

Львівський національний університет імені Івана Франка
Філологічний факультет
Кафедра загального мовознавства

“ЗАТВЕРДЖУЮ”
Декан філологічного
факультету
С.М.Пилипчук
“30” серпня 2018 року



РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

ВСТУП ДО КОМП'ЮТЕРНОЇ ЛІНГВІСТИКИ

галузі знань	<u>03 Гуманітарні науки;</u> <u>01 Освіта/Педагогіка</u>
спеціальності	<u>035 Філологія;</u> <u>014 Середня освіта</u>
спеціалізації	<u>035.01 Українська мова та література;</u> <u>014.01 Українська мова і література</u>
факультет	<u>філологічний</u>

Львів – 2018 рік

Робоча програма навчальної дисципліни «Вступ до комп'ютерної лінгвістики» для студентів за спеціальностями 035 Філологія; 014 Середня освіта, спеціалізаціями 035.01 Українська мова та література; 014.01 Українська мова і література. – Львів: ЛНУ імені Івана Франка, 2018. – 13 с.

Розробник: Бук С.Н., доцент кафедри загального мовознавства,
канд. філол. наук, доцент

Робоча програма розглянута на засіданні кафедри загального мовознавства.

Протокол від “30” серпня 2018 року № 1.

Завідувач кафедри загального мовознавства



(Бацевич Ф.С.)

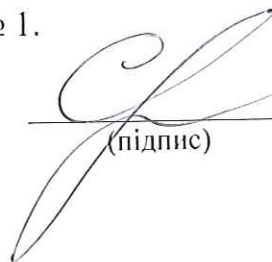
(підпис)

“30” серпня 2018 року

Затверджено Вченою радою філологічного факультету.

Протокол від “30” серпня 2018 року № 1.

Голова



(Пилипчук С.М.)

(підпис)

“30” серпня 2018 року

1.Опис навчальної дисципліни

Найменування показників	Галузь знань, напрям підготовки, освітньо-кваліфікаційний рівень	Характеристика навчальної дисципліни	
		<u>денна форма навчання</u>	
Кількість кредитів – 3	<u>Галузі знань:</u> 03 Гуманітарні науки; 01Освіта/Педагогіка	За вибором студента	
	<u>Спеціальності:</u> 035 Філологія; 014 Середня освіта		
Модулів – 2	<u>Спеціалізації:</u> 035.01 Українська мова та література; 014.01 Українська мова і література	Рік підготовки:	
Змістових модулів – 2		3-й	
Індивідуальне науково-дослідне завдання реферат-есе (назва)		Семестр	
Загальна кількість годин – 90		5-й	
Тижневих годин для денної форми навчання: аудиторних – 2 самостійної роботи студента – 3,6	<u>Освітньо-кваліфікаційний рівень:</u> Бакалавр	Лекції	
		32 год.	
		Практичні, семінарські	
		-	
		Лабораторні	
		-	
		Самостійна робота	
		58 год.	
		Індивідуальні завдання:	
-			
Вид контролю: залік			

Примітка.

Співвідношення кількості годин аудиторних занять до самостійної і індивідуальної роботи становить:

для денної форми навчання – 1:1,8
для заочної форми навчання -

2. Мета та завдання навчальної дисципліни

Курс “Основи комп'ютерної лінгвістики” займає важливе місце у підготуванні кваліфікованих фахівців-філологів, адже застосування комп'ютерних технологій у лінгвістиці відкривають нові шляхи у дослідженні літератури та мови, а також володіють величезним потенціалом у вирішенні багатьох теоретичних завдань мовознавства і практичних завдань обробки текстових даних.

Мета: ознайомлення студентів з основними поняттями комп'ютерної лінгвістики та її методами дослідження.

Найважливішими завданнями курсу є ліквідувати упередженість студентів-філологів у можливості використання комп'ютерів для розв'язання лінгвістичних задач; дати студентам необхідні теоретичні знання про мову і мовлення з позицій комп'ютерної лінгвістики; навчити студентів типовим методам та алгоритмам розв'язання лінгвістичних задач у комп'ютерних системах; ознайомити студентів з типовими лінгвістичними системами.

