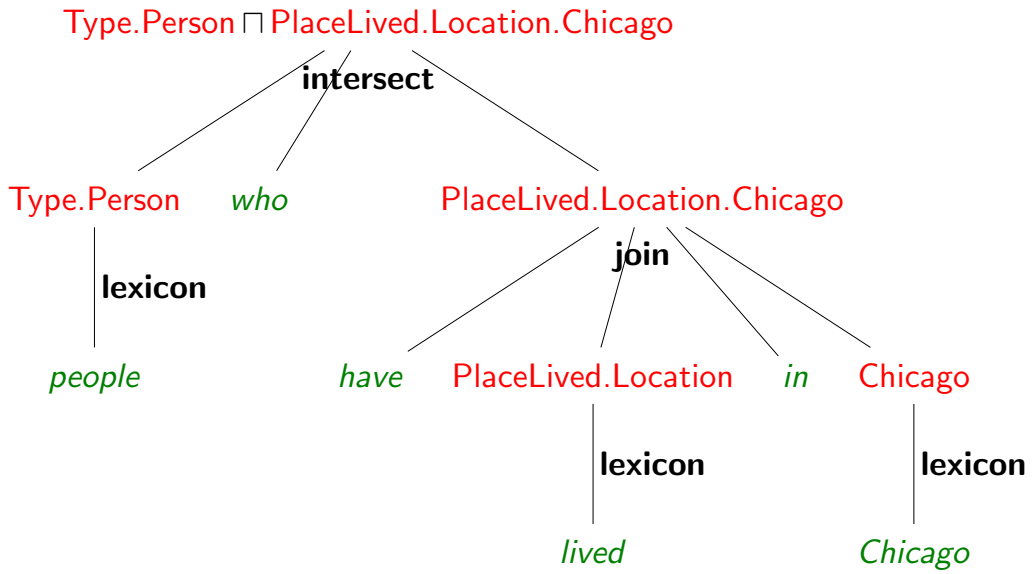
people who have lived in Chicago



set of candidate derivations $\mathcal{D}(x)$

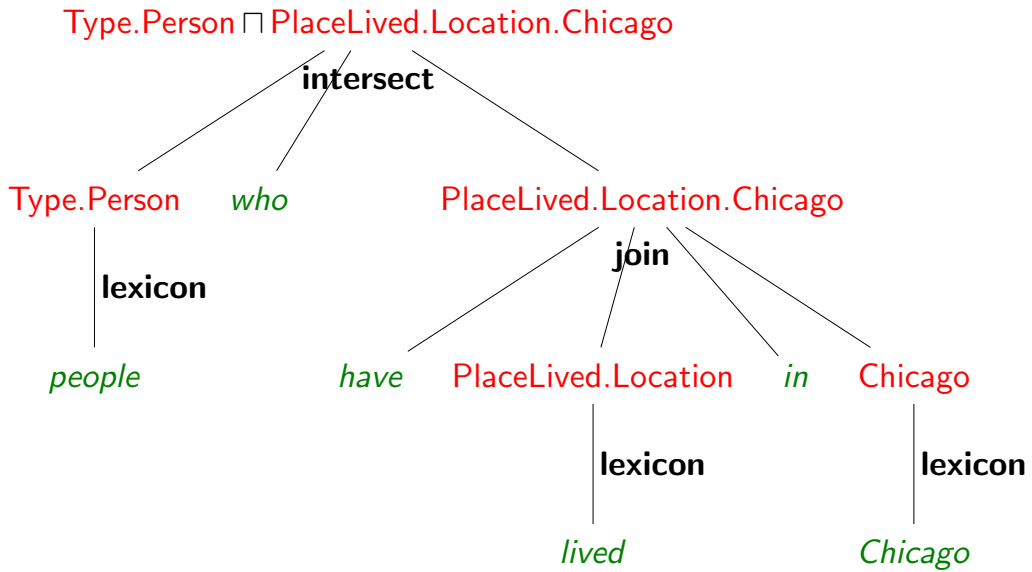


x : utterance
 d : derivation



Feature vector $\phi(x, d) \in \mathbb{R}^F$:

x : utterance
 d : derivation



Feature vector $\phi(x, d) \in \mathbb{R}^F$:

apply join	1
apply intersect	1
apply lexicon	3
skipped IN	1
skipped NN	0
<i>lived</i> maps to PlacesLived.Location	1
<i>lived</i> maps to PlaceOfBirth	0
alignmentScore	1.52
denotation-size=1	1
...	...

Scoring derivations

Feature vector: $\phi(x, d) = [1.3, 2, 0, 1, 0, 0, \dots] \in \mathbb{R}^F$

Scoring derivations

Feature vector: $\phi(x, d) = [1.3, 2, 0, 1, 0, 0, \dots] \in \mathbb{R}^F$

Parameter vector: $\theta = [1.2, -2.7, 3.4, \dots] \in \mathbb{R}^F$

Scoring derivations

Feature vector: $\phi(x, d) = [1.3, 2, 0, 1, 0, 0, \dots] \in \mathbb{R}^F$

Parameter vector: $\theta = [1.2, -2.7, 3.4, \dots] \in \mathbb{R}^F$

Scoring function:

$$\text{Score}_\theta(x, d) = \phi(x, d) \cdot \theta = \sum_{j=1}^F \theta_j \phi_j(x, d)$$

Log-linear model

Candidate derivations: $\mathcal{D}(x)$

Log-linear model

Candidate derivations: $\mathcal{D}(x)$

Model: distribution over derivations d given utterance x

$$p(d \mid x, \theta) = \frac{\exp(\text{Score}_\theta(x, d))}{\sum_{d' \in \mathcal{D}(x)} \exp(\text{Score}_\theta(x, d'))}$$

Learning

Training data:

What's Bulgaria's capital?

Sofia

What movies has Tom Cruise been in?

TopGun, VanillaSky, ...

...

+grammar, +features

Learning

Training data:

What's Bulgaria's capital?

Sofia

What movies has Tom Cruise been in?

TopGun, VanillaSky, ...

...

+grammar, +features

Objective: Maximum likelihood

$$\arg \max_{\theta} \sum_{i=1}^n \log p_{\theta}(y^{(i)} | x^{(i)})$$

Learning

Training data:

What's Bulgaria's capital?

Sofia

What movies has Tom Cruise been in?

TopGun, VanillaSky, ...

...

+grammar, +features

Objective: Maximum likelihood

$$\arg \max_{\theta} \sum_{i=1}^n \log p_{\theta}(y^{(i)} | x^{(i)})$$

Algorithm:

AdaGrad (stochastic gradient with per-feature step size)

Training intuition

Where did Mozart typress?

Vienna

Training intuition

Where did Mozart tupress?

PlaceOfBirth.Mozart

PlaceOfDeath.Mozart

PlaceOfMarriage.Mozart

Vienna

Training intuition

Where did Mozart tupress?

PlaceOfBirth.Mozart ⇒ Salzburg

PlaceOfDeath.Mozart ⇒ Vienna

PlaceOfMarriage.Mozart ⇒ Vienna

Vienna

Training intuition

Where did Mozart tupress?

~~PlaceOfBirth.Mozart → Salzburg~~

PlaceOfDeath.Mozart ⇒ Vienna

PlaceOfMarriage.Mozart ⇒ Vienna

Vienna

Training intuition

Where did Mozart tupress?

~~PlaceOfBirth.Mozart → Salzburg~~

PlaceOfDeath.Mozart ⇒ Vienna

PlaceOfMarriage.Mozart ⇒ Vienna

Vienna

Where did William Hogarth tupress?

Training intuition

Where did Mozart tupress?

~~PlaceOfBirth.Mozart → Salzburg~~

PlaceOfDeath.Mozart ⇒ Vienna

PlaceOfMarriage.Mozart ⇒ Vienna

Vienna

Where did William Hogarth tupress?

PlaceOfBirth.WilliamHogarth

PlaceOfDeath.WilliamHogarth

PlaceOfMarriage.WilliamHogarth

London

Training intuition

Where did Mozart tupress?

~~PlaceOfBirth.Mozart → Salzburg~~

PlaceOfDeath.Mozart ⇒ Vienna

PlaceOfMarriage.Mozart ⇒ Vienna

Vienna

Where did William Hogarth tupress?

PlaceOfBirth.WilliamHogarth ⇒ London

PlaceOfDeath.WilliamHogarth ⇒ London

PlaceOfMarriage.WilliamHogarth ⇒ Paddington

London

Training intuition

Where did Mozart tupress?

~~PlaceOfBirth.Mozart → Salzburg~~

PlaceOfDeath.Mozart ⇒ Vienna

PlaceOfMarriage.Mozart ⇒ Vienna

Vienna

Where did William Hogarth tupress?

PlaceOfBirth.WilliamHogarth ⇒ London

PlaceOfDeath.WilliamHogarth ⇒ London

~~PlaceOfMarriage.WilliamHogarth → Paddington~~

London

Training intuition

Where did Mozart tupress?

~~PlaceOfBirth.Mozart → Salzburg~~

PlaceOfDeath.Mozart ⇒ Vienna

PlaceOfMarriage.Mozart ⇒ Vienna

Vienna

Where did William Hogarth tupress?

PlaceOfBirth.WilliamHogarth ⇒ London

PlaceOfDeath.WilliamHogarth ⇒ London

~~PlaceOfMarriage.WilliamHogarth → Paddington~~

London

Outline

- A semantic parsing framework
- **A closer look at the elements**
 - Logical forms: lambda DCS
 - Lexical coverage
 - Grammar: building logical forms
 - Learning via bootstrapping
 - Leveraging denotations
 - Datasets/results
- Beyond Freebase
- Final remarks

Outline

- A semantic parsing framework
- A closer look at the elements
 - **Logical forms: lambda DCS**
 - Lexical coverage
 - Grammar: building logical forms
 - Learning via bootstrapping
 - Leveraging denotations
 - Datasets/results
- Beyond Freebase
- Final remarks

Lambda DCS

Entity

Chicago

Lambda DCS

Entity

Chicago

Join

PlaceOfBirth.Chicago

Lambda DCS

Entity

Chicago

Join

PlaceOfBirth.Chicago

Intersect

Type.Person \cap PlaceOfBirth.Chicago

Lambda DCS

Entity

Chicago

Join

PlaceOfBirth.Chicago

Intersect

Type.Person \cap PlaceOfBirth.Chicago

Aggregation

count(Type.Person \cap PlaceOfBirth.Chicago)

Lambda DCS

Entity

Chicago

Join

PlaceOfBirth.Chicago

Intersect

Type.Person \sqcap PlaceOfBirth.Chicago

Aggregation

count(Type.Person \sqcap PlaceOfBirth.Chicago)

Superlative

argmin(Type.Person \sqcap PlaceOfBirth.Chicago, DateOfBirth)

Lambda DCS

Entity

Chicago

Join

PlaceOfBirth.Chicago

Intersect

Type.Person \sqcap PlaceOfBirth.Chicago

Aggregation

count(Type.Person \sqcap PlaceOfBirth.Chicago)

Superlative

argmin(Type.Person \sqcap PlaceOfBirth.Chicago, DateOfBirth)

Anaphora

μx . Type.Person \sqcap Children.Influence. x

Lambda DCS

Entity

Chicago

Join

PlaceOfBirth.Chicago

Intersect

Type.Person \sqcap PlaceOfBirth.Chicago

Aggregation

count(Type.Person \sqcap PlaceOfBirth.Chicago)

Superlative

argmin(Type.Person \sqcap PlaceOfBirth.Chicago, DateOfBirth)

Anaphora

μx . Type.Person \sqcap Children.Influence. x

Variable

λx . Gender.Female \sqcap Children.Children. x

Comparison to lambda calculus

Lambda calculus

$\lambda x. \text{Type}(x, \text{Person}) \wedge \exists e. \text{PlacesLived}(x, e) \wedge \text{Location}(e, \text{Chicago})$

Comparison to lambda calculus

Lambda calculus

$\lambda x. \text{Type}(x, \text{Person}) \wedge \exists e. \text{PlacesLived}(x, e) \wedge \text{Location}(e, \text{Chicago})$

Lambda dependency-based compositional semantics (DCS)

$\text{Type. Person} \sqcap \text{PlacesLived. Location. Chicago}$

Comparison to lambda calculus

Lambda calculus

$\lambda x. \text{Type}(x, \text{Person}) \wedge \exists e. \text{PlacesLived}(x, e) \wedge \text{Location}(e, \text{Chicago})$

Lambda dependency-based compositional semantics (DCS)

$\text{Type. Person} \sqcap \text{PlacesLived. Location. Chicago}$

- Eliminate variables to simplify

Comparison to lambda calculus

Lambda calculus

$\lambda x. \text{Type}(x, \text{Person}) \wedge \exists e. \text{PlacesLived}(x, e) \wedge \text{Location}(e, \text{Chicago})$

Lambda dependency-based compositional semantics (DCS)

$\text{Type. Person} \sqcap \text{PlacesLived. Location. Chicago}$

- Eliminate variables to simplify
- Existential quantification by default

Comparison to lambda calculus

Lambda calculus

$\lambda x. \text{Type}(x, \text{Person}) \wedge \exists e. \text{PlacesLived}(x, e) \wedge \text{Location}(e, \text{Chicago})$

Lambda dependency-based compositional semantics (DCS)

$\text{Type. Person} \sqcap \text{PlacesLived. Location. Chicago}$

- Eliminate variables to simplify
- Existential quantification by default
- Replace entity sets ($e \rightarrow t$) rather than truth values (t)

Comparison to lambda calculus

Lambda calculus

$\lambda x. \text{Type}(x, \text{Person}) \wedge \exists e. \text{PlacesLived}(x, e) \wedge \text{Location}(e, \text{Chicago})$

Lambda dependency-based compositional semantics (DCS)

$\text{Type. Person} \sqcap \text{PlacesLived. Location. Chicago}$

- Eliminate variables to simplify
- Existential quantification by default
- Replace entity sets ($e \rightarrow t$) rather than truth values (t)
- Use tree/graph structures (analogy to dependency syntax)

Comparison to lambda calculus

Lambda calculus

$\lambda x. \text{Type}(x, \text{Person}) \wedge \exists e. \text{PlacesLived}(x, e) \wedge \text{Location}(e, \text{Chicago})$

Lambda dependency-based compositional semantics (DCS)

$\text{Type. Person} \sqcap \text{PlacesLived. Location. Chicago}$

- Eliminate variables to simplify
- Existential quantification by default
- Replace entity sets ($e \rightarrow t$) rather than truth values (t)
- Use tree/graph structures (analogy to dependency syntax)
- Inspired by Discourse Representation Theory, description logic, modal logic, similarities to AMR

An example

states west of Texas whose capitals are their largest cities

Outline

- A semantic parsing framework
- A closer look at the elements
 - Logical forms: lambda DCS
 - **Lexical coverage**
 - Grammar: building logical forms
 - Learning via bootstrapping
 - Leveraging denotations
 - Datasets/results
- Beyond Freebase
- Final remarks

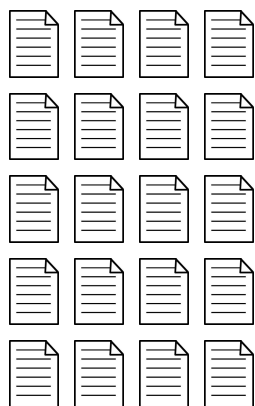
Challenge: lexical coverage

born ⇒ Type.City, PlaceOfBirth, Profession.Lawyer, ...

?