В результаті вивчення даного курсу студент повинен

знати сфери застосування комп'ютерних технологій у мовознавстві; сучасний стан комп'ютерних досліджень у світовому мовознавстві та в Україні; основні поняття та категорійний апарат комп'ютерної лінгвістики; теоретичні основи опису мови і мовлення з позицій комп'ютерної лінгвістики

уміти застосовувати методи комп'ютерної лінгвістики для розв'язання лінгвістичних завдань; працювати з комп'ютерними програмами автоматичного опрацювання текстів; пояснити роботу найпростіших алгоритмів для опрацювання текстів.

Курс пов'язаний з іншими навчальними дисциплінами. До їх числа входять “Сучасна українська мова”, “Загальне мовознавство”, “Основи статистичної лінгвістики”, “Лексикографія та корпусна лінгвістика”. Знання з алгебри та інформатики потрібні в обсязі, що дає середня освіта.

3. Програма навчальної дисципліни

МОДУЛЬ 1

Змістовий модуль 1. Основні поняття та категорійний апарат комп'ютерної лінгвістики. Лінгвістичні моделі

Тема 1. Проблеми комп'ютерної лінгвістики. Предмет, об'єкт та завдання курсу «Основи комп'ютерної лінгвістики».

1. Дефініція поняття “комп'ютерна лінгвістика”. Зв'язок із математичними та мовознавчими дисциплінами.
2. Історія виникнення комп'ютерної лінгвістики.
3. Проблеми, які вивчають у курсі “комп'ютерна лінгвістика” у “східній” та “західній” лінгвістичній традиції.
4. Місце комп'ютерної лінгвістики серед розділів та напрямків мовознавства. (На основі ілюстрації з *Карпіловська Є. А. Вступ до комп'ютерної*

- лінгвістики. С. 12-13; *Bolshakov, Igor. A., Gelbukh Alexander Computational Linguistics. P. 18.*)
5. Стан розвитку комп'ютерної лінгвістики в Україні. Які питання комп'ютерної лінгвістики висвітлено в українських лінгвістичних енциклопедичних словниках (на прикладі УМЕ).
 6. Предмет, об'єкт та завдання вивчення курсу «Основи комп'ютерної лінгвістики».

Рекомендована література

1. *Грязнухіна Т. О.* Автоматичне оброблення тексту // УМЕ. К., 2000. С. 9.
2. *Грязнухіна Т. О.* Автоматичний аналіз // УМЕ. К., 2000. С. 9 – 10.
3. *Дарчук Н. П.* Комп'ютерна лінгвістика (автоматичне опрацювання тексту): підручник.— К.: Видавничо-поліграфічний центр “Київський університет”, 2008.
4. *Клименко Н. Ф.* Машинний переклад // УМЕ. К., 2000. С. 303 – 304.
5. Комп'ютерна лінгвістика – новий етап розвитку теоретичної і прикладної лінгвістики // *Карпіловська Є. А.* Вступ до комп'ютерної лінгвістики.— Донецьк: Юго-Восток, 2003. С. 7 – 19.
6. Компьютерная лингвистика // *Баранов А. Н.* Введение в прикладную лингвистику: Учебное пособие. — М.: Эдиториал УРСП, 2001.— С. 13 – 25.
7. Основні наукові напрямки комп'ютерної лінгвістики і її зв'язок з іншими науками // *Волошин В. Г.* Комп'ютерна лінгвістика: Навч. посібник.— Суми: Університетська книга, 2004. С. 11 – 21.
8. *Пецак М. М.* Машинний фонд української мови // УМЕ. К., 2000. С. 304.
9. *Перебийніс В. С.* Автоматизація лінгвістичних систем // УМЕ. К., 2000. С. 8 – 9.
10. Проблемы моделирования языка в действии. Компьютерная лингвистика // *Алефиренко Н. Ф.* Современные проблемы науки о языке: Учебное пособие.— М.: Флинта: Наука, 2005.— С. 231–234.
11. Products of Computation Linguistics: Present and Prospective // *Bolshakov, Igor. A., Gelbukh Alexander Computational Linguistics. Models, Resources, Applications.*— México, 2004.— P. 53–81.
12. The Oxford Handbook of Computation Linguistics / ed. by *Ruslan Mitkov.*— Oxford Un-ty Press, 2003.— 784 p.
13. What is Computation Linguistics? // *Grishman R.* Computation Linguistics: an introduction (studies in natural language processing).— Cambridge: Cambridge University Press, 1986.— P. 4–9 (ksero).