Solution: alignment

Open information extraction on ClueWeb09:



(Barack Obama, was born in, Honolulu)

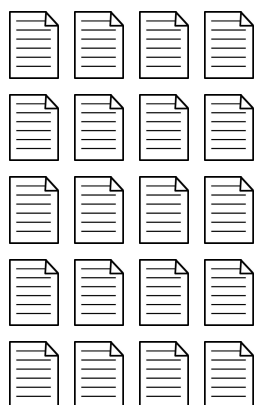
(Albert Einstein, was born in, Ulm)

(Barack Obama, lived in, Chicago)

... 15M triples ...

Solution: alignment

Open information extraction on ClueWeb09:



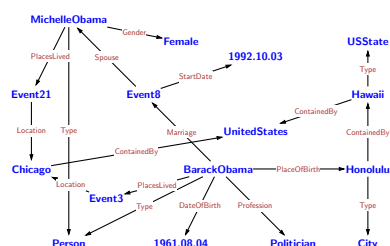
(Barack Obama, was born in, Honolulu)

(Albert Einstein, was born in, Ulm)

(Barack Obama, lived in, Chicago)

... 15M triples ...

Freebase:



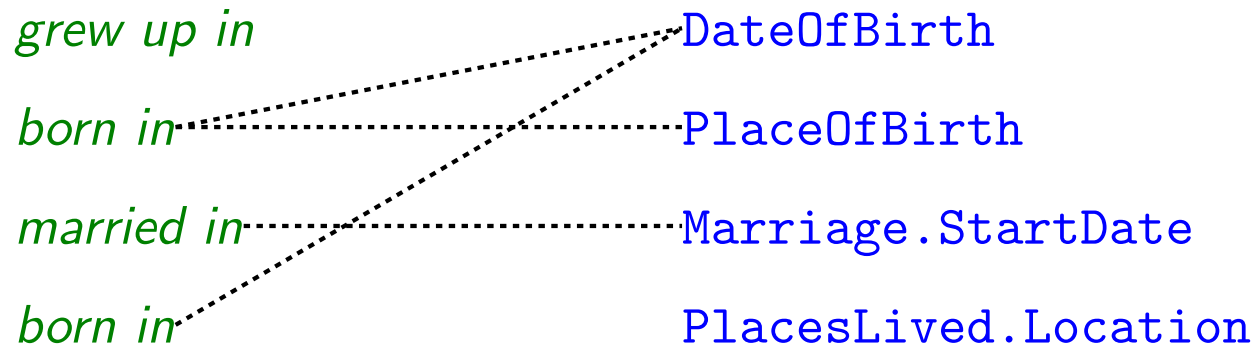
(BarackObama, PlaceOfBirth, Honolulu)

(Albert Einstein, PlaceOfBirth, Ulm)

(BarackObama, PlacesLived.Location, Chicago)

... 400M triples ...

Match text and Freebase predicates



Similar schema matching / alignment ideas [Cai & Yates, 2013, Fader et. al, 2013, Yao & van Durme, 2014; etc.]

Challenge: variability in language

What is the currency in the US?

Challenge: variability in language

What is the currency in the US?

What money do they use in the states?

How do you pay in America?

What's the currency of the US?

What money is accepted in the United States?

What money to take to the US?

...

A solution: paraphrasing

How many people live in Seattle?

paraphrase

What is the population of Seattle?

PopulationOf(Seattle)

850,000

Convert to a text-only problem

Challenge: "sub-lexical compositionality"

grandmother

$\lambda x. \text{Gender.Female} \sqcap \text{Children.Children}.x$

mayor

$\lambda x. \text{GovtPositionsHeld}.(\text{Title.Mayor} \sqcap \text{OfficeOfJurisdiction}.x)$

Challenge: "sub-lexical compositionality"

grandmother

$\lambda x. \text{Gender.Female} \sqcap \text{Children.Children}.x$

mayor

$\lambda x. \text{GovtPositionsHeld}.(\text{Title.Mayor} \sqcap \text{OfficeOfJurisdiction}.x)$

This is "grounding".

Challenge: "sub-lexical compositionality"

*presidents who have served two **non-consecutive** terms*

[requires higher-order quantification]

*presidents who were **previously** vice-presidents*

[anaphora]

***every other** president*

[weird quantification]

Outline

- A semantic parsing framework
- A closer look at the elements
 - Logical forms: lambda DCS
 - Lexical coverage
 - **Grammar: building logical forms**
 - Learning via bootstrapping
 - Leveraging denotations
 - Datasets/results
- Beyond Freebase
- Final remarks

Over-simplifying the grammar

A Real Dumb Grammar

(lexicon) *Chicago* \Rightarrow E : **Chicago**

(lexicon) *people* \Rightarrow E : **Type.Person**

(lexicon) *live* \Rightarrow E \times E : **PlacesLived**

...

(join) E \times E : *b* E : *u* \Rightarrow E : *b.u*

(intersect) E : *u* E : *v* \Rightarrow E : *u* \sqcap *v*

Over-simplifying the grammar

A Real Dumb Grammar

(lexicon) *Chicago* \Rightarrow E : **Chicago**

(lexicon) *people* \Rightarrow E : **Type.Person**

(lexicon) *live* \Rightarrow E \times E : **PlacesLived**

...

(join) E \times E : *b* E : *u* \Rightarrow E : *b.u*

(intersect) E : *u* E : *v* \Rightarrow E : *u* \sqcap *v*

A CCG Grammar

(lexicon) *Chicago* \Rightarrow NP : **PlaceOfBirth**

(lexicon) *people* \Rightarrow N : $\lambda x.$ **Type(x, Person)**

(lexicon) *live* \Rightarrow (N\N)/NP : $\lambda y.$ $\lambda f.$ $\lambda x.f(x) \wedge$ **PlacesLived(x, y)**

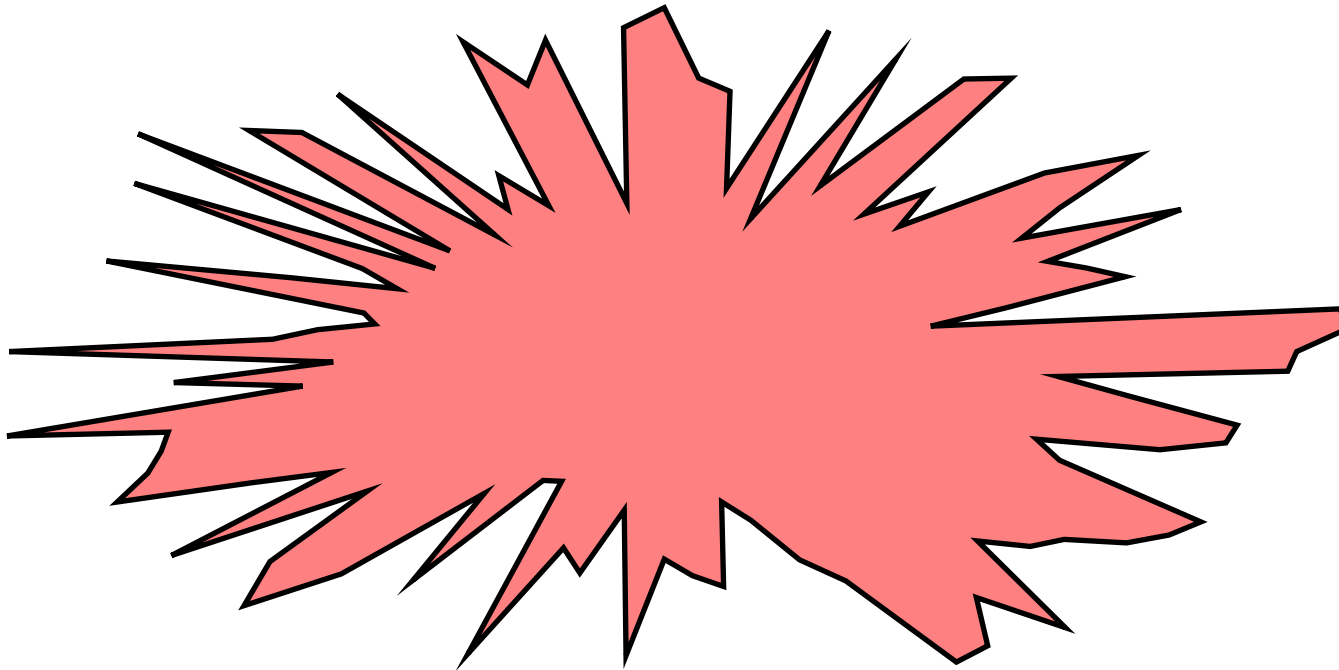
...

(backward) X : *x* Y\X : *f* \Rightarrow Y : *f(x)*

(forward) Y/X : *b* X : *u* \Rightarrow Y : *f(x)*

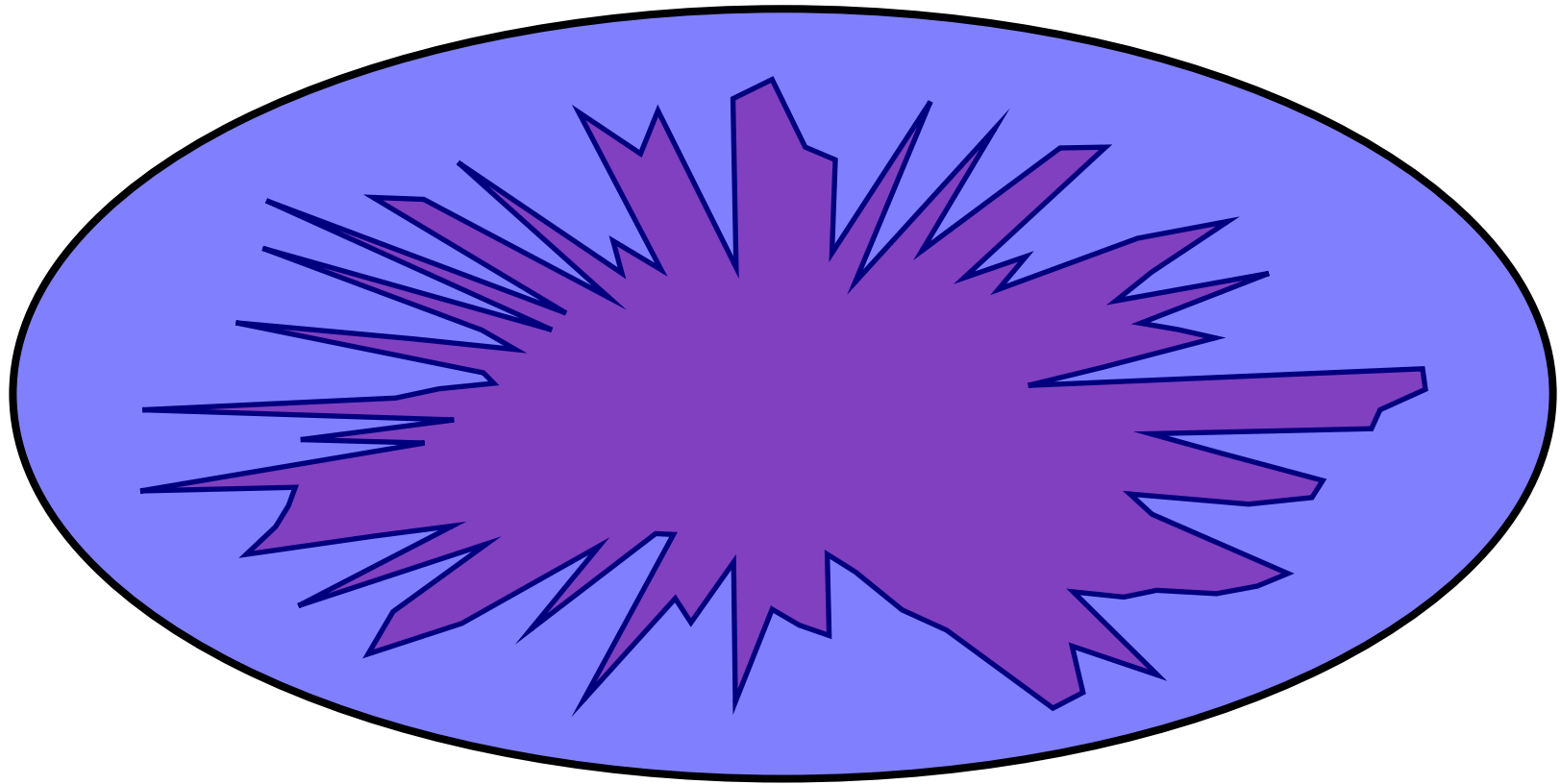
Overapproximation via simple grammars

- Modeling correct derivations requires complex rules



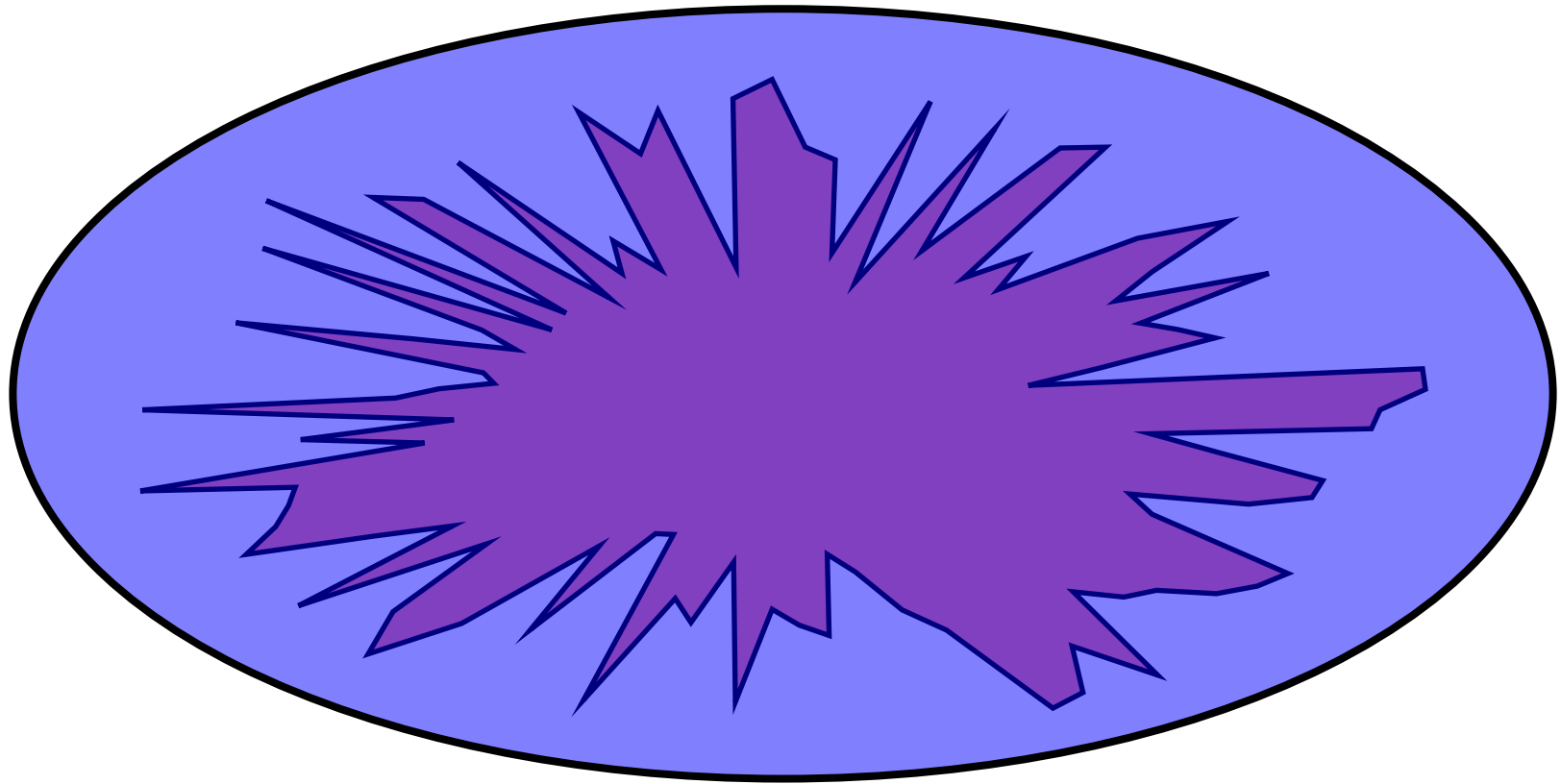
Overapproximation via simple grammars

- Modeling correct derivations requires complex rules
- Simple rules generate overapproximation of good derivations



Overapproximation via simple grammars

- Modeling correct derivations requires complex rules
- Simple rules generate overapproximation of good derivations



- Hard grammar rules \Rightarrow soft/overlapping features

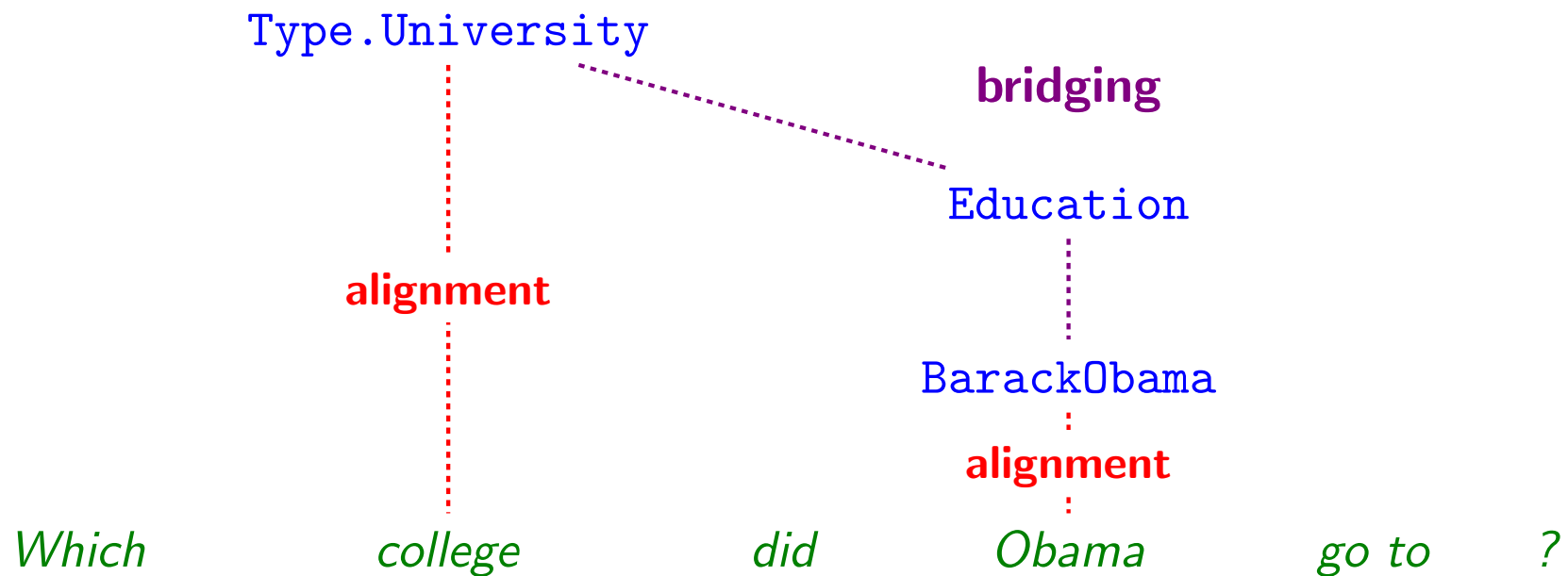
Bridging

Type.University \sqcap Education.BarackObama



Bridging

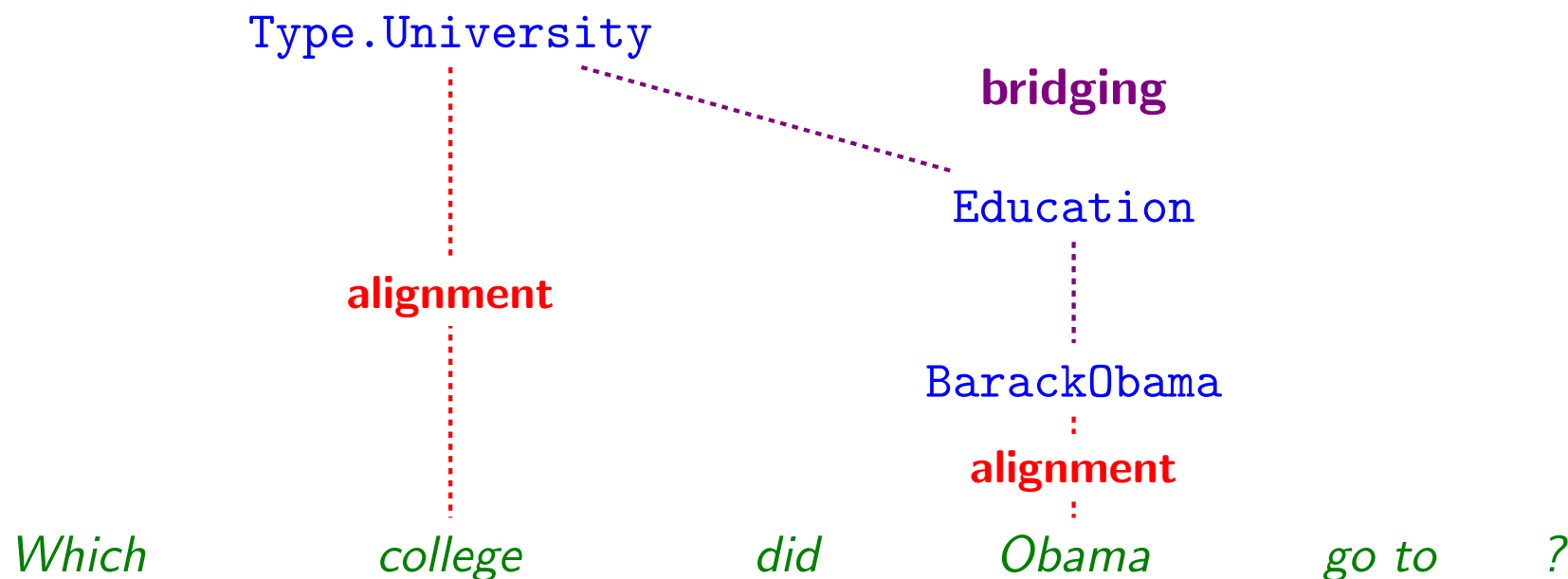
Type.University \sqcap Education.BarackObama



Bridging: use neighboring predicates / type constraints

Bridging

Type.University \sqcap Education.BarackObama



Bridging: use neighboring predicates / type constraints

Start building from parts with more certainty

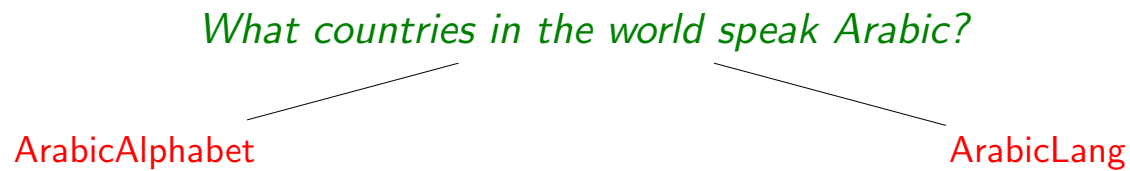
Bridging to nowhere

Search logical forms based on "prior":

What countries in the world speak Arabic?

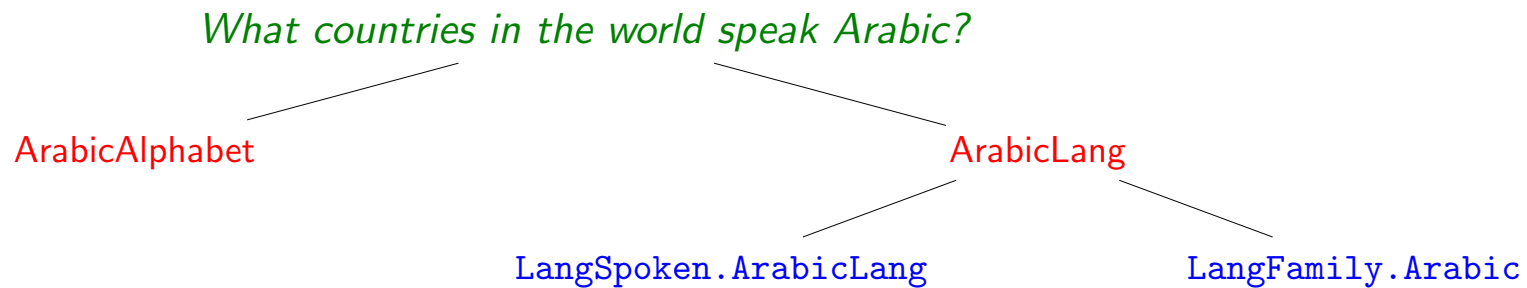
Bridging to nowhere

Search logical forms based on "prior":



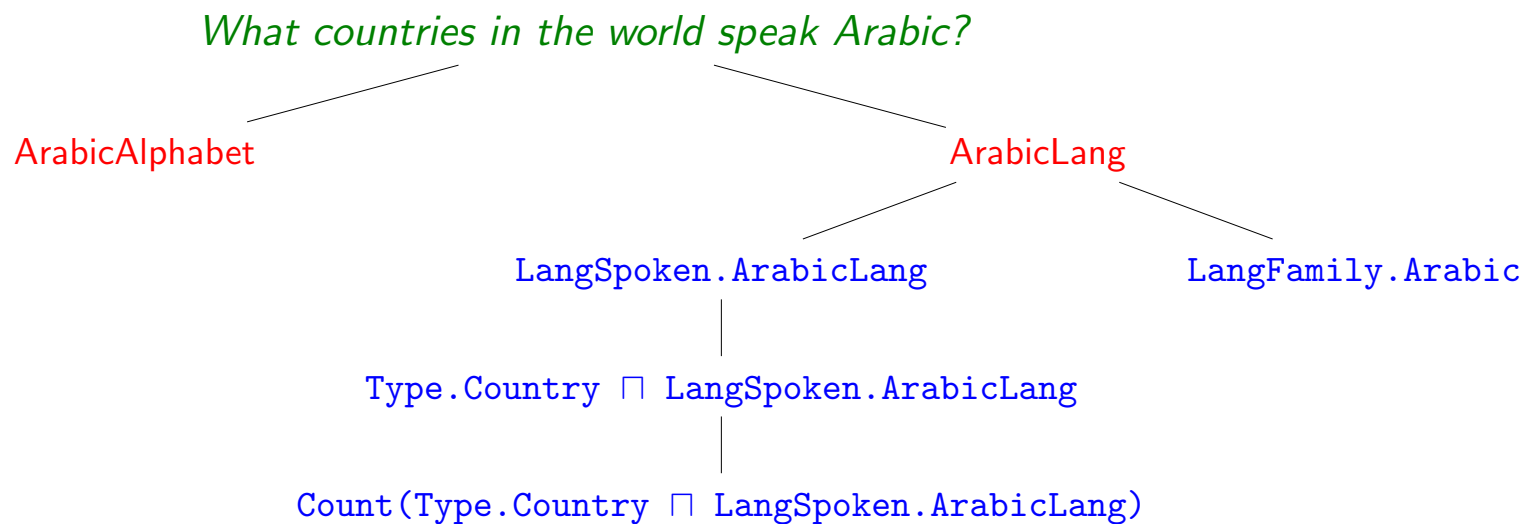
Bridging to nowhere

Search logical forms based on "prior":



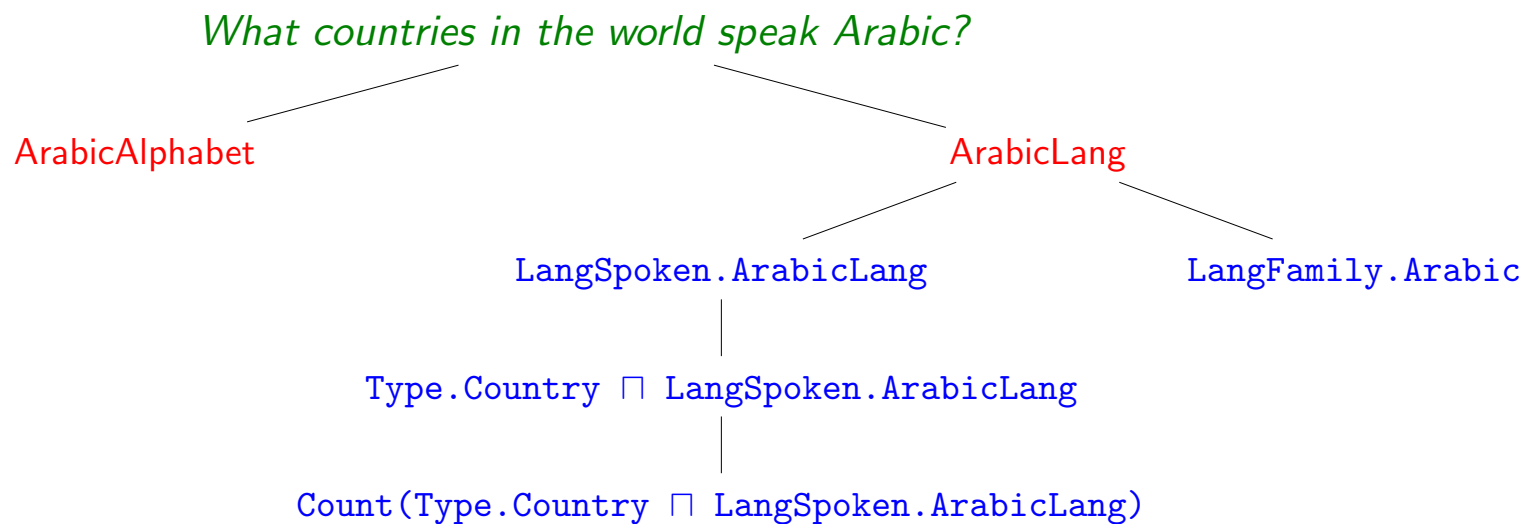
Bridging to nowhere

Search logical forms based on "prior":



Bridging to nowhere

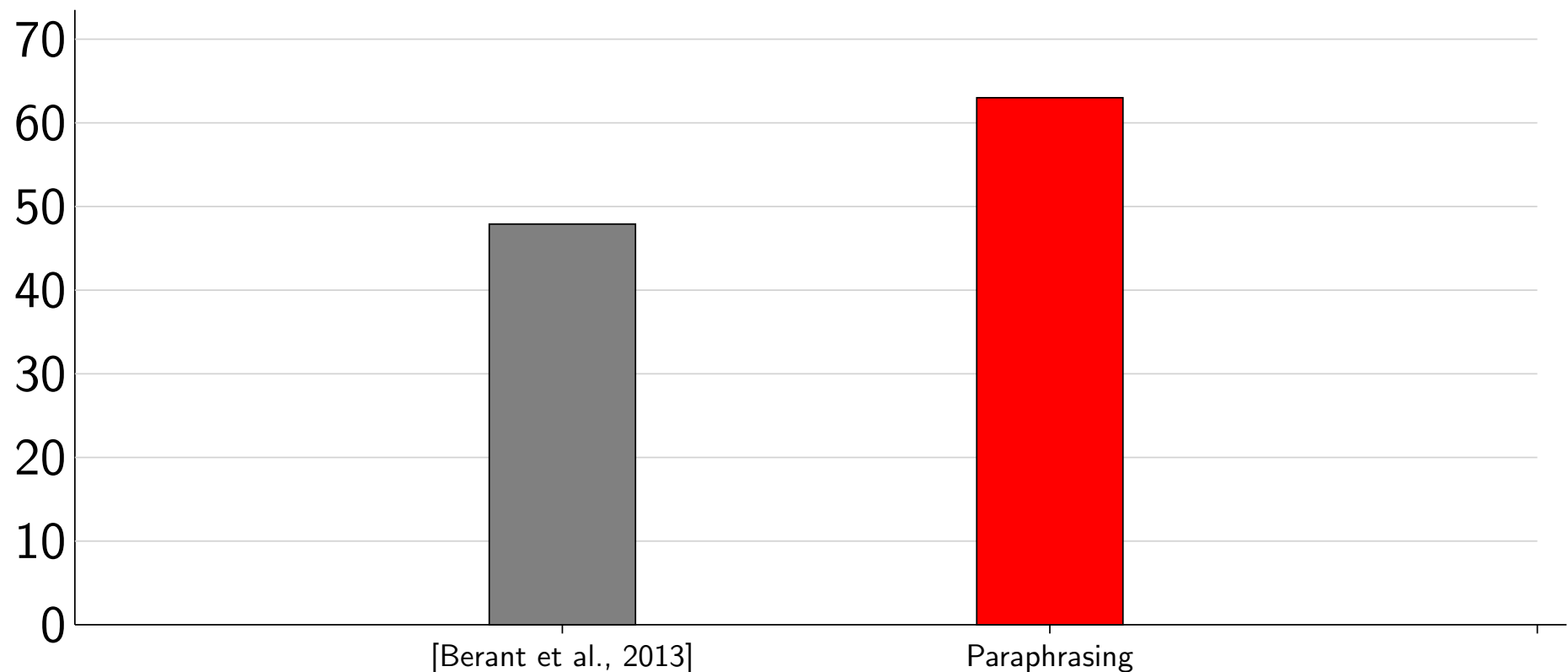
Search logical forms based on "prior":



Start building from parts with more certainty

Oracle on WebQuestions

For what fraction of utterances was a candidate logical form correct?