Тема 2. Лінгвістична модель. Комп'ютерне моделювання мислення

1. Поняття моделі. Вимоги до моделі. Особливості лінгвістичної моделі.
2. Природна мова як модель дійсності та джерело комп'ютерного моделювання.
3. Поняття мислення. Декларативні та процедурні знання.
4. Структури знань: фрейм, сценарій, скрипт, план, семантична сітка
5. Знання, необхідні для моделі світу.
6. Роль словників у моделюванні людського мислення.
7. Міжнародний проект Wordnet. Основні принципи створення, структура та практичне застосування.
8. Міжнародний проект Framenet. Основні принципи створення, структура та практичне застосування. Спільне та відмінне з Wordnet.

Рекомендована література

1. Баранов А. Н. Фреймы и сценарии. Категории искусственного интеллекта в лингвистической семантике. — М., 1987. С. 3–32.
2. Баранов А. Н., Сергеев В. М. Искусственный интеллект и традиционные методы анализа текста описания семантики предложения: опыт типологии // Когнитивные исследования за рубежом. — М., 1990. С. 169–205.
3. Когнітивна наука як наукова дисципліна моделювання людського мислення // Волошин В. Г. Комп'ютерна лінгвістика: Навч. посібник.— Суми: Університетська книга, 2004. С. 154 – 164.
4. Минский М. Фреймы для представления знаний.— М., 1979.
5. Русский Wordnet (структура) <http://www.pgups.ru/WebWN/wordnet.uix>
6. Linguistic Models // Bolshakov, Igor. A., Gelbukh Alexander Computational Linguistics. Models, Resources, Applications.— México, 2004.— P. 129–152. (доступна в Інтернеті)
7. Скороходько Э. С. Сетевое моделирование лексики // Использование ЭВМ в лингвистических исследованиях. К., 1990. 129–161.
8. Структуры представления знаний в языке.— М., 1994.
9. Язык и структуры представления знаний.— М., 1992.
10. Apprehending lexical items: netting and framing // Linde-Usiekiewicz J., Olko M. Multilingual Dictionaries on-line: Reality and Perspectives // Semantyka a konfrontacja językowa. Т. 3.— Warszawa S³awistyczny ośrodek wydawniczy, 2006.— S. 44–48.
11. Framenet <http://framenet.icsi.berkeley.edu/>
12. Wordnet <http://wordnet.princeton.edu/>

Тема 3. Штучний інтелект. Комп'ютерне моделювання спілкування

1. Поняття комунікації. Моделі прямого та опосередкованого мовленнєвого акту.
2. Штучний інтелект. Історичний розвиток поняття.

3. Місце спілкування серед проблем штучного інтелекту
4. Комп'ютерна модель діалогу Дж. Вейценбаума «Еліза».
5. Експеримент М. Макгайра.
6. Висновки експерименту М. Макгайра для теорії комунікації
7. Проблеми створення інтерфейсу природною мовою.

Рекомендована література

1. *Баранов А. Н., Сергеев В. М.* Искусственный интеллект и традиционные методы анализа текста описания семантики предложения: опыт типологии // Когнитивные исследования за рубежом.— М., 1990. С. 169 – 205.
2. *Городецкий Б. Ю.* Компьютерная лингвистика: моделирование языкового общения // Новое в зарубежной лингвистике: [Сб. ст. Переводы]. Вып. 24. Компьютерная лингвистическая / Сост. и вступит. ст. *Б. Ю. Городецкого*.— М.: Прогресс., 1989.— 428 с.— С. 5–31.
3. Моделирование общения // *Баранов А. Н.* Введение в прикладную лингвистику: Учебное пособие. — М.: Эдиториал УРСП, 2001.— С. 20 – 25.
4. Моделювання мовленнєвої діяльності в комп'ютерних діалогових системах // *Карпіловська Є. А.* Вступ до комп'ютерної лінгвістики.— Донецьк: Юго-Восток, 2003. С. 162 – 171.
5. Природний інтелект і штучний інтелект як його модель // *Карпіловська Є. А.* Вступ до комп'ютерної лінгвістики.— Донецьк: Юго-Восток, 2003. С. 101 – 103.
6. Штучний інтелект *Волошин В. Г.* Комп'ютерна лінгвістика: Навч. посібник.— Суми: Університетська книга, 2004. С. 165 – 227.
7. *Bolshakov, Igor. A., Gelbukh Alexander* Computational Linguistics. Models, Resources, Applications.— México, 2004.— 154 p. (доступна в інтернеті)
8. *Grishman R.* Computation Linguistics: an introduction (studies in natural language processing).— Cambridge: Cambridge University Press, 1986.— 193 p. (ksero)
9. The Oxford Handbook of Computation Linguistics / ed. by *Ruslan Mitkov*.— Oxford Un-ty Press, 2003.— 784 p.