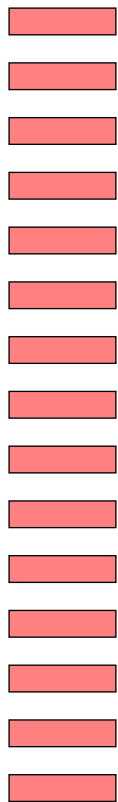
Outline

- A semantic parsing framework
- A closer look at the elements
 - Logical forms: lambda DCS
 - Lexical coverage
 - Grammar: building logical forms
 - **Learning via bootstrapping**
 - Leveraging denotations
 - Datasets/results
- Beyond Freebase
- Final remarks

Bootstrapping from easy examples

Iteration 1

Example 1



...

Example 2



...

Example 3



...

Example 4



...

Example 5

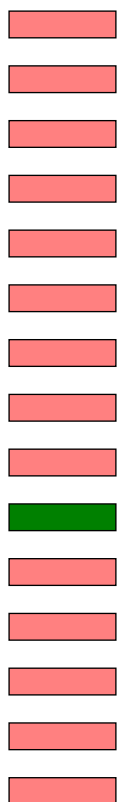


...

Bootstrapping from easy examples

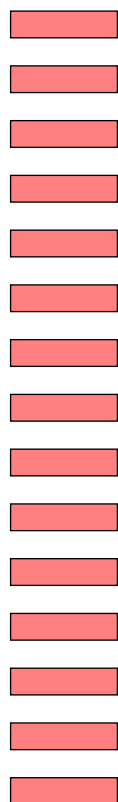
Iteration 2

Example 1



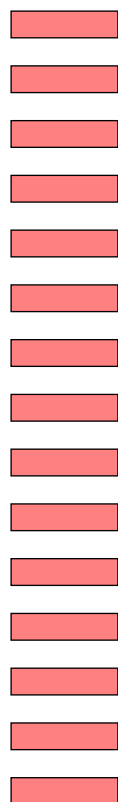
...

Example 2



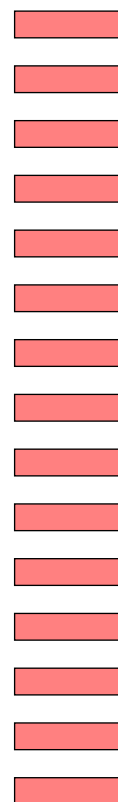
...

Example 3



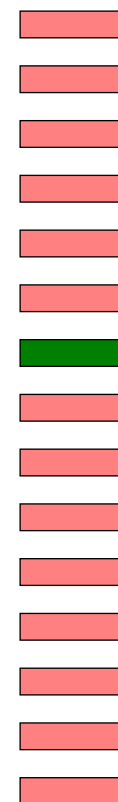
...

Example 4



...

Example 5

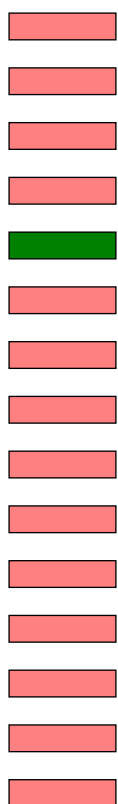


...

Bootstrapping from easy examples

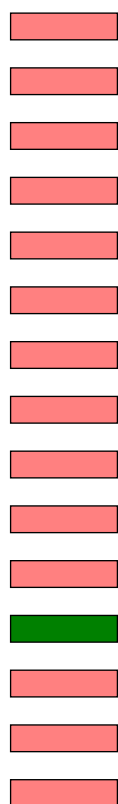
Iteration 3

Example 1



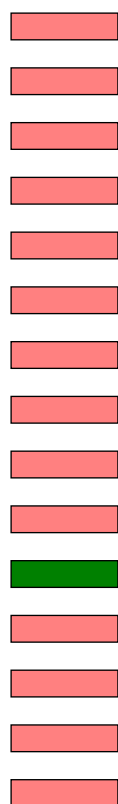
...

Example 2



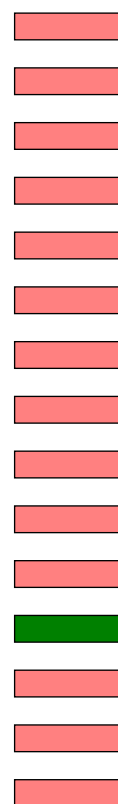
...

Example 3



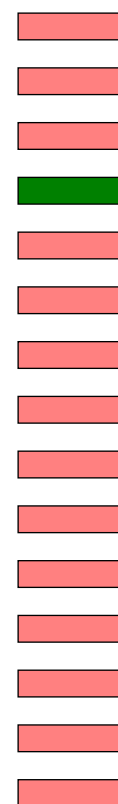
...

Example 4



...

Example 5

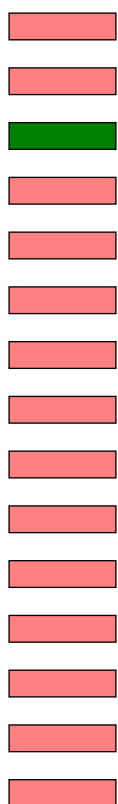


...

Bootstrapping from easy examples

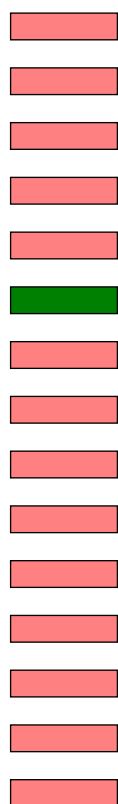
Iteration 4

Example 1



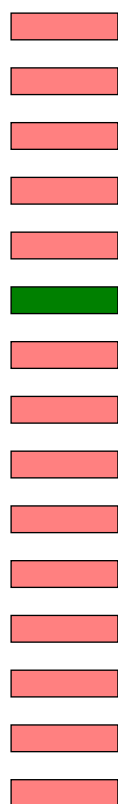
...

Example 2



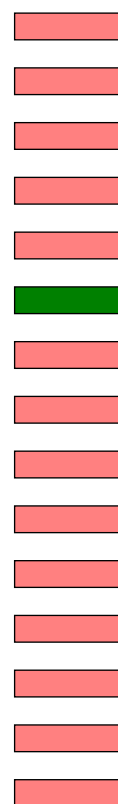
...

Example 3



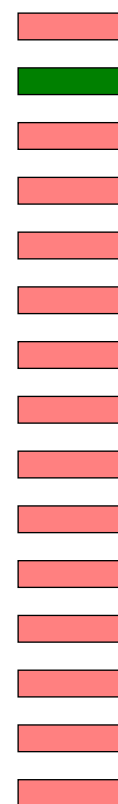
...

Example 4



...

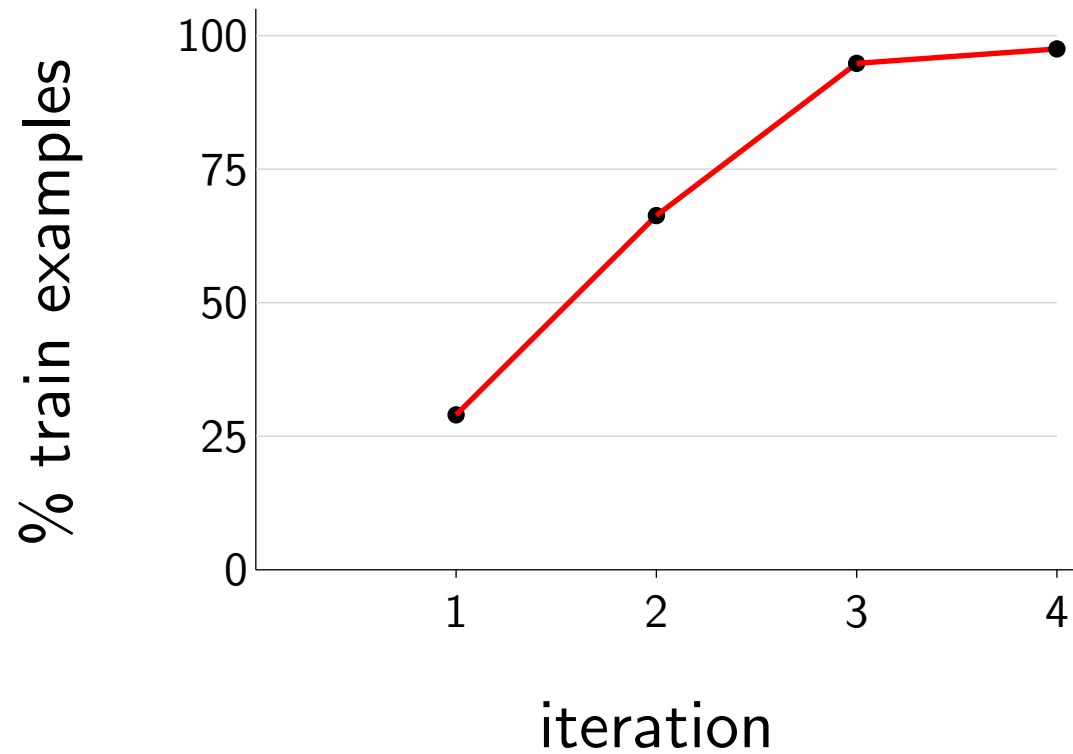
Example 5



...

Bootstrapping from easy examples

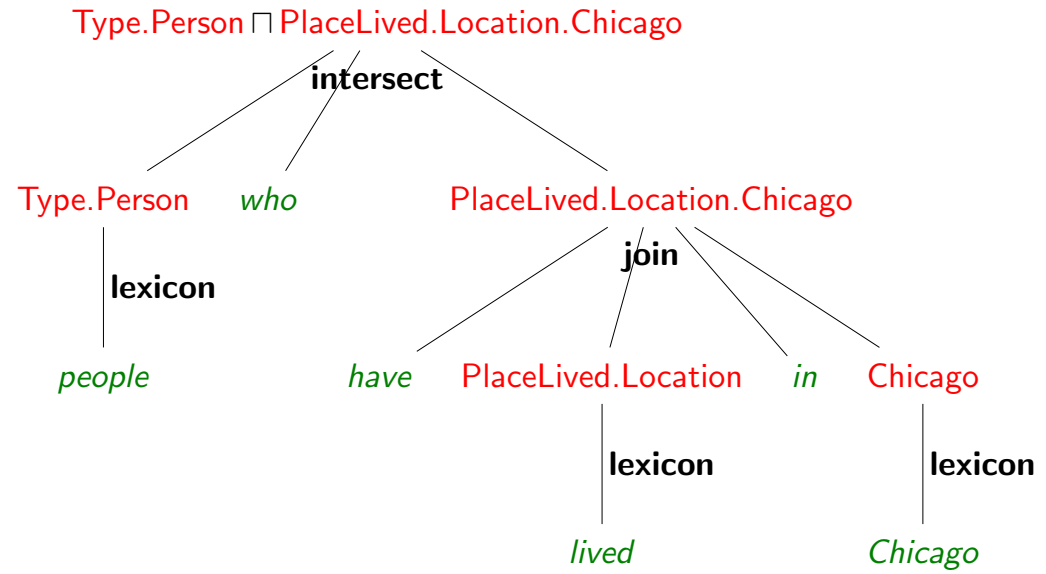
On GeoQuery [Liang et al., 2011]:



Outline

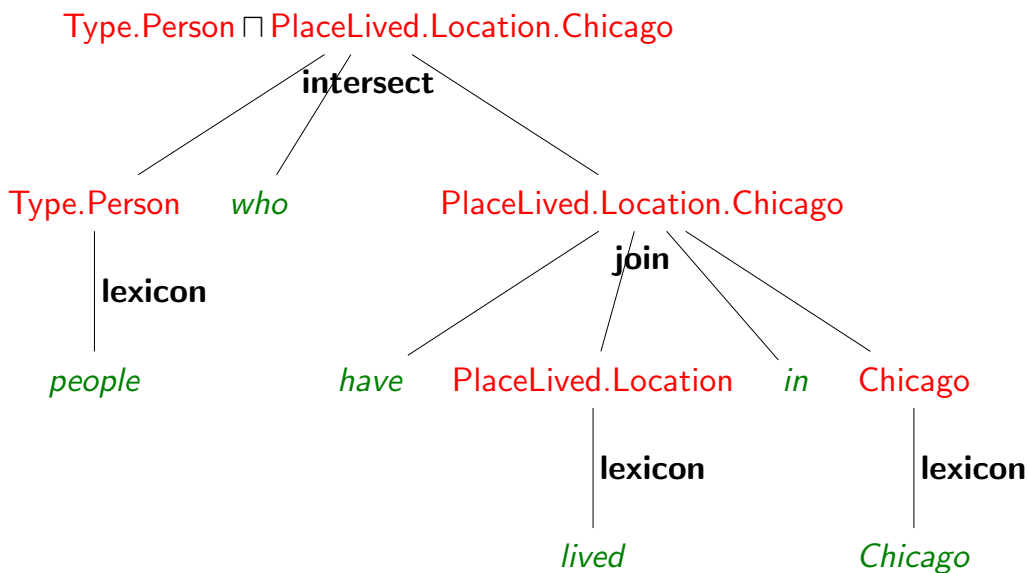
- A semantic parsing framework
- A closer look at the elements
 - Logical forms: lambda DCS
 - Lexical coverage
 - Grammar: building logical forms
 - Learning via bootstrapping
 - **Leveraging denotations**
 - Datasets/results
- Beyond Freebase
- Final remarks

x : utterance
 d : derivation



Feature vector $\phi(x, d) \in \mathbb{R}^F$:

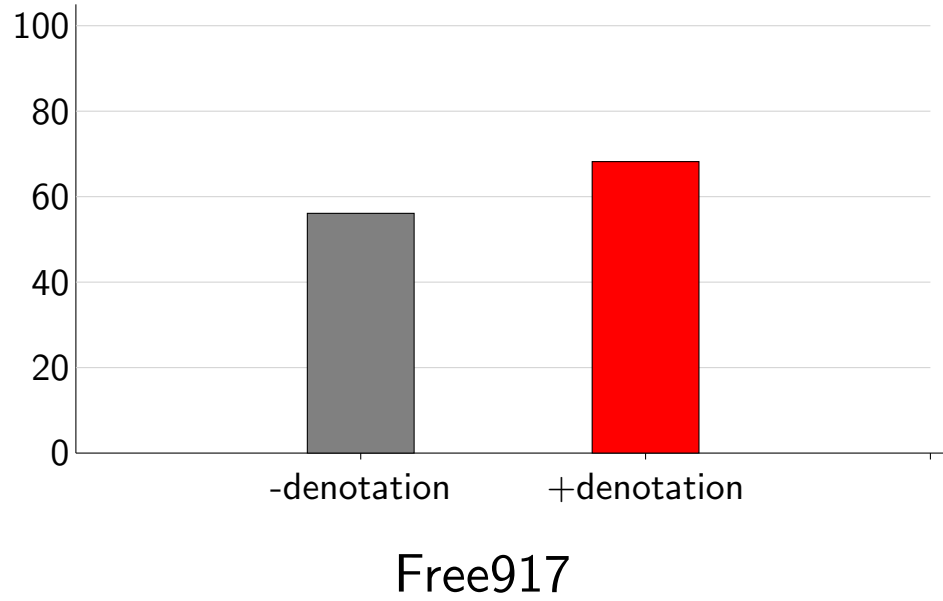
x : utterance
 d : derivation



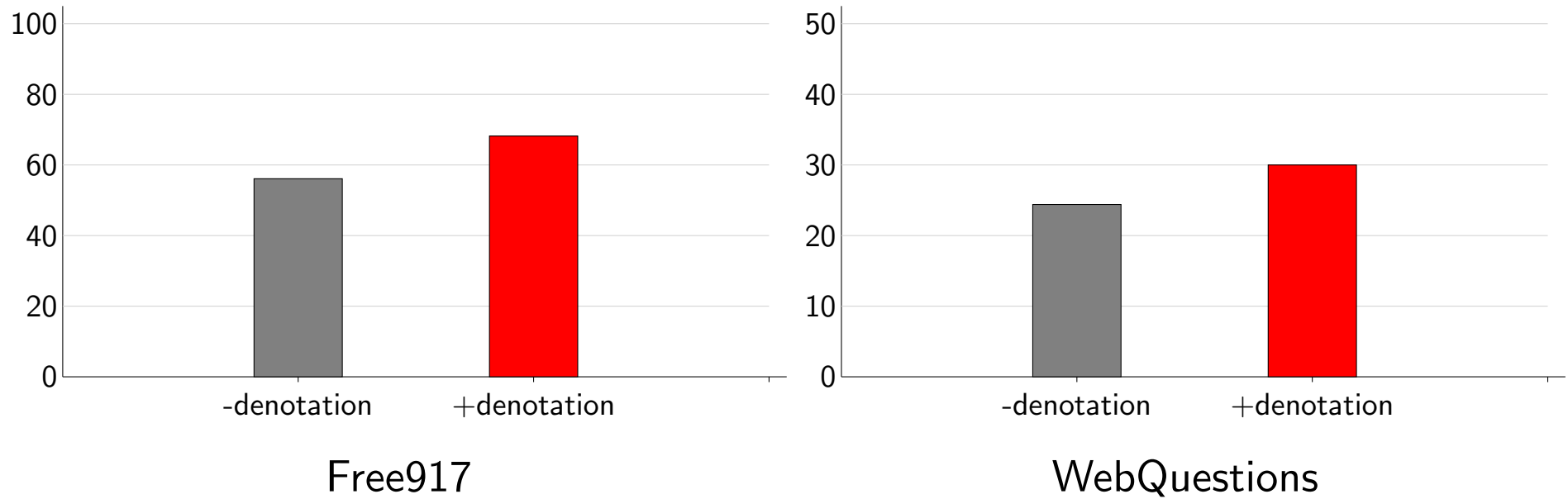
Feature vector $\phi(x, d) \in \mathbb{R}^F$:

apply join	1
apply intersect	1
apply lexicon	3
skipped IN	1
skipped NN	0
<i>lived</i> maps to PlacesLived.Location	1
<i>lived</i> maps to PlaceOfBirth	0
alignmentScore	1.52
denotation-size=1	1
...	...

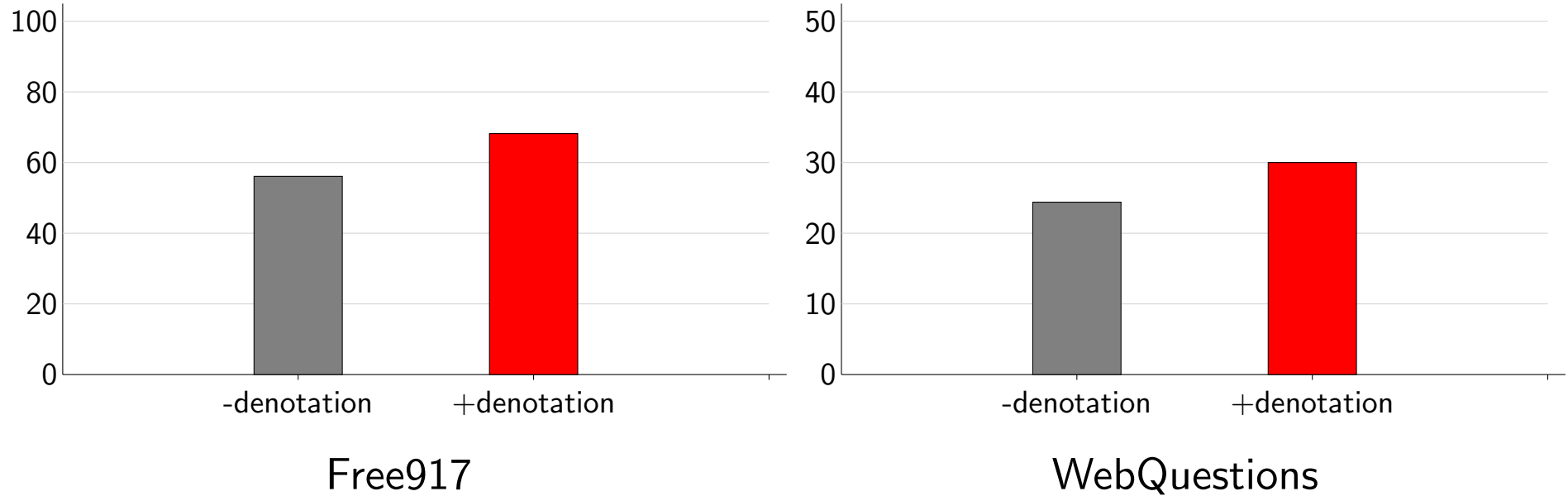
Impact of denotation features



Impact of denotation features



Impact of denotation features



Working with denotations actually provides more information than just logical forms

Outline

- A semantic parsing framework
- A closer look at the elements
 - Logical forms: lambda DCS
 - Lexical coverage
 - Grammar: building logical forms
 - Learning via bootstrapping
 - Leveraging denotations
 - **Datasets/results**
- Beyond Freebase
- Final remarks

Question answering on

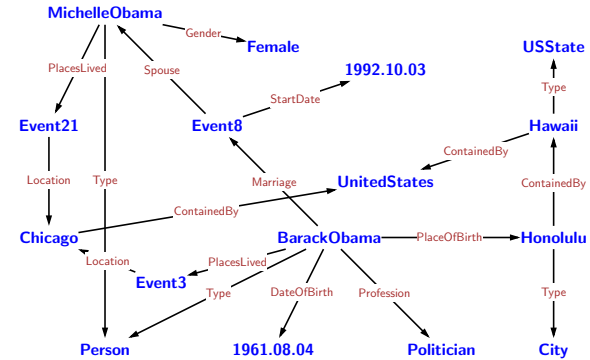


Free917 (1K questions) [Cai & Yates, 2013]

At what institutions was Marshall Hall a professor?

How many companies are traded by the NYSE?

Who was the newscaster in 1948 on CBS evening news?



Question answering on

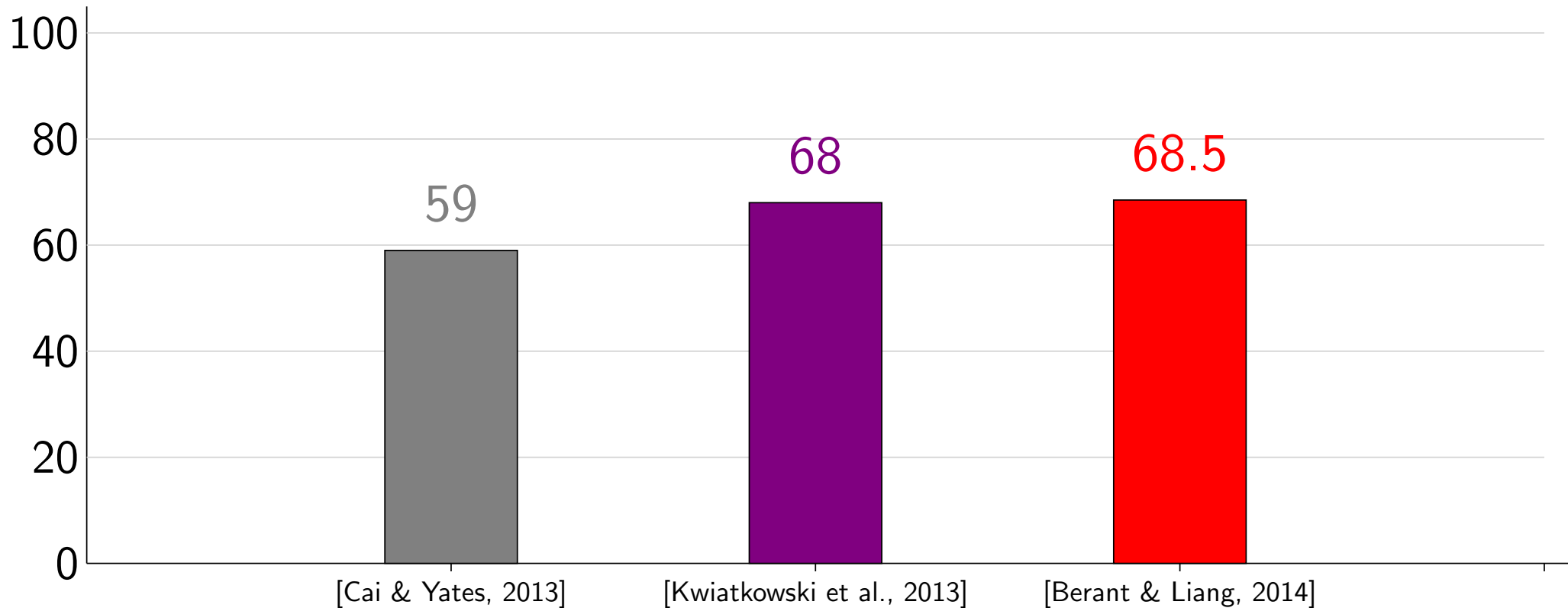
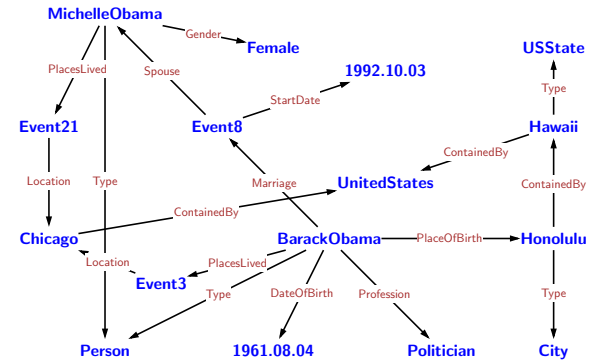


Free917 (1K questions) [Cai & Yates, 2013]

At what institutions was Marshall Hall a professor?

How many companies are traded by the NYSE?

Who was the newscaster in 1948 on CBS evening news?



Dataset collection

How to get natural questions (inputs)?

Dataset collection

How to get natural questions (inputs)?

Strategy: breadth-first search over Google Suggest graph

Dataset collection

How to get natural questions (inputs)?

Strategy: breadth-first search over Google Suggest graph

Where was Barack Obama born?

Dataset collection

How to get natural questions (inputs)?

Strategy: breadth-first search over Google Suggest graph

Where was *Barack Obama* born?

Where was *_* born?  Barack Obama
Lady Gaga
Steve Jobs

Dataset collection

How to get natural questions (inputs)?

Strategy: breadth-first search over Google Suggest graph

Where was *Barack Obama* born?

Where was *_* born?  Barack Obama
Lady Gaga
Steve Jobs

Where was *Steve Jobs* born?

Dataset collection

How to get natural questions (inputs)?

Strategy: breadth-first search over Google Suggest graph

Where was *Barack Obama* born?

Where was *_* born? Google Suggest → Barack Obama
Lady Gaga
Steve Jobs

Where was *Steve Jobs* born?

Where was *Steve Jobs* *_*? Google Suggest → born
raised
on the Forbes list

Dataset collection



How to get natural questions (inputs)?

Strategy: breadth-first search over Google Suggest graph

Where was *Barack Obama* born?

Where was *_* born?  Google Suggest  Barack Obama
Lady Gaga
Steve Jobs

Where was *Steve Jobs* born?

Where was *Steve Jobs* *_*?  Google Suggest  born
raised
on the Forbes list

Where was *Steve Jobs* raised?

Dataset collection

How to get natural questions (inputs)?

Strategy: breadth-first search over Google Suggest graph

Where was *Barack Obama* born?

Where was *_* born? $\xrightarrow{\text{Google Suggest}}$ Barack Obama
Lady Gaga
Steve Jobs

Where was *Steve Jobs* born?

Where was *Steve Jobs* *_*? $\xrightarrow{\text{Google Suggest}}$ born
raised
on the Forbes list

Where was *Steve Jobs* raised?

...