МОДУЛЬ 2

Змістовий модуль 2. Методи автоматичного аналізу тексту

Тема 4. Розрізнення анафори у процесі автоматичного оброблення тексту

1. Поняття анафори, антецедента і кореференції у комп'ютерній лінгвістиці. Види анафори, що не мають відсильної функції.
2. Види анафори за частиномовною належністю.
3. Поняття прямої та непрямої анафори. Спільне та відмінне.
4. Джерела знань, необхідні для визначення анафори. Власнелінгвістичні та екстралінгвістичні.
5. Процес та етапи визначення анафори.
6. Значення виокремлення анафори для проблем комп'ютерної лінгвістики.

Рекомендована література

1. Проблемы моделирования языка в действии. Компьютерная лингвистика // *Алефиренко Н. Ф.* Современные проблемы науки о языке: Учебное пособие.— М.: Флинта: Наука, 2005.— С. 231–234.
2. Кибрик А. Анафора // Энциклопедия Кругосвет доступно з <http://www.krugosvet.ru/articles/76/1007610/print.htm>
3. *Городецкий Б. Ю.* Компьютерная лингвистика: моделирование языкового общения // Новое в зарубежной лингвистике: [Сб. ст. Переводы]. Вып. 24. Компьютерная лингвистическая / Сост. и вступит. ст. *Б. Ю. Городецкого.* М.: Прогресс., 1989. С. 5–31.
4. *Mitkov R.* Anaphora resolution // The Oxford Handbook of Computation Linguistics / ed. by *Ruslan Mitkov.*— Oxford Un-ty Press, 2003.— P. 266–282.
5. The Oxford Handbook of Computation Linguistics / ed. by *Ruslan Mitkov.*— Oxford Un-ty Press, 2003.— 784 p.

Тема 5. Комп'ютерне опрацювання природної мови

1. Дефініція комп'ютерного опрацювання природної мови (КОПМ)
2. Основні завдання КОПМ
3. Проблеми на шляху КОПМ
4. Рівні КОПМ
5. Комп'ютерний морфологічний аналіз тексту (автоматичний морфологічний аналіз — АМА)
6. Типи АМА за характером лінгвістичного забезпечення та способом розпізнавання морфологічної структури тексту
7. Розроблення АМА в Україні
8. Теоретичне та практичне значення АМА для КОПМ

Рекомендована література

1. Автоматичний морфологічний аналіз // *Карпіловська Є. А.* Вступ до комп'ютерної лінгвістики.— Донецьк: Юго-Восток, 2003.
2. Автоматический морфологический анализ текста // Использование ЭВМ в лингвистических исследованиях. К., 1990. С. 129–161.
3. Дарчук Н. П., Алексієнко Л. А., Шевчук І. О. Синтез парадигм дієслова в українській мові // Проблеми українізації комп'ютерів. К. 1994. С. 41 – 42.
4. *Грязнухіна Т. О.* Лінгвістичне забезпечення автоматизованих систем управління // Мовознавство.— 1983.— № 5.— С. 39–43.
5. *Дарчук Н. П.* Комп'ютерна лінгвістика (автоматичне опрацювання тексту): підручник.— К.: Видавничо-поліграфічний центр “Київський університет”, 2008.