AMT annotation \Rightarrow 6.6K question/answer pairs

Question answering on

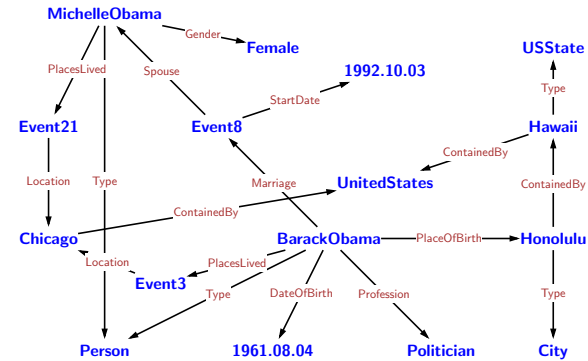


WebQuestions dataset (6K questions) [Berant et al., 2013]

what did obama study in school

where to fly into bali

what was tupac name in juice



Question answering on

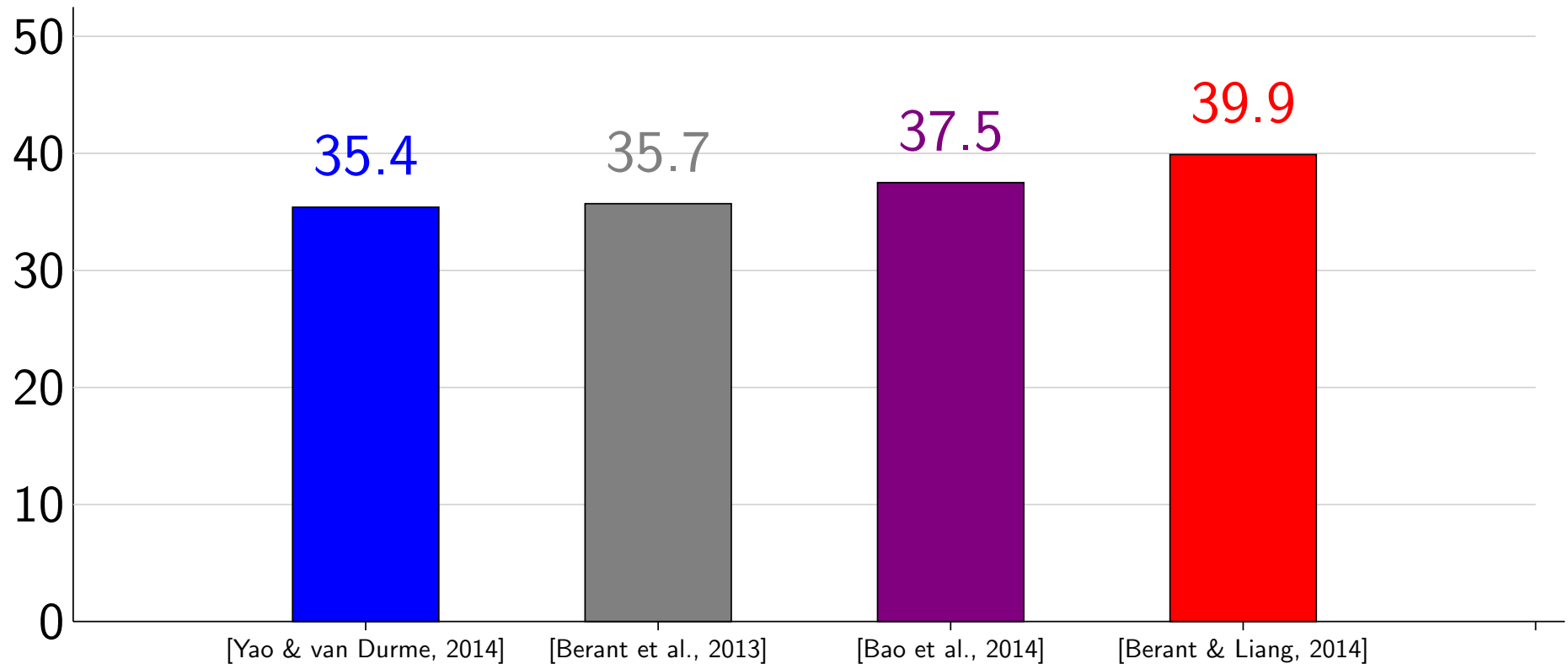
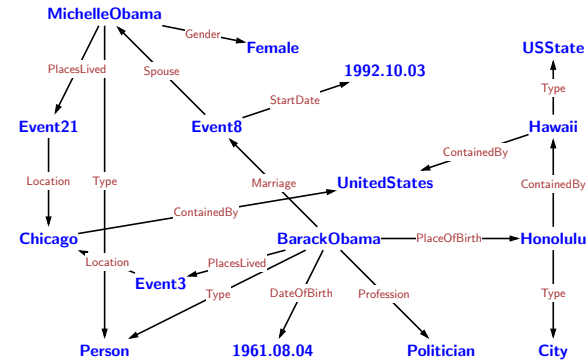


WebQuestions dataset (6K questions) [Berant et al., 2013]

what did obama study in school

where to fly into bali

what was tupac name in juice



Other semantic parsing tasks

Playing computer games [Branavan et al., 2010, 2011]

Following navigational instructions [Tellex et. al 2011; Chen et. al 2012; Artzi et. al 2013]

Understanding visual world [Matuszek et. al; 2012, Krishnamurthy & Kollar, 2013]

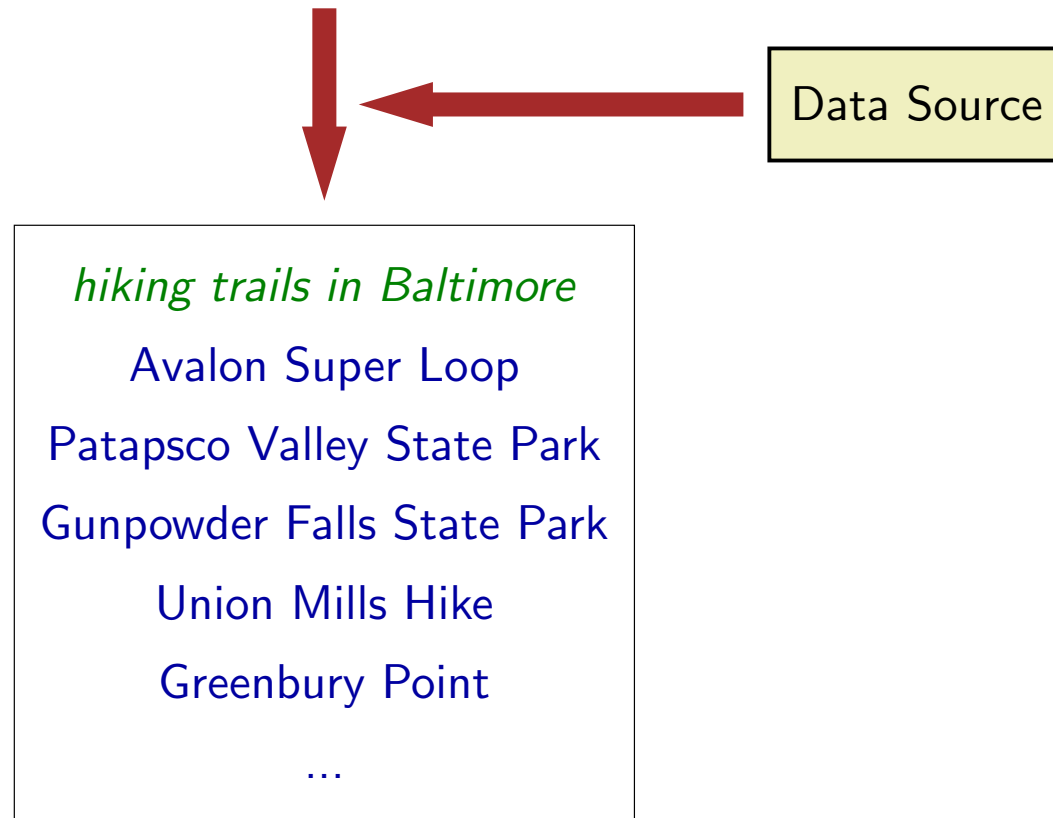
Solving algebra word problems [Kushman et. al, 2013]

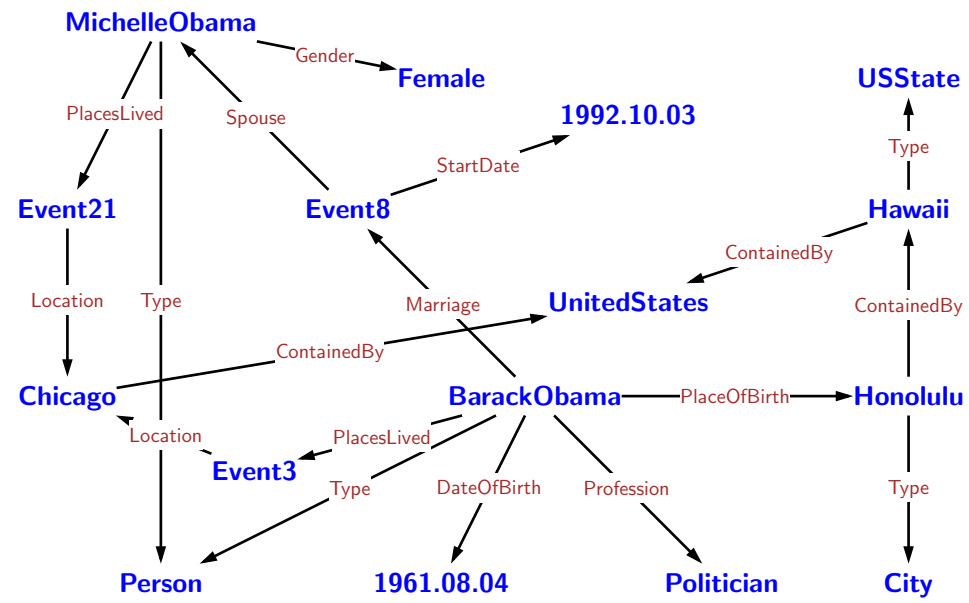
Outline

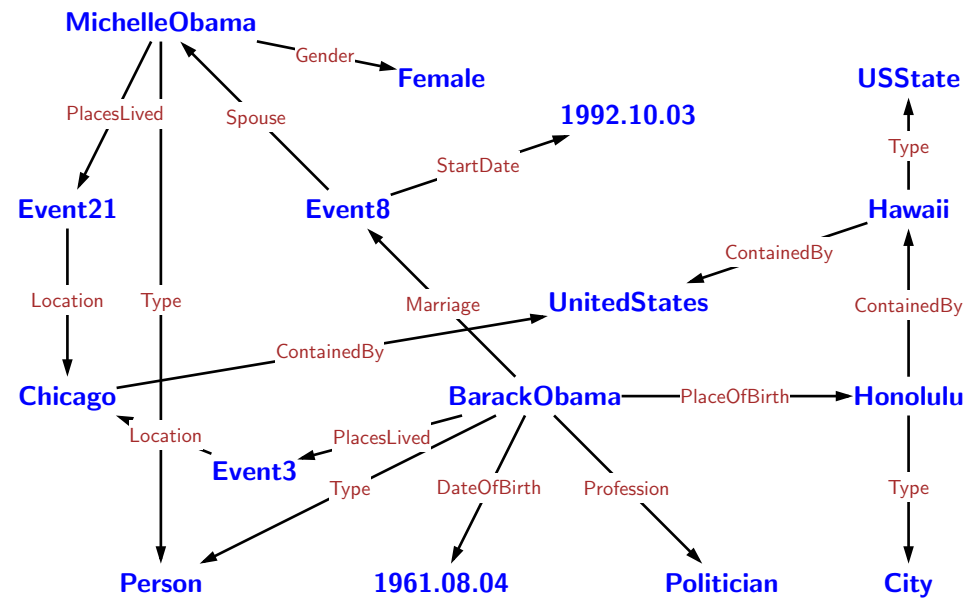
- A semantic parsing framework
- A closer look at the elements
 - Logical forms: lambda DCS
 - Lexical coverage
 - Grammar: building logical forms
 - Learning via bootstrapping
 - Leveraging denotations
 - Datasets/results
- **Beyond Freebase**
- Final remarks

Challenge: incomplete knowledge base

What are the longest hiking trails in Baltimore?







Fewer than 10% general web questions can be answered via Freebase

ACCEPTED LONG PAPERS

ACL 2014

- A Bayesian Mixed Effects Model of Literary Character
David Bamman, Ted Underwood and Noah A. Smith
- A chance-constrained measure of inter-annotator agreement for syntactic
Arne Sippel
- A Decision Theoretic Approach to Natural Language Generation
Nathan McKinley and Saumya May
- A Descriptive Graph-Based Parser for the Abstract Meaning Representation
Jeffrey Flanigan, Sam Thomson, John Carbonell, Goro Doi and Mostafa
- A Generalized Language Model
Rene Plathard, Thomas Goto
- A Joint Graph Model for Phylogenetic
Zhongyi Jin and Hui Zhao
- A Learning-Based Algorithm
Alexander M. Rush, Michael Collins
- A Linear-Time Bottom-Up Disambiguation
Vanesa Wei Feng and Goro Doi
- A practical but linguistically-motivated
Denis Paperno, Theodor A. G. Fischer
- A Provably Correct Learning Algorithm
Shay B. Cohen and Michael Collins

Regular Faculty

54 people

Name

Alex Atkeson

Berant Batzoglou

Gil Bejerman

Michael Bernstein

Dan Borra

David Chaffin

Steve Cooper

Bill Dally

David Diez

Ron Dror













Dawson Engler

Ron Fedkiw

Most Popular Action Feature Films

- Godzilla (2014)**
★★★★★ 7.2/10
The world's most famous monster is pitted against invulnerable creatures who, in humanity's scientific arrogance, threaten our very existence.
Dir: Gareth Edwards With: Aaron Taylor-Johnson, Elizabeth Olsen, Bryan Cranston
Action | Sci-Fi | Thriller
- X-Men: Days of Future Past**
★★★★★
The X-Men send Wolverine to the past in a desperate effort to change history and prevent a future that results in the extinction of his kind.
Dir: Bryan Singer With: Hugh Jackman, James McAvoy, Michael Fassbender
Action | Adventure | Sci-Fi
- The Amazing Spider-Man**
★★★★★
Peter Parker runs the gauntlet against Iron Man, the Vulture, and the Green Goblin.
Dir: Marc Webb With: Andrew Garfield, Emma Stone, Rhys Ifans
Action | Adventure | Sci-Fi
- Transformers: Age of Extinction**
★★★★★
An automobile mechanic discovers a secret about his own past.
Dir: Michael Bay With: Mark Wahlberg, Will Bridges, Anthony Mackie
Action | Adventure | Sci-Fi

#	President	Took office
1	 George Washington (1732-1799) (1789-1797)	April 30, 1789 (3-3)
2	 John Adams (1735-1826) (1797-1801)	March 4, 1797
3	 Thomas Jefferson (1743-1826) (1801-1809)	March 4, 1801
4	 James Madison (1751-1836) (1809-1817)	March 4, 1809

Thursday 12 June			
12:00 (2014) - 13:00 Local Time	 BRAZIL	17:00	CROATIA 
GROUP A Brazil, Croatia, Brazil, Brazil			
Friday 13 June			
13:00 (2014) - 14:00 Local Time	 MEXICO	13:00	CAMEROON 
GROUP B Mexico, Cameroon, Mexico, Mexico			
16:00 (2014) - 17:00 Local Time	 SPAIN	16:00	NETHERLANDS 
GROUP B Spain, Netherlands, Spain, Spain			
18:00 (2014) - 19:00 Local Time	 CHILE	18:00	AUSTRALIA 
GROUP B Chile, Australia, Chile, Chile			
Saturday 14 June			
13:00 (2014) - 14:00 Local Time	 COLOMBIA	13:00	GREECE 
GROUP C Colombia, Greece, Colombia, Colombia			
16:00 (2014) - 17:00 Local Time	 URUGUAY	16:00	COSTA RICA 
GROUP B Uruguay, Costa Rica, Uruguay, Uruguay			

Semantic parsing on the web

Input:

- query x

hiking trails near Baltimore

- web page w

Semantic parsing on the web

Input:

HOME | EXPLORE | MOBILE APPS | CREATE TRIP | MY EVERYTRAIL GO

(Update Current Location) Login | Signup

Hiking near Baltimore, Maryland

Like 49 people like this. Tweet 1

This list shows the most popular Hiking near Baltimore, Maryland based on user reviews, votes, and mobile downloads. Plan your next trip with EveryTrail guides by downloading a guide to your mobile phone with the EveryTrail iPhone or Android app.

Sort: show community trips

Filter Trails

Guides

Avalon Super Loop - Patapsco State Park

Patapsco State Park, Maryland, United States (7.5 miles away)

★★★★★

Difficult: 12.7 miles, Full day

lots of ruins, waterfalls, trains, and river views

Do the entire Avalon Patapsco state park in 1 day! This loop covers the majority of the Avalon area, with multiple ruins, waterfalls and other artifacts to find along the way. Starting at the parking lot, you hike up the road a ways to the Ridge trail sign. The next leg is the maintenance loop which has an old old tractor to look at and some...

Patapsco Valley State Park - Hilton Area 8 Miles/Moderate

Catonsville, Maryland, United States (7.7 miles away)

Moderate: 7.8 miles, Half day

8 mile circuit hike including sections in the Avalon, Orange Grove and Glen Artney areas of PVSP.

OVERVIEW: One of the more scenic routes in the Patapsco Valley State Park in the Hilton Area which includes multiple stream crossings, viewings and waterfalls including Cascade waterfalls, two swinging bridge crossings, Ilchester Overlook, and Bloedes Dam. This is a moderate hike and can be hiked in either direction. Counterclockwise is an easier hike...

Popular places for Hiking

- Hiking in Maryland
- Hiking in Patapsco Valley State Park
- Hiking in Calvert Cliffs State Park
- Hiking in Patuxent River State Park

Semantic parsing on the web

Input:

EveryTrail

HOME | EXPLORE | MOBILE APPS | CREATE TRIP | MY EVERYTRAIL

Search **GO**

(Update Current Location) [Login](#) | [Signup](#)

Hiking near Baltimore, Maryland

Like 49 people like this. **Tweet** 1

This list shows the most popular Hiking near Baltimore, Maryland based on user reviews, votes, and mobile downloads. Plan your next trip with EveryTrail guides by downloading a guide to your mobile phone with the EveryTrail iPhone or Android app.

Sort: show community trips

Filter Trails

Guides

Avalon Super Loop - Patapsco State Park

Patapsco State Park, Maryland, United States (7.5 miles away)

★★★★★

Difficult: 12.7 miles, Full day

lots of ruins, waterfalls, trains, and river views

Do the entire Avalon Patapsco state park in 1 day! This loop covers the majority of the Avalon area, with multiple ruins, waterfalls and other artifacts to find along the way. Starting at the parking lot, you hike up the road a ways to the Ridge trail sign. The next leg is the maintenance loop which has an old old tractor to look at and some...

Patapsco Valley State Park - Hilton Area 8 Miles/Moderate

Patapsco Valley State Park, Maryland, United States (8 miles away)

Moderate: 7.8 miles, Half day

8 mile circuit hike including sections in the Avalon, Orange Grove and Glen Artney areas of PVSP.

OVERVIEW: One of the more scenic routes in the Patapsco Valley State Park in the Hilton Area which includes multiple stream crossings, viewings and waterfalls including Cascade waterfalls, two swinging bridge crossings, Ilchester Overlook, and Bloedes Dam. This is a moderate hike and can be hiked in either direction. Counterclockwise is an easier hike...

Google [Terms of Use](#)

Popular places for Hiking

[Hiking in Maryland](#)

[Hiking in Patapsco Valley State Park](#)

[Hiking in Calvert Cliffs State Park](#)

[Hiking in Patuxent River State Park](#)

Semantic parsing on the web

Input:

- query x

hiking trails near Baltimore

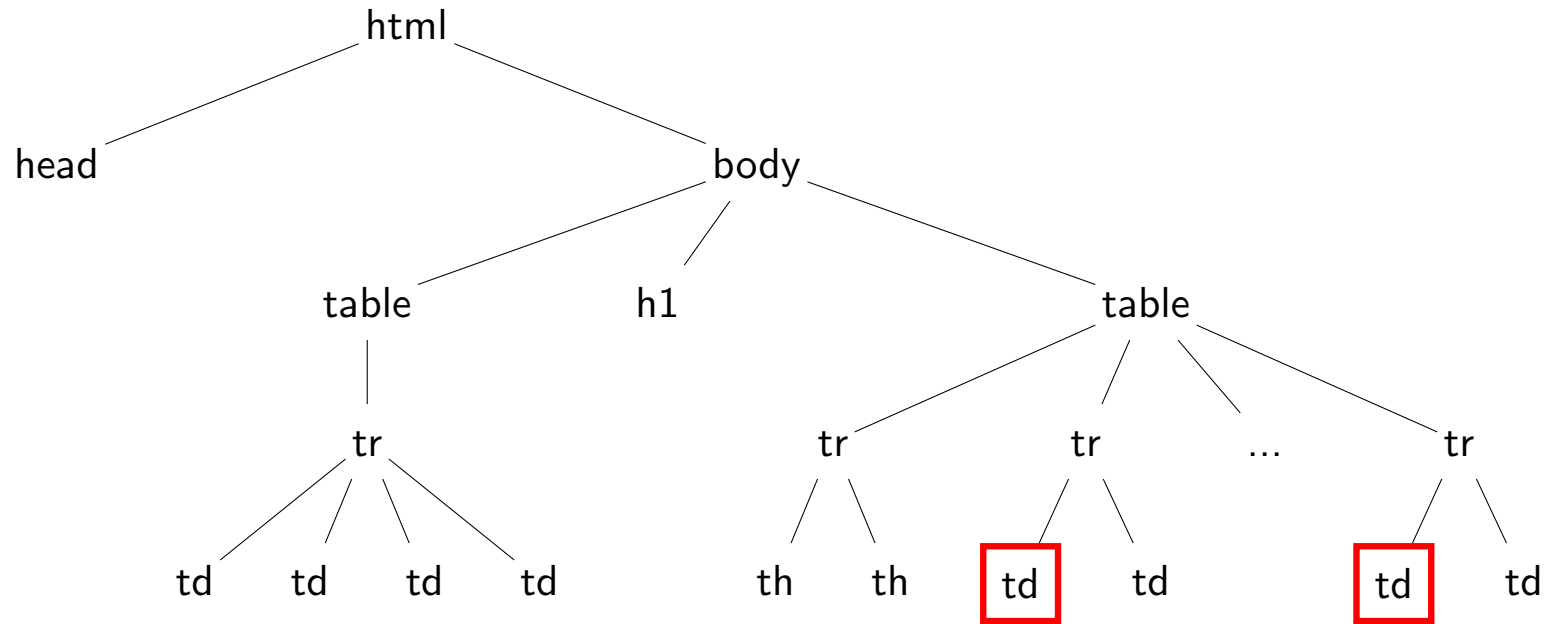
- web page w

Output:

- list of entities y

[Avalon Super Loop, Patapsco Valley State Park, ...]

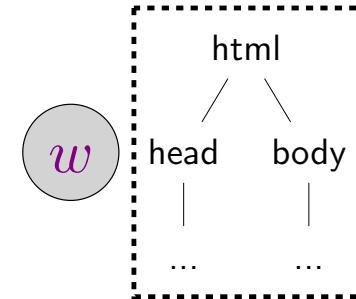
Logical forms: XPath expressions



$z = /html[1]/body[1]/table[2]/tr/td[1]$

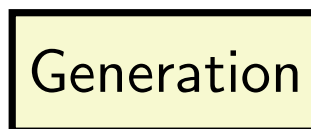
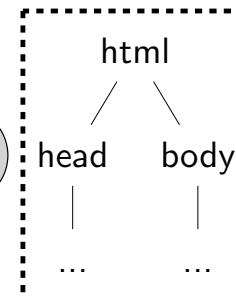
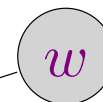
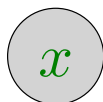
Framework

*hiking trails
near Baltimore* x

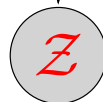


Framework

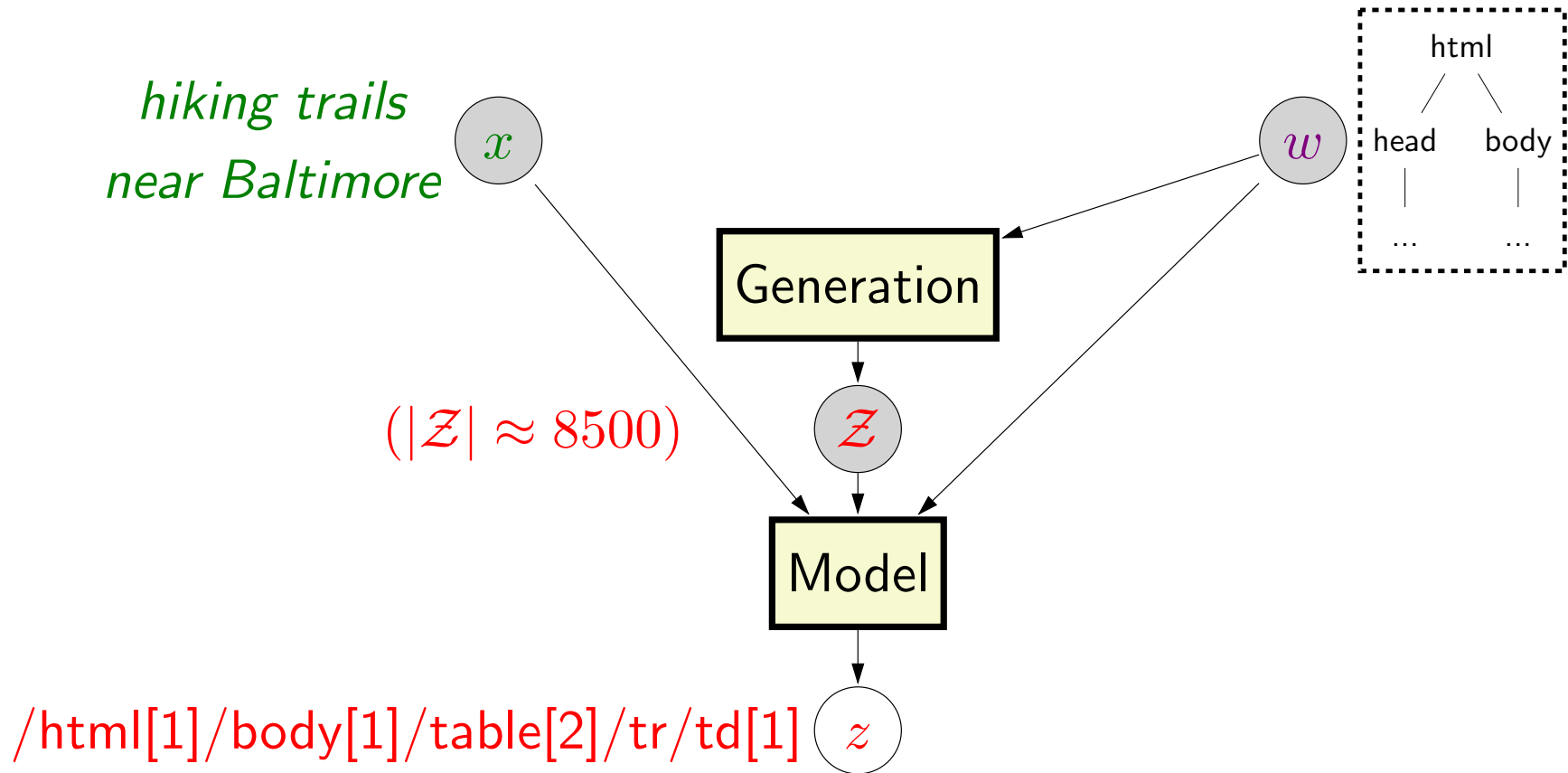
*hiking trails
near Baltimore*



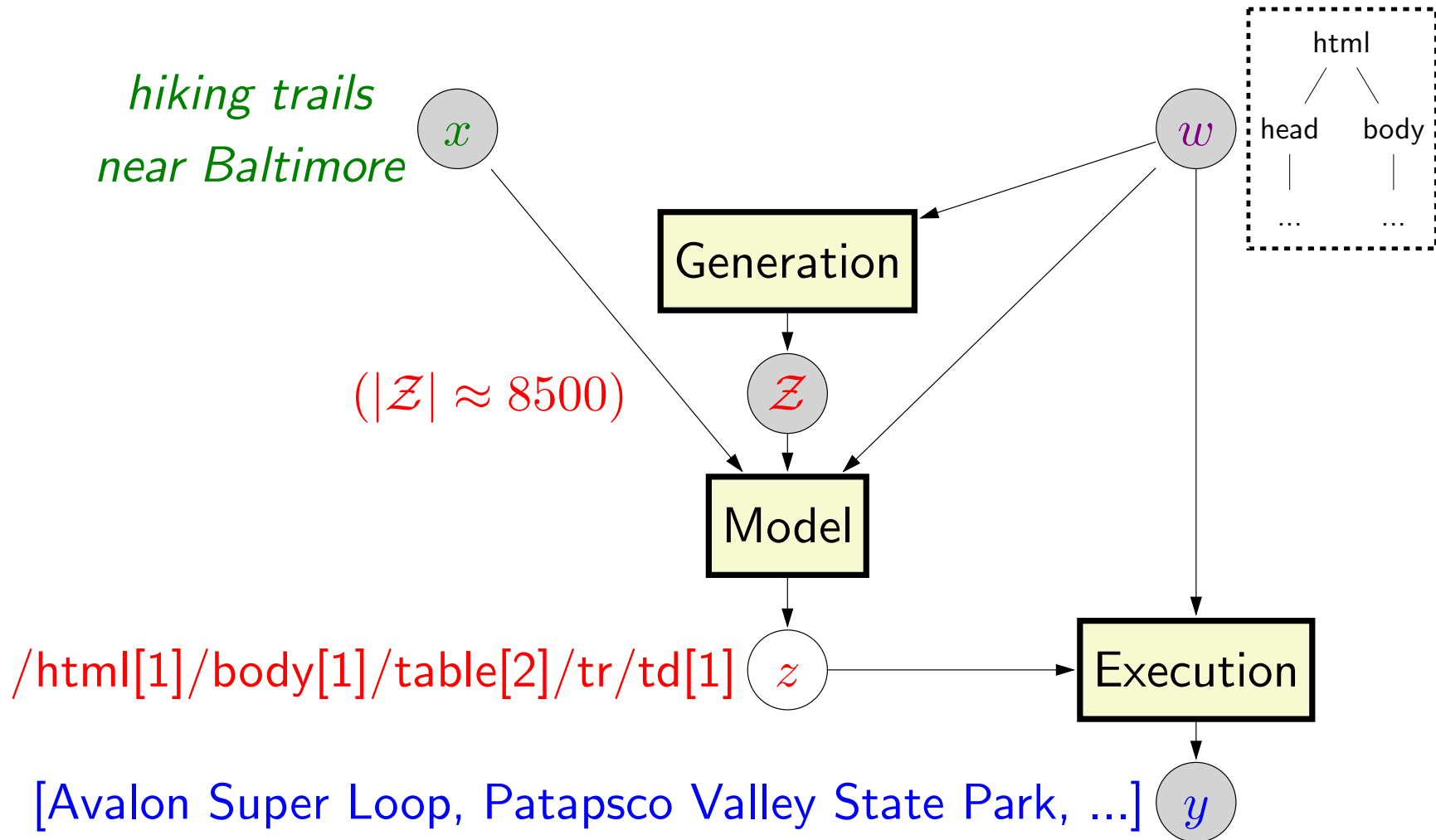
$(|\mathcal{Z}| \approx 8500)$



Framework



Framework



Features

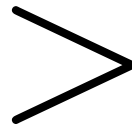
$$p_{\theta}(z \mid x, w) \propto \exp\{\theta^{\top} \phi(x, w, z)\}$$

Features

$$p_{\theta}(z | x, w) \propto \exp\{\theta^{\top} \phi(x, w, z)\}$$

Structural Features: captures context

No	President	Took office
1	 George Washington (1732–1799) [11][12][13]	April 30, 1789 [n 2]
2	 John Adams (1735–1826) [15][16][17]	March 4, 1797
3	 Thomas Jefferson (1743–1826) [18][19][20]	March 4, 1801



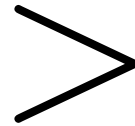
The listing below is complete for the current government of the USA. For this country, however, there were prior governments (including that under the **Articles of Confederation**). Prior to George Washington as first president under the current constitution, there were **twelve people** in leadership over the government of the United States of America who held the title of "President". Also during the **Civil War**, there was the position of **"President of the Confederate States of America"** in an entity separate from the USA, and this position was held by **one person**.