6. Лук'янчук С. Комп'ютерна модель парадигматичних класів дієслів // Українське мовознавство.— 2000.— Вип. 22.—С. 82–85.
7. Олексієнко Л., Дарчук Н. Лематизація парадигм іменників української мови // Проблеми українізації комп'ютерів.— К., 1993.— С. 62–65.
8. Пецяк М. М. Нариси з комп'ютерної лінгвістики.— Ужгород: Видавництво закарпаття, 1999.— 200 с.
9. Natural Language Processing // Free on-line encyclopedia Wikipedia

Тема 6. Автоматичний синтаксичний аналіз

1. Мета автоматичного синтаксичного аналізу (АСА), його типи (за сферою застосування, за вихідними елементами, за способом виконання)
2. АСА за безпосередніми складниками речення.
3. Графічне представлення синтаксичних структур
4. Теоретичне та практичне значення АСА для КОПМ

Рекомендована література

1. Автоматичний синтаксичний аналіз тексту // Карпіловська Є. А. Вступ до комп'ютерної лінгвістики.— Донецьк: Юго-Восток, 2003.С. 139-143
2. Грязнухина Т.А., Дарчук Н.П. и др. Синтаксический анализ на ЭВМ. - К. 1993.
3. Падучева Е.В. О способах представления синтаксической структуры предложения // Вопр. языкознания, 1964. - № 2. -С.99-113.
4. Севбо И.П. Графическое представление синтаксических структур и стилистическая диагностика. - К., 1981. - 190 с.
5. Фитиалов С.Я. О моделировании синтаксиса в структурной лингвистике.- М., 1962.- С. 100-114.
6. Frost, R., Hafiz, R. and Callaghan, P. (2007) « Modular and Efficient Top-Down Parsing for Ambiguous Left-Recursive Grammars » 10th International Workshop on Parsing Technologies (IWPT), ACL-SIGPARSE , Pages: 109—120, June 2007, Prague.

4. Структура навчальної дисципліни

Назви змістових модулів і тем	Кількість годин					
	Денна форма					
	Усього	у тому числі				
		л	п	лаб	інд	ср
1	2	3	4	5	6	7
МОДУЛЬ 1						
Змістовий модуль 1. Основні поняття та категорійний апарат комп'ютерної лінгвістики. Лінгвістичні моделі						
Тема 1. Проблеми комп'ютерної лінгвістики. Предмет, об'єкт та завдання курсу «Основи комп'ютерної лінгвістики».	13	4				9
Тема 2. Лінгвістична модель. Комп'ютерне моделювання мислення	15	4				11
Тема 3. Штучний інтелект. Комп'ютерне моделювання спілкування	19	8				11
<i>Разом – зм. модуль 1</i>	47	16				31
МОДУЛЬ 2						
Змістовий модуль 2. Методи автоматичного аналізу тексту						
Тема 4. Розрізнення анафори у процесі автоматичного оброблення тексту	13	4				9
Тема 5. Комп'ютерне опрацювання природної мови	13	4				9
Тема 6. Автоматичний синтаксичний аналіз	17	8				9
<i>Разом – зм. модуль 2</i>	43	16				27
<i>Разом</i>	90	32				58

5. Самостійна робота

№ з/п	Назва теми	Кількість годин
1	Міжнародний проект Wordnet. Основні принципи створення, структура та практичне застосування.	4
2	Міжнародний проект Framenet? Основні принципи створення, структура та практичне застосування.	4
3	Структури знань: фрейм, сценарій, скрипт, план, семантична сітка	4
4	Стан розвитку комп'ютерної лінгвістики в Україні. Які питання комп'ютерної лінгвістики висвітлено в українських лінгвістичних енциклопедичних словниках (на прикладі УМЕ)?	6
5	Комп'ютерна модель діалогу Дж. Вейценбаума «Еліза». Експеримент М. Макгайра.	4
6	Джерела знань, необхідні для визначення анафори. Власнелінгвістичні та екстралінгвістичні.	4
7	Проблеми на шляху комп'ютерного опрацювання природної мови	4
8	Розпізнавання мовлення: мета, проблеми та практичне значення	4
9	Переведення усного мовлення в текст: мета, проблеми та практичне значення	4
10	Спрощення тесту: мета, проблеми та практичне значення	4
11	Комп'ютерний переклад: рівні, проблеми та практичне значення	4
12	Український мовно-інформаційний фонд: структура, проекти, перспективи	4
13	Інтегрована лексикографічна система УМІФ словники України http://corp.ulif.org.ua/dictua : структура, функції, практичне застосування	4
14	Лінгвістичний портал MOVA.info : структура, функції, практичне застосування	4
	Разом	58

6. Методи навчання

Пояснювально-ілюстративний метод, метод проблемного викладання, евристичний метод, дослідницький метод.