Features

$$p_{\theta}(z | x, w) \propto \exp\{\theta^{\top} \phi(x, w, z)\}$$

Denotation Features: captures content

hiking trails near Baltimore
Avalon Super Loop
Patapsco Valley State Park
Gunpowder Falls State Park
Rachel Carson Conservation Park
Union Mills Hike
...



hiking trails near Baltimore
Home
About Baltimore Tour
Pricing
Contact
Online Support
...

Dataset

We created the OPENWEB dataset with diverse queries and web pages.

airlines of italy

natural causes of global warming

lsu football coaches

bf3 submachine guns

badminton tournaments

foods high in dha

technical colleges in south carolina

songs on glee season 5

singers who use auto tune

san francisco radio stations

Dataset

We created the OPENWEB dataset with diverse queries and web pages.

The screenshot shows the Ranker website with a search for "All Italian Airlines". The results list the top 7 airlines:

Rank	Name	Hubs
1	Air Dolomiti	Munich Airport, Verona Villafranca Airport, Tre Venezie Orio al Serio Airport
2	Air Europe	Malpensa Airport
3	Air Italy	Malpensa Airport, Leonardo da Vinci-Fiumicino Airport, Verona Villafranca Airport
4	Air One	Malpensa Airport
5	Air Vallée	Phma Airport, Asolo Airport, Federico Fellini International Airport
6	Alidaunia	Foggia "Gino Lisa" Airport
7	Alitalia-Linee Aeree Italiane	Malpensa Airport, Leonardo da Vinci-Fli Airport

airlines of italy

The first diagram, titled "The Greenhouse Effect", shows a sun emitting rays that hit the Earth's surface. Some rays are reflected, while others are absorbed and re-emitted as heat, which is trapped by greenhouse gases in the atmosphere. The second diagram, titled "Air Pollution", shows a person carrying a large sack of trash, with a cloud of smoke or pollution rising from the sack.

10. Greenhouse Effect

Greenhouse effect is the process in which the atmosphere of the Earth trap some of the heat coming from the sun, making the Earth warm but due to burning fuels, cutting trees, the concentration of heat on Earth is increased to abnormal levels making greenhouse effect as one of the major causes of global warming. Carbon Dioxide, methane, nitrous oxide are the greenhouse gases which helps to keep the Earth warm. It is a natural phenomenon that takes place with the adequate concentrations of the greenhouse gases. But when the concentration of these gases rises, they disturb the climatic conditions, making the Earth more warm. These gases are not able to escape, which is the cause of worldwide increase in temperature. So the balance of carbon dioxide and other gases should be maintained so that it does not become the major reason of global warming.

9. Air Pollution

The harmful gases emitted from the vehicles and factories and the greenhouse gases cause pollution in the air and these gases get trapped in the atmosphere. The smoke, rather in in the atmosphere forms a layer full of harmful

natural causes of global warming

The screenshot shows the LSU Football Coaches page for the 2013-14 season. It features a grid of 12 coaches with their names and titles:

Name	Title
Les Miles	Head Coach
Cam Cameron	Offensive Coordinator/Quarterbacks Coach
John Chavis	Defensive Coordinator
Frank Wilson	Running Backs Coach/Recruiting Coordinator
Steve Ensminger	Tight Ends Coach
Brick Haley	Defensive Line Coach
Adam Henry	Wide Receivers Coach
Thomas McGaughey	Special Teams Coordinator
Corey Raymond	Defensive Backs Coach
Greg Studrawa	Offensive Line Coach
Steve Kragthorpe	Administrator
Tommy Moffitt	Strength & Conditioning Coordinator
Dr. Sam Nader	Assistant Athletics Director - Football Coordinator

lsu football coaches

Dataset Statistics

2773 examples

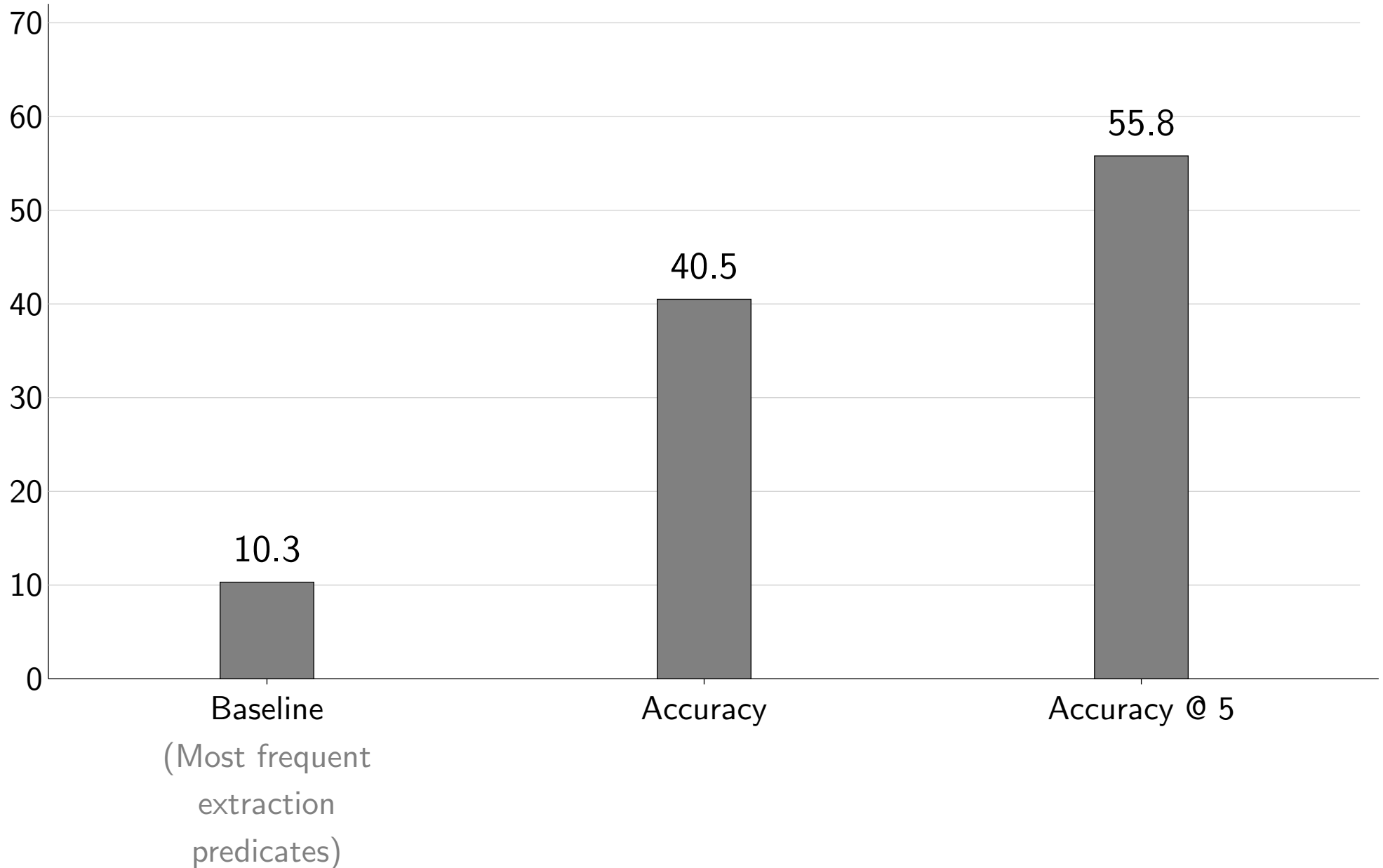
2269 unique queries

894 unique headwords ← long tail!

1483 unique web domains ← long tail!

(≠ wrapper induction)

Main Results



Examples of Correct Predictions

Query: *nobel prize winners*

The Nobel Prize in Physics 2013
François Englert and **Peter W. Higgs**
"for the theoretical discovery of a mechanism that contributes to our understanding of the origin of mass of subatomic particles, and which recently was confirmed through the discovery of the predicted fundamental particle, by the ATLAS and CMS experiments at CERN's Large Hadron Collider"

The Nobel Prize in Physics 2012
Serge Haroche and **David J. Wineland**
"for ground-breaking experimental methods that enable measuring and manipulation of individual quantum systems"

The Nobel Prize in Physics 2011
Saul Perlmutter, Brian P. Schmidt and **Adam G. Riess**
"for the discovery of the accelerating expansion of the universe through observations of distant supernovae"

/html[1]/body/div/div[2]/div/div/div/h6/a/text

Examples of Coverage Errors

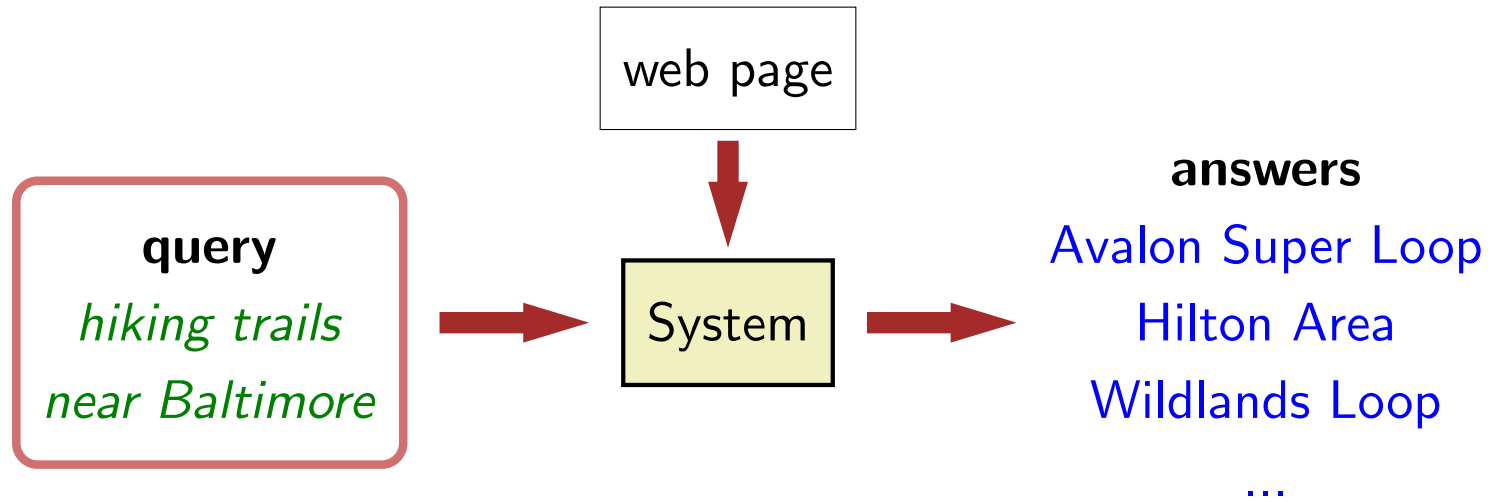
Query: *hedge funds in new york*

Rank ↕	Firm	Headquarters ↕
1	Bridgewater Associates	 Westport, CT
2	Man Group	 London
3	J.P. Morgan Asset Management	 New York
4	Brevan Howard Asset Management	 London
5	Och-Ziff Capital Management Group	 New York
6	Paulson & Co.	 New York
7	BlackRock Advisors	 New York

/html/body/div[3]/div[3]/div[4]/.../table/tbody/tr/td[2]/a

Need compositionality!

Summary



A framework for extracting entities from a **natural language query** and a single web page

Outline

- A semantic parsing framework
- A closer look at the elements
 - Logical forms: lambda DCS
 - Lexical coverage
 - Grammar: building logical forms
 - Learning via bootstrapping
 - Leveraging denotations
 - Datasets/results
- Beyond Freebase
- **Final remarks**

Semantics

[utterance: user input]



semantic parsing

[intermediate semantic representations (text? logical forms? vectors?)]



execute

[denotation: user output]

Semantic representations as a means to an end

Long-term engineering vision

hiking trails near Baltimore

Long-term engineering vision

hiking trails near Baltimore under 5 miles

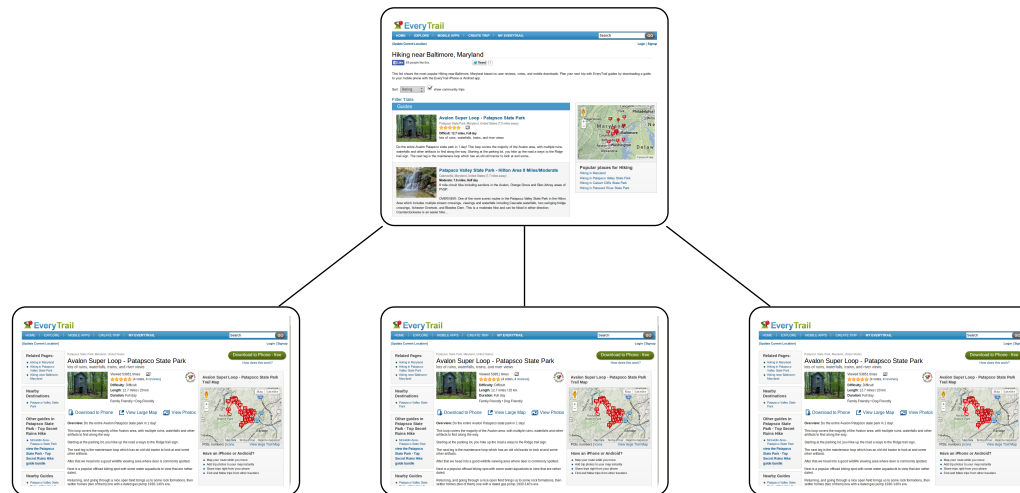
Long-term engineering vision

dog-friendly hiking trails near Baltimore under 5 miles

Long-term engineering vision

dog-friendly hiking trails near Baltimore under 5 miles

`/body/div[2]/(CLICK)/body/div/(SEARCH)/body/div[3]`



Conclusions

- **Learning and computation** are important to constrain representation

Conclusions

- **Learning and computation** are important to constrain representation
- Learning from denotations framework in which logical forms are **intermediate**

Conclusions

- **Learning and computation** are important to constrain representation
- Learning from denotations framework in which logical forms are **intermediate**
- Need **richer representations** that connect to real world with richer language

Code and data online

<http://www-nlp.stanford.edu/software/sempr/>

<http://www-nlp.stanford.edu/software/web-entity-extractor-ACL2014/>

Code and data online

<http://www-nlp.stanford.edu/software/sempr/>

<http://www-nlp.stanford.edu/software/web-entity-extractor-ACL2014/>

Collaborators

Jonathan Berant (post-doc)

Andrew Chou (masters)

Roy Frostig (Ph.D.)

Panupong Pasupat (Ph.D.)

Code and data online

<http://www-nlp.stanford.edu/software/sempr/>

<http://www-nlp.stanford.edu/software/web-entity-extractor-ACL2014/>

Collaborators

Jonathan Berant (post-doc)

Andrew Chou (masters)

Roy Frostig (Ph.D.)

Panupong Pasupat (Ph.D.)

Thank you!