7. Методи контролю

Контроль засвоєння матеріалу включає поточний контроль (контрольні роботи за двома змістовими модулями, $15 \times 2 = 30$ балів), оцінку самостійної роботи (20 балів) — разом за семестр 50 балів — та залік, що оцінюється в 50 балів. Сумарна оцінка, таким чином, виставляється за 100-бальною шкалою.

8. Розподіл балів, що присвоюються студентам

Оцінювання роботи студента здійснюється за схемою:

1. Максимальна оцінка, яку може отримати студент, – 100 балів, з них 50 – за роботу протягом семестру, 50 – за знання, продемонстровані на заліку.
2. Робота студента протягом семестру оцінюється за показниками:
 - а) самостійна робота – 20 балів;

- б) два модулі (поточне тестування) по 15 балів – 30 балів.
3. Підсумковий залік є обов'язковим і проводиться в комбінованій (усній і письмовій) формі. Студент повинен письмово виконати тестове завдання, яке оцінюється 30 балами, а також відповісти на 2 теоретичні питання, кожне з яких оцінюється 10 балами.

1. Розподіл балів, які отримують студенти

Поточне тестування та самостійна робота		Самостійна робота	Підсумковий тест (залік)	Сума
Змістовий модуль 1	Змістовий модуль 2			
Т.1-Т.3	Т.4-Т.6	20	50	100
15	15			

Шкала оцінювання: національна та ECTS

Сума балів за всі види навчальної діяльності	Оцінка ECTS	Оцінка за національною шкалою	
		для екзамену, курсового проекту (роботи), практики	для заліку
90 – 100	A	відмінно	зараховано
81-89	B	добре	
71-80	C		
61-70	D	задовільно	
51-60	E		
21-50	FX	незадовільно з можливістю повторного складання	не зараховано з можливістю повторного складання
0-20	F	незадовільно з обов'язковим повторним вивченням дисципліни	не зараховано з обов'язковим повторним вивченням дисципліни

9. Методичне забезпечення

1. *Бовтенко М. А.* Комп'ютерна лінгводидактика. — М., 2005. 216 с.
2. *Гузь І.О.* Методичні матеріали щодо забезпечення самостійно роботи студентів з дисципліни “Основи інформатики та прикладної лінгвістики” (для бакалаврів)ю — К.: ДП «Вид. дім «Персонал», 2013. — 67 с.

10. Рекомендована література

Базова

1. *Карпіловська Є. А.* Вступ до комп'ютерної лінгвістики.— Донецьк: Юго-Восток, 2003.
2. *Дарчук Н. П.* Комп'ютерна лінгвістика (автоматичне опрацювання тексту): підручник.— К.: Видавничо-поліграфічний центр “Київський університет”, 2008.
3. *Пецак М. М.* Нариси з комп'ютерної лінгвістики.— Ужгород: Видавництво закарпаття, 1999.— 200 с.

4. *Баранов А. Н.* Введение в прикладную лингвистику: Учебное пособие. — М.: Эдиториал УРСП, 2001.

Допоміжна

1. *Волошин Володимир Григорович* Комп'ютерна лінгвістика: Навч. посібник.— Суми: Університетська книга, 2004.
2. *Лук'янчук С.* Комп'ютерна модель парадигматичних класів дієслів // Українське мовознавство.— 2000.— Вип. 22.—С. 82–85.
3. *Bolshakov, Igor. A., Gelbukh Alexander* Computational Linguistics. Models, Resources, Applications.— Мéxico, 2004.
4. The Oxford Handbook of Computation Linguistics / ed. by *Ruslan Mitkov*.— Oxford Un-ty Press, 2003.

11. Інформаційні ресурси

1. Framenet <http://framenet.icsi.berkeley.edu/>
2. Wordnet <http://wordnet.princeton.edu/>
3. Русский Wordnet (структура) <http://www.pgups.ru/WebWN/wordnet.uix>
4. Лінгвістичний портал MOVA.info
5. Інтегрована лексикографічна система УМІФ словники України <http://lcorp.ulif.org.ua/>
6. <https://www.academia.edu/>
7. Library of Congress. [electronic resource]. – access mode : <http://www.loc.gov/>